energies

MDPI

Article

Short-Term Load Forecasting for Secondary Substations in Electrical Distribution Networks

Diego Andreotti *, Matteo Spiller, Giuliano Rancilio and Marco Merlo

Department of Energy, Politecnico di Milano, Via Lambruschini 4, I-20156 Milano, Italy; matteo.spiller@polimi.it (M.S.); giuliano.rancilio@polimi.it (G.R.); marco.merlo@polimi.it (M.M.)
* Correspondence: diego.andreotti@polimi.it

Abstract

The increasing penetration of distributed energy resources (DERs) and new electric loads associated with decarbonization is pushing Distribution System Operators (DSOs) towards more proactive management of Medium Voltage (MV) and Low Voltage (LV) networks. In this context, short-term load forecasting (STLF) at the secondary substation (SS) level is becoming increasingly relevant for network operation and planning. However, conventional approaches typically use exclusively the data at the SS level, without exploiting information available across the distribution network. This paper proposes a data-driven hierarchical forecasting framework that combines substation-level predictions with aggregated forecasts of the underlying connection points. SS are characterized according to their operating conditions, based on the balance between annually consumed and produced energy, to investigate how these affect predictability. Forecasts from one to five days ahead are obtained using horizon-specific Random Forest (RF) models combining autoregressive, meteorological, and calendar information. Minimum Trace (MinT) reconciliation then ensures coherence between substation-level and connection-point forecasts. The analysis reveals markedly different forecasting behavior across operating conditions, with consumption-dominated substations proving considerably more predictable than their generation-dominated counterparts, as reflected in median day-ahead weighted absolute percentage errors of 9.2% and 79%, respectively. Hierarchical reconciliation follows the same pattern, delivering its most consistent gains for passive substations, with a median day-ahead MAE reduction of 2.8%, reaching up to 4.6%, while its benefit gradually fades as local generation grows. The results provide DSOs with practical indications for adapting forecasting and reconciliation strategies across heterogeneous distribution networks.

Keywords: distribution networks; hierarchical reconciliation; load forecasting; machine learning; power distribution; random forests

Check for updates

Academic Editor: Marcin Kaminski

Received: 22 September 2026
Revised: 28 October 2026
Accepted: 3 November 2026
Published: 7 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In recent years, power distribution systems have undergone significant transformations due to the increasing integration of Renewable Energy Sources (RESs), Energy Storage Systems (ESSs), and the electrification of sectors such as transport and heating through Electric Vehicles (EVs) and Heat Pumps (HPs) [1,2]. These Distributed Energy Resources (DERs), increasingly connected to Medium Voltage (MV) and Low Voltage (LV) networks, are progressively reshaping the operating conditions of distribution grids by introducing more variable, heterogeneous, and potentially bidirectional power flows [3]. Consequently, Distribution System Operators (DSOs) are required to move toward a more proactive

management approach [4]. The ability to anticipate the evolution of electrical demand and power flows becomes increasingly important to identify potential operating issues before they occur and to plan suitable corrective actions in advance. Accurate short-term load forecasting (STLF) therefore represents a key enabling tool for distribution system operation [5]. Forecasts ranging from the next few hours to a few days ahead can support the early identification of critical operating conditions, the scheduling of flexibility resources, and the efficient use of existing network infrastructure. However, the operational value of these forecasts depends not only on their temporal horizon, but also on the spatial level at which load is predicted.

Distribution system operation involves decisions at different aggregation levels, ranging from the overall network to individual feeders, substations, and local network areas [6]. Consequently, the spatial granularity of load forecasts plays a crucial role in determining the type of operational information they can provide [7]. While highly aggregated forecasts provide an overall view of expected system loading, forecasts at lower aggregation levels allow local variations in demand and potential network constraints to be captured more effectively.

1.1. State-of-the-Art and Research Objectives

The problem under investigation is schematically illustrated in Figure 1. DSOs can access and manage measurements collected by the meters installed at the end users' connection points, as well as measurements provided by the meter associated with the MV/LV distribution transformer. The objective is to forecast the power profile measured at the transformer level, which represents the aggregated power exchange between the LV distribution network and the upstream MV grid resulting from the consumption and generation of the connected users. Such information could enable more efficient and reliable operation of the MV grid.

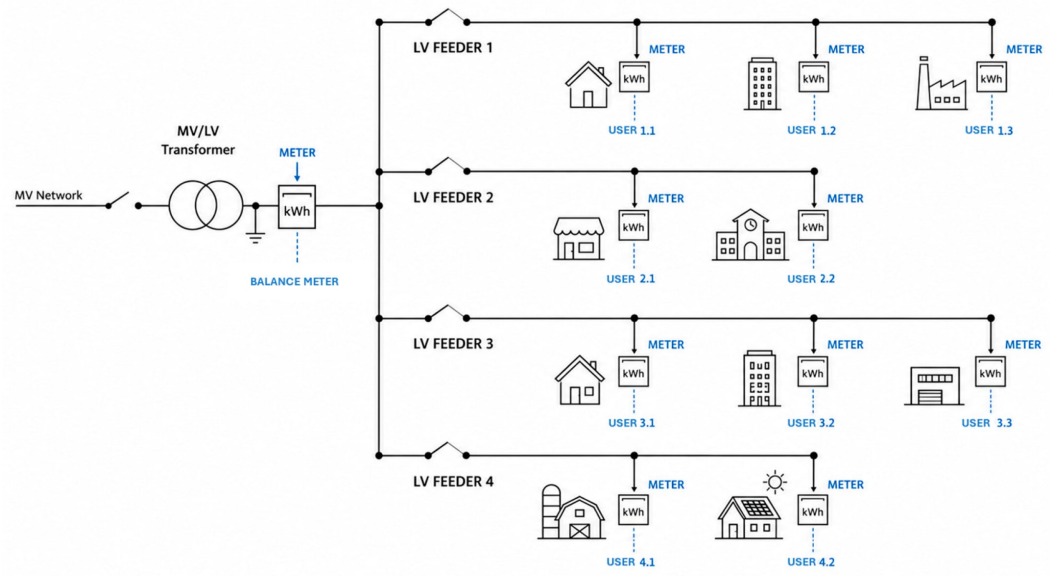


Figure 1. Reference structure of the LV grids under investigation.

Considerable research has been devoted to STLF in distribution networks. Conventional statistical approaches, including linear regression and autoregressive models, remain widely used because of their limited computational requirements and straightforward implementation [8]. Autoregressive terms are often combined with calendar and meteorological information to capture temporal periodicities and weather-dependent variations [8,9]. More recently, machine learning techniques have gained increasing attention because of

their ability to model nonlinear relationships between historical measurements and heterogeneous input variables. Random Forest (RF) and Gradient Boosting (GB) have been successfully applied to distribution-level forecasting, including MV feeders and distribution transformers [8,10], while Artificial Neural Networks (ANNs) and recurrent architectures, particularly Long Short-Term Memory (LSTM) networks, have been investigated to capture more complex temporal dependencies [11,12]. Global forecasting approaches, where a single model is trained across multiple distribution-level time series, have also been proposed to improve scalability by exploiting common patterns among different loads [13]. Neural-network-based methods have further been applied to load forecasting at LV substations with distributed generation [14].

Despite these advances, forecasting at the distribution level remains challenging, as lower aggregation levels are characterized by increasingly heterogeneous consumption and generation patterns [15]. As illustrated in Figure 1, different types of users may be connected to the LV feeders, and different distribution grids may therefore exhibit heterogeneous user compositions. Consequently, forecasting performance can vary significantly with the characteristics of the considered load or generation profile [16–19]. This heterogeneity becomes particularly relevant when forecasting methods are applied to large populations of secondary substations (SSs) [13,20].

A further challenge arises when forecasts are produced across multiple aggregation levels of the distribution network. Forecasts obtained independently for individual customers, substations, and higher-level aggregates are not necessarily coherent with the physical aggregation structure of the network. Forecast reconciliation methods address this issue by adjusting independently generated base forecasts to enforce consistency across aggregation levels [21]. Among them, the Minimum Trace (MinT) method proposed in [22] exploits the covariance structure of forecast errors to obtain coherent predictions while minimizing the overall forecast error variance. Its application to distribution systems has been investigated in [23], where forecasts at customer, phase, and distribution transformer levels were reconciled using Advanced Metering Infrastructure (AMI) data. However, reconciliation introduces additional computational requirements, which may become critical when the procedure is deployed over a large number of substations.

Although the literature provides a wide range of forecasting and reconciliation techniques, fewer studies jointly address heterogeneous SSs, multi-day forecasting, and hierarchical reconciliation within the same framework. Practical DSO applications therefore require forecasting approaches that combine accuracy, robustness across heterogeneous operating conditions, and computational scalability. In this context, it is also important to assess how forecasting performance varies across substations with different consumption and generation characteristics and, consequently, under which conditions the resulting forecasts can be considered reliable.

1.2. Paper Contribution

The gaps just highlighted motivate the analysis performed in this paper, which develops and evaluates a forecasting framework combining RF regression and MinT reconciliation. The proposed approach is tailored to real-world metering data and systematically assessed across SSs with heterogeneous operating characteristics, with particular emphasis on the forecasting requirements of DSOs. The main contributions of this work are as follows:

- A characterization of SSs according to their consumption and generation behavior, distinguishing passive, hybrid, and active operating profiles.
- The development of a STLF framework over a five-day horizon, based on autoregressive, meteorological, and calendar features.

- The assessment of the forecasting methodology on real distribution system measurements over an extended one-year time window.
- The formulation of a measurement-based hierarchy in which the residual between the substation meter and the connection-point meters (error index) is modeled as an explicit node, which allows MinT reconciliation to be applied to real metering data, together with an assessment of its accuracy benefit across operating classes and forecasting horizons and of its computational cost.

The rest of the paper is organized as follows: Section 2 presents the proposed methodology, including SSs classification, short-term forecasting, and hierarchical reconciliation. Section 3 presents the case study and discusses the results in terms of SSs classification, forecasting performance across the different operating classes, forecasting drivers, the impact of hierarchical reconciliation, and the robustness of the results with respect to the evaluation period. Finally, Section 4 summarizes the main findings, conclusions, and limitations.

2. Methodology

The proposed methodology is organized into sequential stages. Each SS is first classified based on the ratio between its annual energy consumption and production. The classification is then complemented by an autocorrelation analysis of the substation power profile, which characterizes the temporal structure of each class and supports the interpretation of the forecasting results. Second, a five-day-ahead STLF framework is developed based on a family of RF models, with a dedicated feature engineering and forecasting for each day. Third, a hierarchical reconciliation stage based on the MinT method is applied to ensure coherence between the substation level forecast and the forecasts of the underlying connection point clusters, while assessing its impact on forecasting performance. The three stages are detailed in the following subsections.

2.1. Secondary Substation Classification

The dataset provided by the DSO consists of real-world measurements collected during the operation of its distribution network. For each SS, it includes measurements from the balance meter on the MV side of the transformer, describing the aggregated active power consumed (A1) and produced (A2), together with the corresponding measurements from all underlying connection points. All measurements cover the whole of 2025 at quarter-hourly resolution (35,040 samples per series). Given the heterogeneous operating conditions across substations, a preliminary classification is introduced to characterize their different load and generation profiles.

Let E^{A1} and E^{A2} denote the energy consumed and produced by the balance-metering point over the reference year, obtained by integrating the corresponding quarter-hourly power profiles as in Equation (1):

$$E^{A1} = \sum_{t=1}^{T} P_t^{A1} \Delta t, \quad E^{A2} = \sum_{t=1}^{T} P_t^{A2} \Delta t \tag{1}$$

where $T = 35{,}040$ is the number of quarter-hour samples in one year and $\Delta t = 0.25$ h. The classification index ρ is defined in Equation (2) as the ratio between consumed and produced energy:

$$\rho = \frac{E^{A1}}{E^{A2}} \tag{2}$$

Each substation is then assigned to one of three classes:

- Passive: $\rho > 10$;
- Hybrid: $1 \leq \rho \leq 10$;
- Active: $\rho < 1$.

The resulting classes represent different balances between consumption and local generation. Passive substations are characterized by energy consumption largely exceeding production, hybrid substations exhibit more comparable levels of consumed and produced energy, whereas active substations are characterized by annual production exceeding consumption.

The two thresholds have a clear physical interpretation. The threshold ρ distinguishes net energy importers from net exporters on an annual basis. Below this threshold, the substation injects more energy into the MV grid than it withdraws, meaning that reverse power flow dominates annual energy exchanges. The value $\rho = 10$ identifies substations in which local production does not exceed 10% of consumption, so that generation acts only as a minor perturbation of an otherwise conventional load profile. As shown in Section 3.2, the analyzed substations form three well-separated groups in terms of ρ, so that the resulting classification does not depend on the exact values of the thresholds.

Beyond describing the operating characteristics of the substations, this classification provides a basis for investigating whether different consumption-generation patterns are associated with different levels of predictability. To make this link explicit, the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the balance series are computed.

For a stationary series $\{x_t\}$ of mean $\overline{x}$, the Correlation Function (CF) measures the linear dependence between observations separated by k time steps. The sample autocorrelation at lag k is given by Equation (3).

$$r_k = \frac{\sum_{t=1}^{T-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{T}(x_t - \overline{x})^2} \tag{3}$$

While the ACF captures both direct and indirect correlations transmitted through intermediate lags, the PACF isolates the linear dependence between x_t and x_{t-k} after removing the effects of the intermediate lags. To obtain the PACF, an autoregressive model of order k is considered, as expressed in Equation (4).

$$x_t = c + \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \cdots + \phi_{kk}x_{t-k} + \varepsilon_t \tag{4}$$

The autoregressive coefficients $\phi_{k1}, \ldots, \phi_{kk}$ are determined by solving the Yule-Walker equations reported in Equation (5).

$$\begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \ldots & r_{k-2} \\ r_2 & r_1 & 1 & \ldots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \ldots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_k \end{bmatrix} \tag{5}$$

The PACF at lag k is then given by ϕ_{kk}, as expressed in Equation (6).

$$PACF(k) = \phi_{kk} \tag{6}$$

The ACF and PACF therefore characterize the linear dependence of the balance series on its past values and provide an indication of the relevance of autoregressive information for forecasting.

2.2. Short-Term Load Forecasting Framework

The forecasting framework provides load predictions from one to five days ahead. Because the information available at the prediction instant differs for each horizon, a single multi-output model is inadequate. Multiple-input multiple-output architectures were

considered but discarded due to their higher computational cost [24]. Instead, five independent RF regressors $\{RF_k\}$, $k = 1, \ldots, 5$, have been trained, each associated with a specific forecasting horizon from D-1 to D-5.

2.2.1. Feature Engineering

For each horizon k, three families of features are used to predict the target series:

- Autoregressive terms;
- Weather variables;
- Calendar variables.

The autoregressive terms are constructed from two families of past values, motivated by the two dominant periodicities of the load. These include daily and weekly lags to capture short-term persistence and recurring weekly patterns. Lag selection depends on the forecasting horizon k. Since observations between days D and D-k are unavailable, only lags satisfying $\ell \geq 96k$ are used. The resulting lags for the five models are reported in Table 1.

Table 1. Autoregressive lags retained for each forecasting model.

Model	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
RF_1 (D-1)	96 (1d)	192 (2d)	288 (3d)	384 (4d)	672 (1w)
RF_2 (D-2)	192 (2d)	288 (3d)	384 (4d)	480 (5d)	672 (1w)
RF_3 (D-3)	288 (3d)	384 (4d)	480 (5d)	576 (6d)	672 (1w)
RF_4 (D-4)	384 (4d)	480 (5d)	576 (6d)	672 (1w)	1344 (2w)
RF_5 (D-5)	480 (5d)	576 (6d)	672 (1w)	1344 (2w)	2016 (3w)

The weather features are retrieved from the OpenMeteo service by querying the geographic coordinates of each substation [25]. Archived operational forecasts are used to ensure consistency with the information available at the time of prediction. For each forecasting horizon D-k, both training and testing use forecasts issued k days before the target time, preventing the use of information that would not have been available when the load forecast was issued. Seven variables are collected as follows:

i. Air temperature at 2 m;
ii. Relative humidity;
iii. Precipitation;
iv. Wind speed at 10 m;
v. Wind direction at 10 m;
vi. Surface pressure;
vii. Shortwave radiation.

Finally, calendar information is represented through cyclical sine-cosine encoding. Denoting by $q \in \{0, \ldots, 95\}$ the quarter-hour of the day (period 96), by $w \in \{0, \ldots, 6\}$ the day of the week (period 7), and by $m \in \{1, \ldots, 12\}$ the month (period 12), the three variables are encoded as in Equation (7)–(9):

$$\left(\sin\frac{2\pi q}{96}, \cos\frac{2\pi q}{96}\right) \tag{7}$$

$$\left(\sin\frac{2\pi q}{96}, \cos\frac{2\pi q}{96}\right) \tag{8}$$

$$\left(\sin\frac{2\pi q}{96}, \cos\frac{2\pi q}{96}\right) \tag{9}$$

The trigonometric features are then complemented by a binary indicator identifying weekends and national holidays, thereby accounting for differences between working and non-working days.

2.2.2. Model, Training and Validation

RF is adopted as the base learner for its robustness to non-linearities and overfitting [26]. It combines multiple regression trees, with the final forecast obtained as their ensemble average, as given by Equation (10):

$$\hat{y}_t = \frac{1}{B}\sum_{b=1}^{B} h_b(x_t) \tag{10}$$

where h_b is the b-th tree and x_t the engineered feature vector at instant t. Averaging predictions across multiple, weakly correlated trees reduces prediction variance and improves the stability of the resulting model [26].

The forecasting procedure is applied separately to the two power-flow directions, with independent RF models developed for withdrawn (A1) and injected (A2) power, allowing the forecasting performance to be evaluated according to the operating class of each SS.

The first 305 days of 2025 (about 84%) are used for training, and the last 60 days (2 November–31 December, about 16%) are used as a test set, while the last 60 days of the training period (3 September–1 November) serve as a validation window for hyperparameter selection, preserving the temporal order of the series. This approach prevents information leakage by ensuring that future observations are not used during model training [27]. To assess the robustness of the results across different test periods, a rolling-origin evaluation is also conducted at the D-1 horizon (Section 3.3). Six forecast origins are considered at monthly intervals from July to December. At each origin, the models are retrained on all preceding data and evaluated over the following month.

Model accuracy is assessed on the validation set through different metrics. Denoting by y_i the measured value, by $\hat{y}_i$ the forecast, by $\overline{y}$ the mean of the measurements, and by n the number of test samples, the metrics are defined in Equation (11)—(14):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{11}$$

$$MAPE = \frac{100}{|\mathcal{J}|}\sum_{i\in\mathcal{J}}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{12}$$

$$WAPE = 100\,\frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i|} \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{14}$$

The Mean Absolute Error (MAE) quantifies the absolute magnitude of the forecasting error. The Mean Absolute Percentage Error (MAPE) expresses the error relative to the measured values, while R^2 measures the share of the target variance reproduced by the model. Since the MAPE diverges when the measured power approaches zero, which systematically occurs for injected power at night and during low-wind periods, it is computed only on the set $\mathcal{J}$ of samples whose absolute value exceeds 5% of the peak of the test series; the share of

samples retained (coverage) is reported where relevant. The weighted absolute percentage error (WAPE) is always defined, and is therefore used to compare series with near-zero values and substations of different sizes.

The RF forecasts are evaluated against a seasonal naive benchmark that requires no training and uses only information available at forecast issuance. The daily naive benchmark assigns each quarter-hour the value recorded at the same time on the most recent observed day. The improvement over this benchmark is expressed by the skill score, Skill = 100·(1–$WAPE_{RF}/WAPE_{naive}$), which is positive when the RF outperforms the benchmark.

RF hyperparameters, such as the number of trees, the maximum depth, and the minimum number of samples per leaf, are tuned by a grid search over the values listed in Table 2.

Table 2. Random Forest hyperparameter grid explored by the grid-search procedure.

No. of Trees	Max Depth	Min Samples/Leaf	Max Features
200	10	1	"sqrt"
300	20	5	0.5
500	30	10	1
-	None	-	-

The hyperparameter grid search was performed at the D-1 horizon on three series selected to represent the different classes and power-flow directions: SS1 (A1), SS12 (A2), and SS2 (A2). Although the optimal hyperparameters varied across these series, a common configuration was adopted for all models to ensure a consistent setup. For the three series considered, this configuration yielded validation MAE values only 1.5–4.1% higher than their respective minima. The selected settings were 300 trees, a maximum depth of 20, a minimum of 10 samples per leaf, all features considered at each split, and bootstrap sampling. A fixed random seed was used to ensure reproducibility.

2.3. Hierarchical Reconciliation Framework

Forecasts generated independently at different aggregation levels do not necessarily satisfy the physical consistency constraints of the network. Hierarchical reconciliation addresses this issue by adjusting the base forecasts according to the underlying aggregation structure. In the considered setting, the SS represents the aggregate level, while the connection points constitute its underlying components. Accordingly, at each time step t, the power measured at the substation can be expressed through Equation (15):

$$P_t^{SS} = Err_t + \sum_{k=1}^{N_{Users}} P_{k,t}^{Users} \quad (15)$$

where Err_t, the error index, is the measured residual between the substation meter and the sum of the connection-point meters, computed separately for A1 and A2. It collects losses, unmetered loads, and metering errors. Being measured over the whole historical period, it is forecast by its own RF like any other node; including it as a bottom-level node makes the aggregation constraint exact on real metering data. For reconciliation, its predictability is more relevant than its magnitude, as an accurate forecast allows this discrepancy to be effectively accounted for.

2.3.1. Clustering of the Connection Points

Since each SS may include a large number of underlying users, forecasting them individually would significantly increase the computational effort. The connection points

are therefore grouped according to their maximum recorded power P_{max} into three classes, defined by $P_{max} \leq 3.3$ kW, $3.3 < P_{max} \leq 6.6$ kW and $P_{max} > 6.6$ kW. $P_{\max}$ is the maximum quarter-hourly power recorded by each connection-point meter over the year, on withdrawn power for A1 and on injected power for A2. The thresholds correspond to the common Italian contractual powers, so that the clusters approximately group customers by contractual power.

Within each class, the individual profiles are summed to obtain an aggregated time series, for which a dedicated RF model is trained.

The base forecasts are collected in the vector $\hat{y}_t$, defined in Equation (16), which includes the SS forecast, the three cluster forecasts and the error index forecast.

$$\hat{y}_t = \left[\hat{y}_t^{SS} \quad \hat{y}_t^{<3.3\text{ kW}} \quad \hat{y}_t^{3.3-6.6\text{ kW}} \quad \hat{y}_t^{>6.6\text{ kW}} \quad \hat{y}_t^{Err}\right] \tag{16}$$

The coherence constraint is represented through the summing matrix S, which maps the bottom-level vector b_t, containing the three cluster series and the error index, into the complete vector y_t, as defined in Equation (17).

$$S = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; y_t = S\, b_t, \; b_t = \begin{bmatrix} y_t^{<3.3\text{ kW}} \\ y_t^{3.3-6.6\text{ kW}} \\ y_t^{>6.6\text{ kW}} \\ y_t^{Err} \end{bmatrix} \tag{17}$$

According to this structure, coherence requires the SS forecast to equal the sum of the forecasts associated with the three connection-point clusters and the error index.

2.3.2. Minimum Trace Reconciliation

Since the independently generated base forecasts do not necessarily satisfy this constraint, they are reconciled by projecting them onto the coherent space defined by S. The reconciled forecast vector $\tilde{y}_t$ is obtained according to Equation (18), where G is an $m \times n$ matrix mapping the n base forecasts into the m bottom-level series ($n = m + 1$).

$$\tilde{y}_t = S\, G\, \hat{y}_t = P\hat{y}_t \tag{18}$$

The MinT approach determines G by minimizing the total variance of the reconciled forecast errors while preserving unbiasedness. Let W denote the covariance matrix of the base forecast errors $e_t = y_t - \hat{y}_t$. Under the constraint $GS = I$, the MinT optimization problem can be written according to Equation (19).

$$\min_G tr\left(SGWG^T S^T\right), \;\; with\; GS = I \tag{19}$$

Solving this problem with Lagrange multipliers yields the reconciliation matrix in Equation (20):

$$G = \left(S^T W^{-1} S\right)^{-1} S^T W^{-1} \tag{20}$$

The covariance matrix W is estimated from a historical sample of T base forecast errors. At each time step t, the errors associated with the SS, the three connection-point clusters, and the error index are collected in the vector e_t. The sample covariance estimator is therefore given by Equation (21).

$$W = \frac{1}{T}\sum_{t=1}^{T} e_t e_t^T, \;\; e_t = \left[e_t^{SS} \quad e_t^{<3.3\text{ kW}} \quad e_t^{3.3-6.6\text{ kW}} \quad e_t^{>6.6\text{ kW}} \quad e_t^{Err}\right]^T \tag{21}$$

By accounting for both individual error variances and cross-covariances, W allows MinT to exploit the joint error structure of the base forecasts. The resulting forecasts are coherent by construction, ensuring consistency between the SS forecast and its lower-level

components. The reconciliation framework requires additional forecasting models for the three connection-point clusters and the error index. Most of the associated computational burden therefore arises from generating the additional base forecasts, whereas the MinT projection itself introduces a comparatively limited computational overhead. This aspect becomes particularly relevant when the methodology is extended to a large population of substations and multiple forecasting horizons.

3. Results

This section evaluates the proposed methodology on a real distribution system dataset. The analysis covers the identification of substation operating profiles, the short-term forecasting performance, and the impact of hierarchical reconciliation on forecast accuracy. All computations were performed on a ThinkPad T16 Gen 2 laptop (Lenovo, Beijing, China) with an Intel Core i7-1355U processor (10 cores, 12 threads, up to 5.0 GHz; Intel Corporation, Santa Clara, CA, USA), 32 GB of RAM, and Windows 11Pro (Microsoft Corporation, Redmond, WA, USA).

3.1. Case Study and Data

The methodology is assessed on a dataset comprising 17 SSs located in central-southern Italy, whose geographical distribution is shown in Figure 2. For each SS, the dataset covers one year of active power measurements with a 15 min resolution, including both consumed (A1) and produced (A2) power measured at the MV side of the transformer, together with the corresponding measurements of all underlying connection points.

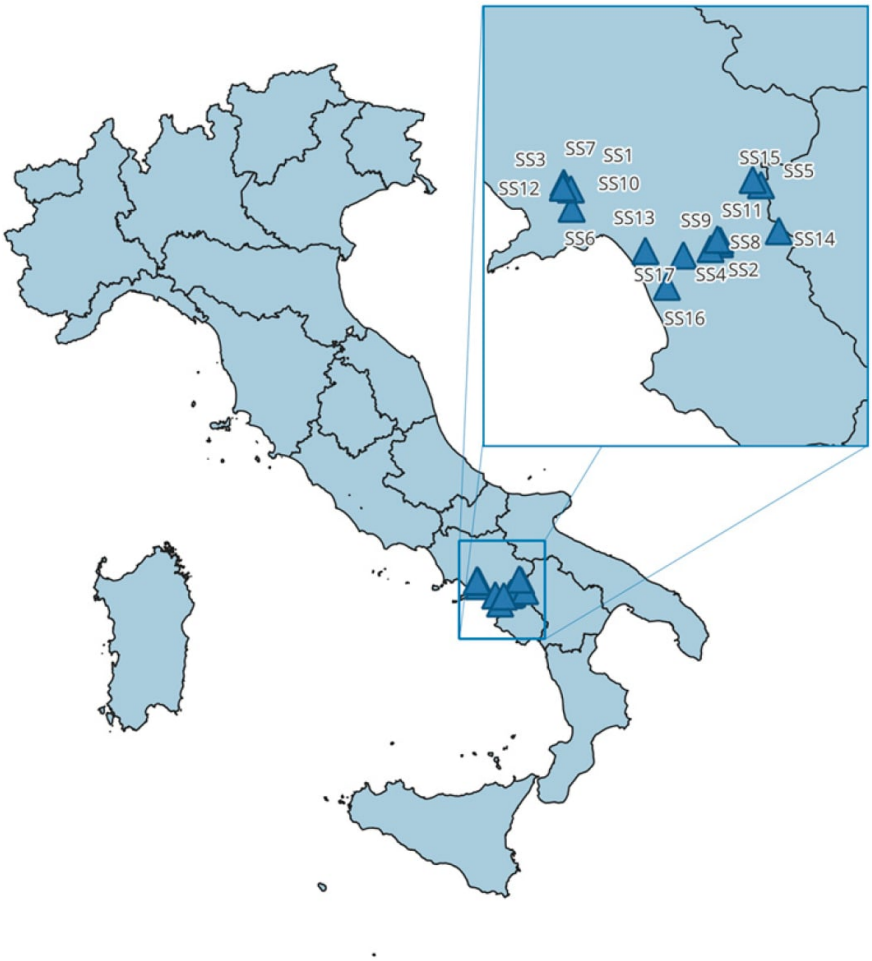


Figure 2. Geographical distribution of the analyzed secondary substations.

The composition of the analyzed SSs is shown in Figure 3 in terms of the number of connection points associated with consumption (A1) and production (A2).

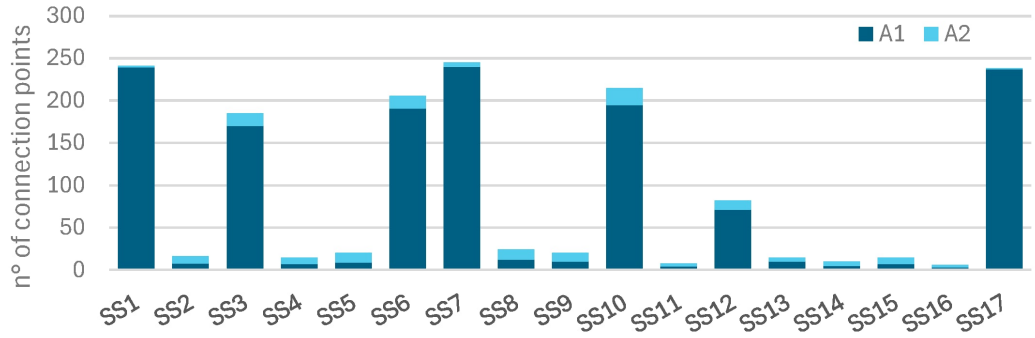


Figure 3. Distribution of withdrawn (A1) and injected (A2) connection points across the analyzed secondary substations.

The dataset includes datasets with markedly different sizes and user compositions, ranging from small SSs with only a few connection points to larger ones supplying more than 200 connection points. Moreover, the relative presence of consumption and production connection points varies considerably across the analyzed SSs, providing a heterogeneous set of operating conditions for the subsequent forecasting analysis. All substations belong to a single DSO and area, which limits the generality of the quantitative results. Nevertheless, the SS spans the three operating classes defined in the methodology, a wide range of sizes, and both PV and wind generation.

3.2. Secondary Substation Classification Results

Applying the energy-ratio criterion defined in Equation (2) to the 17 SSs reveals markedly different operating conditions across the analyzed distribution area. Figure 4 shows, for each substation, the annual consumed energy (A1), produced energy (A2), their ratio ρ, and the resulting operating class.

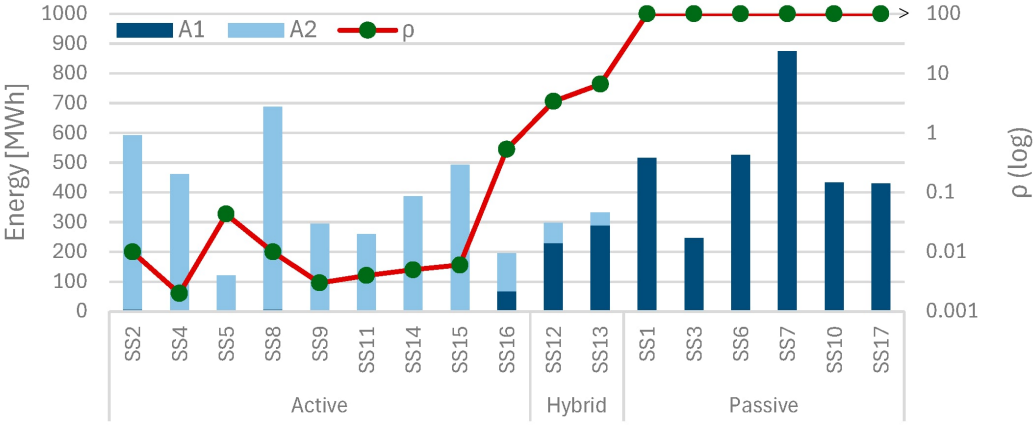


Figure 4. Energy balance and operating classification of the analyzed secondary substations.

Overall, the analysis distinguishes consumption-dominated passive substations, generation-dominated active substations, and intermediate hybrid cases where both consumption and local generation contribute significantly to the annual energy balance. These differences are also reflected in the measured power profiles. As shown by the representative examples in Figure 5, passive substations exhibit a regular and recurrent daily pattern in withdrawn power (A1), whereas injected power (A2) is shown for hybrid and active sub-

stations. Hybrid substations display a recognizable daily generation pattern, concentrated during daylight hours and characterized by a bell-shaped profile, consistent with a predominant photovoltaic contribution. Conversely, active substations exhibit more irregular generation, with large variations throughout the day and no clearly repeatable daily pattern, consistent with the predominance of wind generation. The temporal profiles therefore complement the annual energy balance by revealing distinct operating patterns across the three classes. These differences are further confirmed by the corresponding autocorrelation functions. Passive substations exhibit pronounced ACF peaks at approximately 24 h and its multiples, reflecting the strong daily periodicity of consumption. A similarly clear structure is observed for hybrid substations, where photovoltaic generation produces recurring daily peaks. In contrast, active substations show a much weaker autocorrelation structure, with a rapid decay and limited periodicity at longer lags, reflecting the greater variability of wind generation.

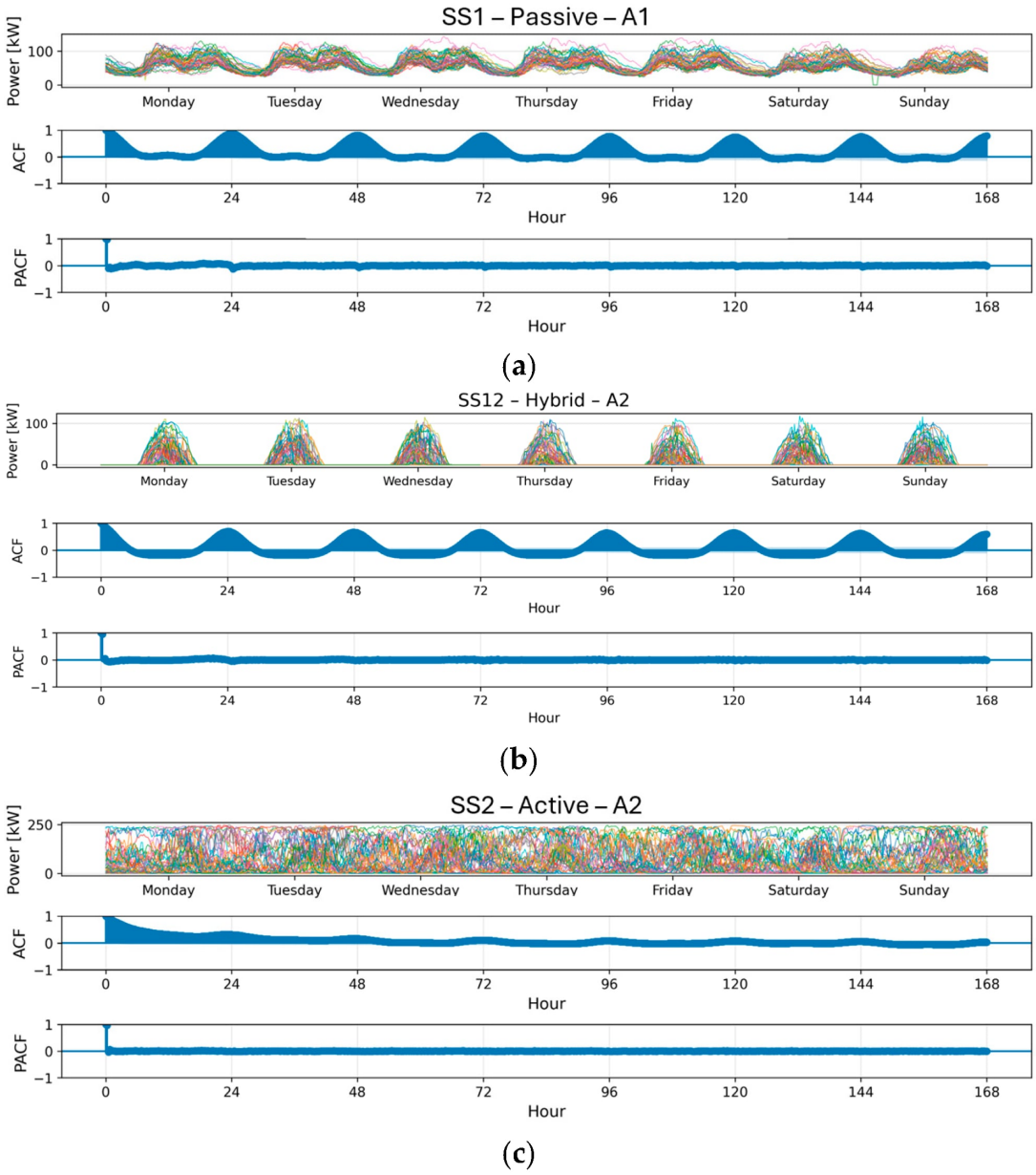


Figure 5. Weekly power profiles with the corresponding ACF and PACF for (**a**) a passive (SS1), (**b**) a hybrid (SS12), and (**c**) an active (SS2) secondary substation.

The classification therefore captures not only different balances between consumption and generation, but also distinct temporal structures that directly affect forecasting. The strong autocorrelation of passive and photovoltaic-dominated hybrid substations provides useful information for autoregressive forecasting, whereas wind-dominated active substations are less predictable from historical power measurements alone. These differences provide a physical and statistical basis for interpreting the forecasting performance of the three classes in the following analysis.

3.3. Short-Term Load Forecasting

Forecasting performance is analyzed separately for passive, hybrid, and active SSs to highlight the differences associated with their operating characteristics.

3.3.1. Passive Substations

Passive SSs represent the conventional load forecasting case, as their power exchange is predominantly determined by the consumption of the underlying connection points. Their profiles exhibit regular daily and weekly patterns, providing a strong temporal structure that can be exploited by the autoregressive and calendar features of the forecasting framework.

Figure 6 shows an example for SS1 over five days of the validation period. The five horizon-specific models reproduce the main evolution of the measured load, including the overnight minimum, the morning increase, and the daily load cycles. Differences among the forecasts become more evident during the periods of rapid load variation and around the daily peaks. The lower panel confirms that the largest errors occur mainly during these high load periods, whereas the forecasts remain closer to the measurements during the more stable overnight conditions.

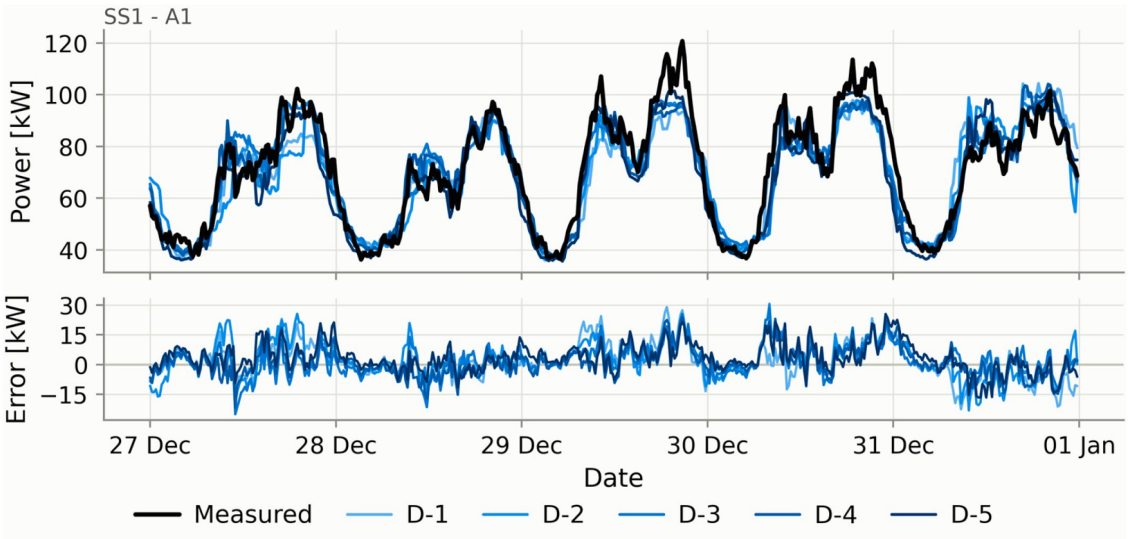


Figure 6. Validation for the passive substation SS1: measured load (A1) and the five model predictions (top) and the corresponding errors (bottom).

Table 3 extends the analysis to all passive substations and reports MAE, MAPE, WAPE, and R^2 for the five forecasting horizons. At D-1, the MAE ranges from 4.37 kW for SS3 to 8.14 kW for SS7, while the MAPE remains between 7.04% and 14.50%. The performance progressively deteriorates as the forecast horizon increases. At D-5, the MAE ranges from 5.01 to 12.02 kW and the MAPE from 9.87% to 16.08%. This trend is observed across all six passive substations.

Table 3. Test set MAE, MAPE, WAPE, R2 of the 5 forecasting models (D-1 to D-5) for passive substations.

	Metric	D-1	D-2	D-3	D-4	D-5
SS1	MAE [kW]	5.829	6.168	6.293	6.785	8.114
	MAPE [%]	8.71	9.07	9.37	10.16	11.73
	WAPE [%]	9.02	9.55	9.74	10.50	12.56
	R^2 [-]	0.868	0.858	0.852	0.828	0.748
SS3	MAE [kW]	4.365	4.529	4.424	4.717	5.012
	MAPE [%]	14.50	15.10	14.57	15.09	15.67
	WAPE [%]	13.98	14.51	14.17	15.11	16.05
	R^2 [-]	0.763	0.742	0.751	0.704	0.659
SS6	MAE [kW]	6.596	7.026	7.235	7.818	9.240
	MAPE [%]	8.97	9.44	9.68	10.27	11.97
	WAPE [%]	9.46	10.08	10.38	11.21	13.25
	R^2 [-]	0.807	0.782	0.775	0.733	0.649
SS7	MAE [kW]	8.136	8.743	9.222	10.597	12.025
	MAPE [%]	7.04	7.50	7.91	8.84	9.87
	WAPE [%]	7.42	7.98	8.41	9.67	10.97
	R^2 [-]	0.900	0.878	0.865	0.821	0.770
SS10	MAE [kW]	6.915	7.344	7.535	9.557	10.422
	MAPE [%]	11.68	12.18	12.42	14.97	16.08
	WAPE [%]	11.65	12.37	12.69	16.09	17.55
	R^2 [-]	0.821	0.792	0.775	0.648	0.588
SS17	MAE [kW]	4.661	4.819	4.921	5.119	6.051
	MAPE [%]	8.41	8.63	8.80	8.98	10.18
	WAPE [%]	8.82	9.12	9.32	9.69	11.46
	R^2 [-]	0.877	0.867	0.863	0.847	0.783

Overall, the results indicate that the proposed approach provides consistent short-term forecasts for passive substations, while confirming the expected loss of accuracy as the available historical information becomes more distant from the target day. The relatively gradual degradation from D-1 to D-5 suggests that the recurring daily and weekly structure of passive demand remains informative even at longer forecasting horizons.

3.3.2. Hybrid Substations

Hybrid substations are characterized by significant contributions from both consumption (A1) and local generation (A2), which are therefore forecast separately.

Figure 7 shows the validation results for SS12 over five representative days. For A1, the five horizon-specific models reproduce the main load pattern, with the largest errors occurring around daily peaks and sharp variations. For A2, the models correctly capture the photovoltaic generation windows but show larger discrepancies in peak magnitude. In particular, the measured peaks vary considerably from day to day, while the forecasts tend to reproduce a more regular generation pattern.

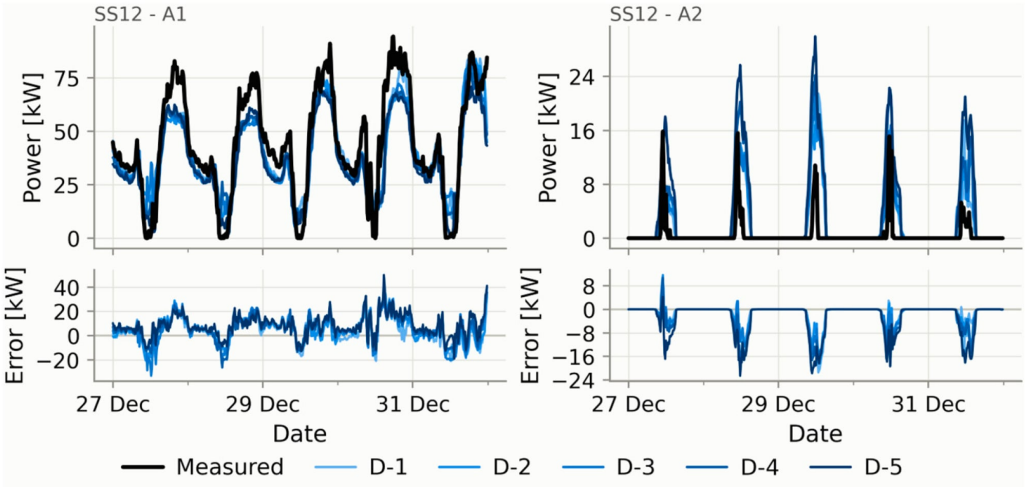


Figure 7. Validation for the hybrid substation SS12: measured load (A1 and A2) and the five model predictions (top) and the corresponding errors (bottom).

Table 4 extends the analysis to both hybrid substations and reports MAE, MAPE, WAPE, and R^2 for A1 and A2 across the five forecasting horizons.

Table 4. Test set MAE, MAPE, WAPE, R2 of the 5 forecasting models (D-1 to D-5) for hybrid substations.

	Target	Metric	D-1	D-2	D-3	D-4	D-5
SS12	A1	MAE [kW]	6.631	7.308	7.717	8.080	8.970
		MAPE [%]	18.70	19.98	20.42	21.23	22.98
		WAPE [%]	18.65	20.55	21.70	22.72	25.22
		R^2 [-]	0.772	0.739	0.723	0.690	0.653
	A2	MAE [kW]	2.081	2.243	2.403	2.770	3.089
		MAPE [%]	74.59	80.92	83.20	83.53	101.72
		WAPE [%]	117.67	126.85	135.88	156.59	174.68
		R^2 [-]	0.391	0.240	0.194	0.044	−0.095
SS13	A1	MAE [kW]	10.003	10.476	11.237	11.369	11.615
		MAPE [%]	21.07	22.46	23.24	24.09	24.82
		WAPE [%]	25.67	26.88	28.84	29.18	29.81
		R^2 [-]	0.693	0.653	0.602	0.591	0.580
	A2	MAE [kW]	1.296	1.376	1.502	1.625	1.683
		MAPE [%]	54.56	55.15	57.79	64.17	64.95
		WAPE [%]	57.16	60.70	66.23	71.66	74.23
		R^2 [-]	0.709	0.678	0.621	0.559	0.546

For A1, the D-1 MAE is 6.63 kW for SS12 and 10.00 kW for SS13, with corresponding MAPE values of 18.70% and 21.07%. Accuracy progressively decreases toward D-5, where the MAE reaches 8.97 kW and 11.62 kW, respectively. Compared with passive substations, the withdrawal forecast therefore exhibits higher relative errors, while maintaining a similar gradual degradation with the forecasting horizon.

For A2, the absolute errors are considerably smaller, with D-1 MAE values of 2.08 kW for SS12 and 1.31 kW for SS13. However, the corresponding MAPE values are much higher, ranging from 54.56% to 74.59% at D-1 and reaching up to 101.72% at D-5.

Overall, the results show that the forecasting framework retains good capability in reproducing the temporal evolution of withdrawn power in hybrid substations, although with higher errors than in the passive case. The injected power forecast captures the occurrence and timing of the daily generation periods, but reproducing their amplitude is more challenging. The distinction between A1 and A2 therefore highlights the different forecasting behavior of the consumption and generation-related components within the same operating class.

3.3.3. Active Substations

Active substations represent the most challenging forecasting case due to the highly variable behavior of their injected power (A2).

Figure 8 shows the validation results for SS2 over five representative days. The five horizon-specific models capture the general evolution of A2 and the main periods of high and low generation but show larger discrepancies during rapid variations and transitions between operating levels. In particular, the forecasts tend to smooth the measured profile, failing to fully reproduce the magnitude and timing of short-term fluctuations. Consequently, the largest errors occur during the most abrupt changes in injected power.

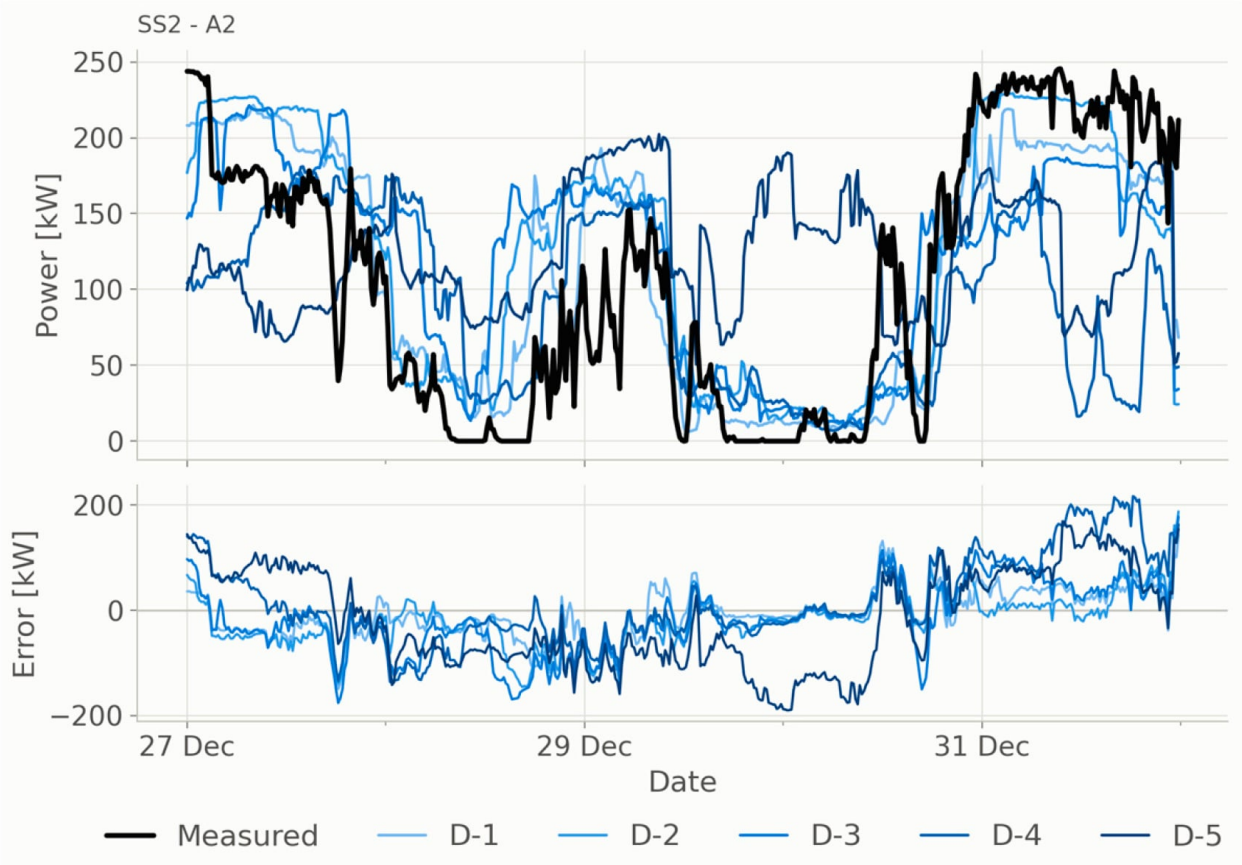


Figure 8. Validation for the active substation SS2: measured load (A2) and the five model predictions (top) and the corresponding errors (bottom).

Table 5 extends the analysis to all active substations and reports MAE, MAPE, WAPE, and R^2 across the five forecasting horizons.

Table 5. Test set MAE, MAPE, WAPE, R2 of the 5 forecasting models (D-1 to D-5) for active substations.

	Metric	D-1	D-2	D-3	D-4	D-5
SS2	MAE [kW]	37.41	39.48	47.25	49.64	74.81
	MAPE [%]	81.85	74.00	83.21	83.92	140.92
	WAPE [%]	67.73	71.47	85.54	89.86	135.44
	R^2 [-]	0.478	0.457	0.250	0.175	−0.650
SS4	MAE [kW]	27.74	28.56	35.04	38.03	43.88
	MAPE [%]	73.58	67.96	75.40	81.72	101.84
	WAPE [%]	68.73	70.76	86.81	94.22	108.71
	R^2 [-]	0.382	0.351	0.030	0.015	−0.196
SS5	MAE [kW]	12.05	12.56	14.75	17.39	17.31
	MAPE [%]	64.82	61.88	65.31	70.31	70.39
	WAPE [%]	90.77	94.61	111.16	131.04	130.46
	R^2 [-]	0.388	0.389	0.296	0.081	0.047
SS8	MAE [kW]	52.80	51.09	56.26	63.01	80.24
	MAPE [%]	76.51	70.61	75.92	78.18	92.28
	WAPE [%]	92.23	89.24	98.27	110.07	140.15
	R^2 [-]	0.318	0.410	0.322	0.143	−0.288
SS9	MAE [kW]	15.90	16.68	17.27	22.82	25.62
	MAPE [%]	76.57	71.44	68.20	91.29	101.25
	WAPE [%]	62.36	65.42	67.75	89.51	100.52
	R^2 [-]	0.453	0.497	0.467	0.062	−0.009
SS11	MAE [kW]	18.35	18.90	22.50	23.43	27.38
	MAPE [%]	73.07	70.60	75.29	77.74	83.29
	WAPE [%]	79.02	81.35	96.87	100.88	117.89
	R^2 [-]	0.400	0.412	0.236	0.140	−0.078
SS14	MAE [kW]	31.42	30.55	35.54	43.07	42.14
	MAPE [%]	80.92	70.38	74.67	102.95	90.60
	WAPE [%]	86.23	83.82	97.52	118.18	115.64
	R^2 [-]	0.299	0.293	0.155	−0.129	−0.089
SS15	MAE [kW]	33.47	35.33	42.74	43.21	50.70
	MAPE [%]	71.20	76.79	88.23	87.53	97.60
	WAPE [%]	85.80	90.57	109.56	110.76	129.96
	R^2 [-]	0.332	0.240	0.048	0.028	−0.258
SS16	MAE [kW]	3.49	3.55	4.02	4.56	4.96
	MAPE [%]	68.61	62.74	71.16	81.15	91.56
	WAPE [%]	60.37	61.47	69.58	78.93	85.80
	R^2 [-]	0.646	0.677	0.575	0.479	0.412

At D-1, the MAE varies considerably across the class, from 3.49 kW for SS16 to 52.80 kW for SS8, while the MAPE ranges from 64.82% to 81.85%. Unlike the passive and hybrid

cases, the accuracy degrades markedly with the horizon (median WAPE from 79% at D-1 to 118% at D-5), following the decreasing accuracy of the weather forecasts on which these models rely.

Overall, the results indicate that injected power forecasting in active substations is substantially more challenging than load forecasting in passive substations. More importantly, the degradation from D-1 to D-5 follows the accuracy of the weather forecasts rather than the loss of recent measurements, and the dominant limitation is the difficulty of reproducing the short-term variability of the generation profile from the information available to the forecasting model. This is particularly relevant for the active substations considered here, whose injected power is predominantly associated with wind generation. Accurate wind power forecasting generally requires dedicated forecasting approaches combining meteorological information with plant-specific characteristics, such as wind conditions at hub height, turbine technology and power curves, and other operational data [28]. The absence of this information limits the ability of the present general-purpose framework to reproduce the highly variable generation profiles observed in this class.

Finally, Table 6 compares the RF with the daily naive benchmark. The RF outperforms the benchmark in 85 of the 95 cases. RF skill is about 20% for passive substations up to D-4 and decreases to 7% at D-5, where the naive is slightly more accurate only for SS10; for active substations it drops from 33% at D-1 to 2% at D-5 as the weather forecasts lose accuracy. The naive is also more accurate for the photovoltaic injection of SS12 at all horizons and SS16 at D-1 and D-5, whose low winter output is well predicted by the previous day.

Table 6. Median test-set WAPE [%] of the RF and of the daily naive benchmark, and median skill [%] of the RF over the benchmark. For the hybrid class, the values of SS12 and SS13 are given in parentheses.

Class	Target	Forecast	D-1	D-2	D-3	D-4	D-5
Passive	A1	RF	9.2	9.8	10.1	10.9	12.9
		Daily naive	11.9	14.0	13.9	14.1	14.5
		Skill [%]	21.0	24.2	22.7	18.0	7.4
Hybrid	A1	RF	22.2 (18.6; 25.7)	23.7 (20.5; 26.9)	25.3 (21.7; 28.8)	25.9 (22.7; 29.2)	27.5 (25.2; 29.8)
		Daily naive	31.3 (21.1; 41.4)	41.2 (24.2; 58.2)	41.3 (24.5; 58.1)	41.7 (25.6; 57.9)	43.1 (27.5; 58.8)
		Skill [%]	24.9 (11.8; 38.0)	34.4 (15.1; 53.8)	30.9 (11.4; 50.4)	30.3 (11.1; 49.6)	28.8 (8.2; 49.3)
Hybrid	A2	RF	87.4 (117.7; 57.2)	93.8 (126.8; 60.7)	101.1 (135.9; 66.2)	114.1 (156.6; 71.7)	124.5 (174.7; 74.2)
		Daily naive	93.0 (84.8; 101.1)	120.8 (99.1; 142.5)	127.1 (105.9; 148.2)	123.2 (102.2; 144.2)	120.1 (101.2; 139.1)
		Skill [%]	2.4 (-38.8; 43.5)	14.7 (−28.0; 57.4)	13.5 (−28.3; 55.3)	−1.4 (−53.2; 50.3)	−13.0 (−72.6; 46.6)
Active	A2	RF	79.0	81.4	96.9	100.9	117.9
		Daily naive	115.2	127.6	128.7	123.7	123.2
		Skill [%]	33.5	39.0	27.7	18.0	1.7

Moreover, the rolling-origin evaluation, shown in Table 7, confirms the ranking of the classes in every window (median WAPE of 8.3%–15.8% for passive and 49%–74% for active substations), and the RF outperforms the daily naive benchmark in 103 of the 114 cases; most exceptions occur in July and August, when the training period does not yet include summer conditions.

Table 7. Rolling-origin evaluation (D-1): median WAPE [%] of the RF and of the daily naive benchmark in monthly test windows.

Class	Target	Forecast	Jul	Aug	Sep	Oct	Nov	Dec	All
Passive	A1	RF	15.8	12.4	9.2	8.3	8.7	9.4	10.0
		Daily naive	14.6	14.2	12.9	11.9	12.1	12.0	12.5
Hybrid	A1	RF	27.1	37.1	23.0	22.3	20.4	23.8	23.4
		Daily naive	35.1	27.3	32.1	34.0	32.7	30.9	29.4
Hybrid	A2	RF	69.7	58.6	41.7	57.3	66.4	139.7	57.3
		Daily naive	93.8	58.6	78.1	102.4	99.7	97.9	84.1
Active	A2	RF	56.9	64.6	52.2	48.9	74.2	68.2	62.1
		Daily naive	82.8	82.9	89.8	90.7	126.6	120.0	93.0

3.4. Forecasting Drivers

The differences in forecasting performance across the operating classes can be interpreted by examining the information exploited by the trained models. To this purpose, the RF feature importance is grouped into three families: autoregressive, weather, and calendar variables. Figure 9 reports the relative contribution of each family in the D-1 models for the relevant combinations of operating class and power-flow direction.

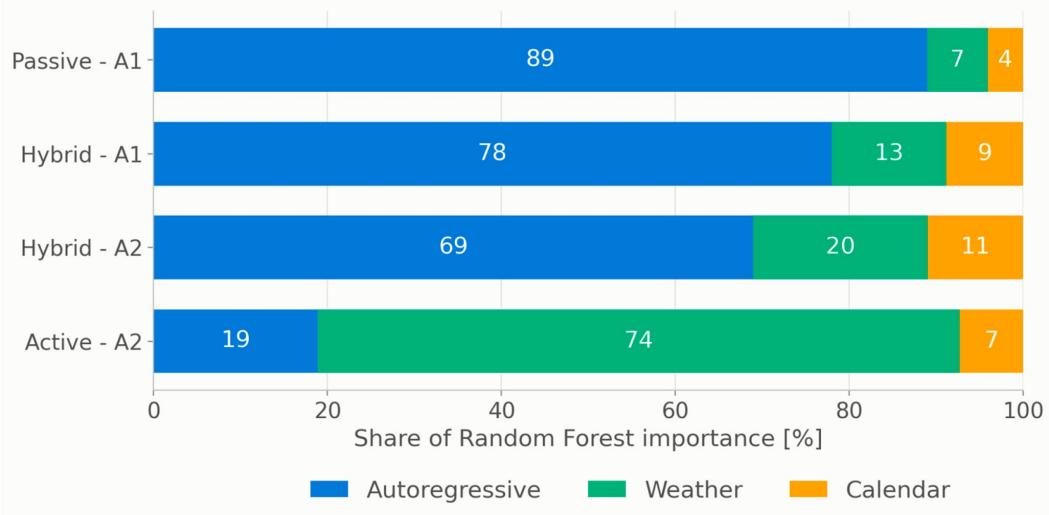


Figure 9. Distribution of RF feature importance across the analyzed operating classes and power-flow directions (D-1 models).

A clear distinction emerges between consumption and generation forecasting. For A1, autoregressive features dominate in both passive and hybrid substations, accounting for

89% and 78% of the total importance, respectively. This is consistent with the regular temporal structure of consumption observed in these classes. For hybrid A2, the autoregressive contribution decreases to 69%, while meteorological variables increase to 20%, reflecting the greater influence of weather conditions on the predominantly photovoltaic generation. A markedly different distribution is observed for active substations. For A2, meteorological variables become dominant, accounting for 74% of the total importance, while the autoregressive contribution drops to 19%. This is consistent with the predominantly wind-driven generation of this class, whose evolution is more strongly influenced by meteorological conditions than by past power measurements.

These results also help explain the differences observed across forecasting horizons. The strong autoregressive contribution in passive and hybrid substations is consistent with the progressive loss of accuracy from D-1 to D-5, as recent historical information becomes less accessible at longer horizons. Conversely, active SSs rely primarily on meteorological information, so that their accuracy across horizons follows that of the weather forecasts, which decreases with the lead time.

3.5. Hierarchical Reconciliation

The hierarchical reconciliation procedure is illustrated using SS1 as a representative example. The results are analyzed separately for the three operating classes, as their effectiveness depends on the composition of the underlying hierarchy. Figure 10 shows the distribution of connection points across the three power clusters for each SS. Passive substations generally include a large number of connection points distributed across multiple clusters, allowing individual behavior to be averaged within each aggregate. For instance, SS1 includes 139 points below 3.3 kW and 76 between 3.3 and 6.6 kW, compared with only 3 above 6.6 kW. In contrast, active substations concentrate their injection in the largest power cluster, typically comprising only 1 to 6 connection points. Their hierarchy therefore consists of a few large generators, providing little aggregation compared with passive substations.

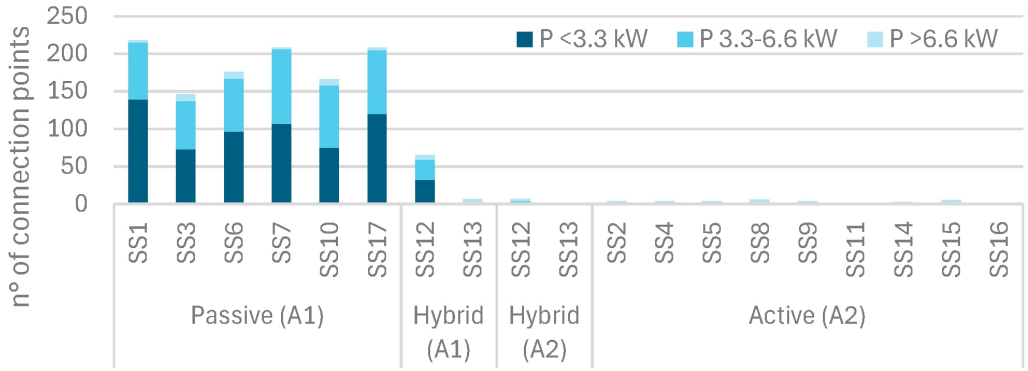


Figure 10. Distribution of connection points among the three power clusters (<3.3 kW, 3.3–6.6 kW, and >6.6 kW) for each substation and modeled power-flow direction, grouped by operating class.

Figure 11 shows the four lower-level nodes of the hierarchy for SS1 over a representative five-day period. Their forecasting performance varies considerably across nodes. The two lower power clusters, which aggregate a larger number of connection points, achieve R^2 values of 0.84 and 0.85, whereas the largest-power cluster, composed of fewer and more irregular connection points, reaches only 0.40. The error index also exhibits a recurring daily pattern and is forecasted with an R^2 of 0.64. These differences are relevant for reconciliation, as the effectiveness of the MinT projection depends on the predictive information provided by the lower-level forecasts.

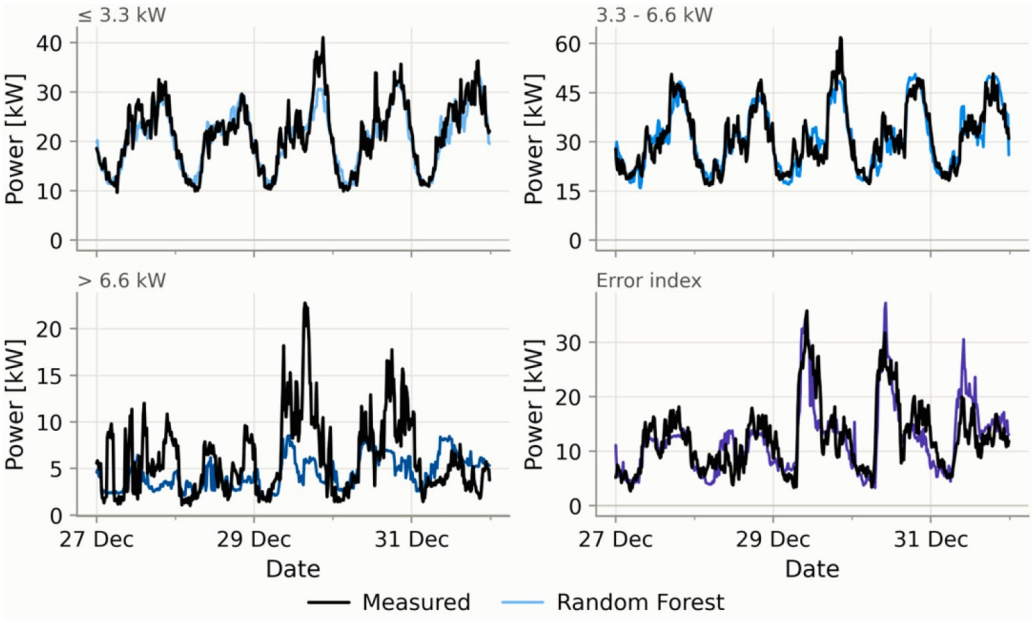


Figure 11. Forecasting performance of the lower-level hierarchy for SS1 over the selected validation period.

3.5.1. Passive Substations

Passive substations provide favorable conditions for hierarchical reconciliation, as their connection point aggregates retain the regular temporal structure of the substation load. Figure 12 compares the base (RF) and reconciled (RF+MinT) day-ahead forecasts for SS1. Both closely reproduce the measured load profile, with the largest deviations occurring around the sharpest peaks. Reconciliation mainly reduces local forecast errors rather than altering the overall profile, with the RF+MinT errors generally confined to a narrower range and fewer pronounced deviations. MinT therefore provides a targeted correction of the base forecast while preserving its overall temporal evolution.

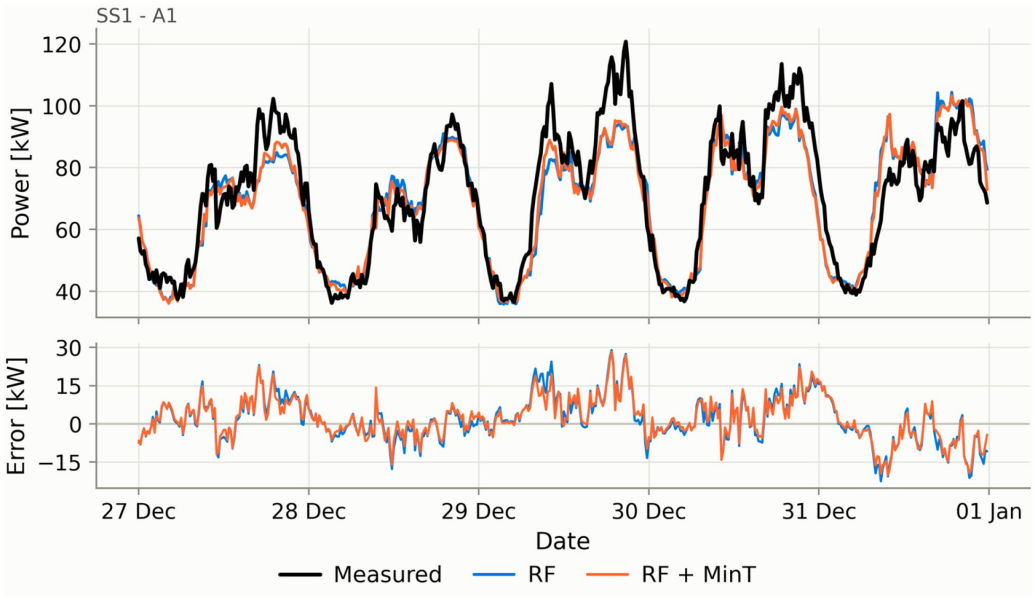


Figure 12. Comparison of base and MinT day-ahead forecasts for passive substation SS1.

Table 8 extends the comparison to all passive substations. Reconciliation reduces the MAE on all the substations, with a median change of −2.8% and a largest change of −4.6%,

and it improves the MAPE on the same cases, with a median change of −0.25 percentage points. The improvement is consistent across the class and both metrics, although moderate in magnitude, indicating that reconciliation primarily corrects local deviations while preserving the overall forecast profile.

Table 8. Day-ahead forecasting performance before and after MinT for passive substations.

	MAE RF [kW]	MAE RF+MinT [kW]	ΔMAE [%]	MAPE RF [%]	MAPE RF+MinT [%]	ΔMAPE [p.p.]	R^2 RF	R^2 RF+MinT	ΔR^2
SS1	5.829	5.699	−2.2	8.71	8.55	−0.15	0.868	0.875	+0.007
SS3	4.365	4.183	−4.2	14.50	13.96	−0.54	0.763	0.784	+0.021
SS6	6.596	6.294	−4.6	8.97	8.52	−0.45	0.807	0.828	+0.021
SS7	8.136	7.904	−2.9	7.04	6.79	−0.24	0.900	0.904	+0.004
SS10	6.915	6.721	−2.8	11.68	11.42	−0.25	0.821	0.835	+0.013
SS17	4.661	4.537	−2.6	8.41	8.16	−0.25	0.877	0.883	+0.006

3.5.2. Hybrid Substations

In hybrid substations, reconciliation is performed separately for the two power-flow directions, allowing its impact on A1 and A2 forecasting to be assessed independently. Figure 13 compares the base and reconciled forecasts of SS12 for both power-flow directions. For A1, reconciliation introduces moderate corrections throughout the profile, with the largest differences occurring around daily peaks and rapid load variations. For A2, the effect is more localized but also more pronounced, with MinT mainly modifying the magnitude of the daily generation peaks. These corrections do not systematically bring the forecast closer to the measured profile, indicating that the effectiveness of reconciliation differs between the two power-flow directions.

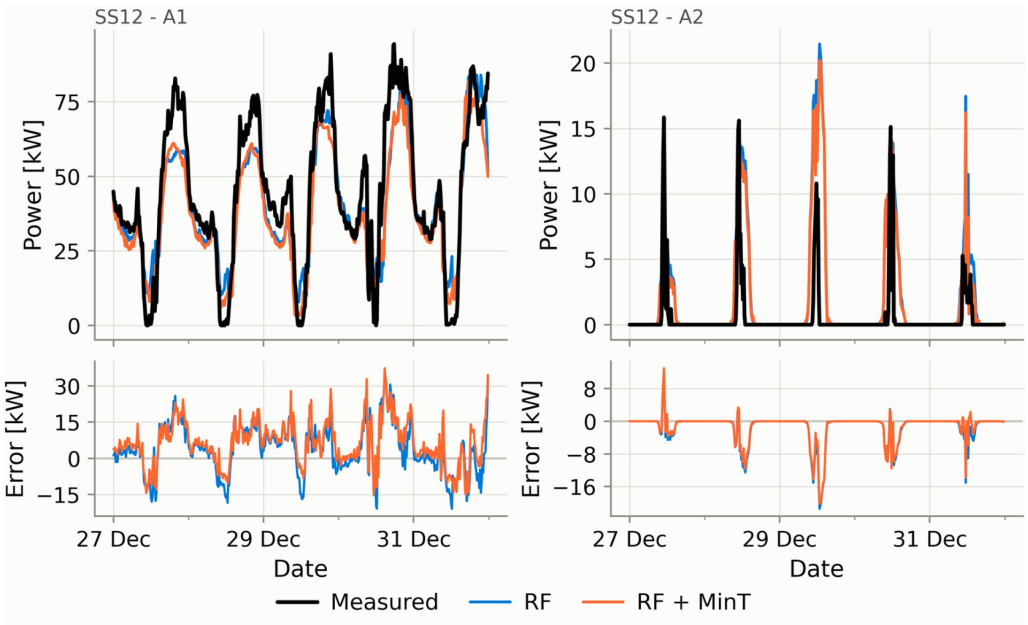


Figure 13. Hybrid substation SS12: comparison of day-ahead base and MinT-reconciled forecasts for withdrawn power (A1) and injected power (A2), with corresponding errors.

Table 9 reports the reconciliation results for both targets of the two hybrid substations. For A1, the MAE changes by −1.9% for SS12 and +4.7% for SS13; for A2, it increases by 19.5% and 14.2%. Reconciliation therefore provides no benefit for hybrid substations and degrades their injection forecast.

Table 9. Day-ahead forecasting performance before and after MinT reconciliation for hybrid substations, reported separately for withdrawn (A1) and injected (A2) power.

	Target	MAE RF [kW]	MAE RF+MinT [kW]	ΔMAE [%]	MAPE RF [%]	MAPE RF+MinT [%]	ΔMAPE [p.p.]	R^2 RF	R^2 RF+MinT	ΔR^2
SS12	A1	6.631	6.506	−1.9	18.70	18.31	−0.39	0.772	0.783	+0.011
	A2	2.081	2.486	+19.5	74.59	99.45	+24.85	0.391	0.174	−0.217
SS13	A1	10.003	10.472	+4.7	21.07	21.86	+0.79	0.693	0.675	−0.018
	A2	1.296	1.480	+14.2	54.56	55.31	+0.74	0.709	0.706	−0.003

3.5.3. Active Substations

For active substations the analysis is restricted to the injected power. Figure 14 shows that the reconciled forecast of SS2 departs from the base forecast in some periods, remaining above the measured injection on 29–30 December and below it on 31 December. Since the hierarchy consists of the error index and a single cluster of generators, which carries almost the same information as the substation meter, MinT cannot add predictive information.

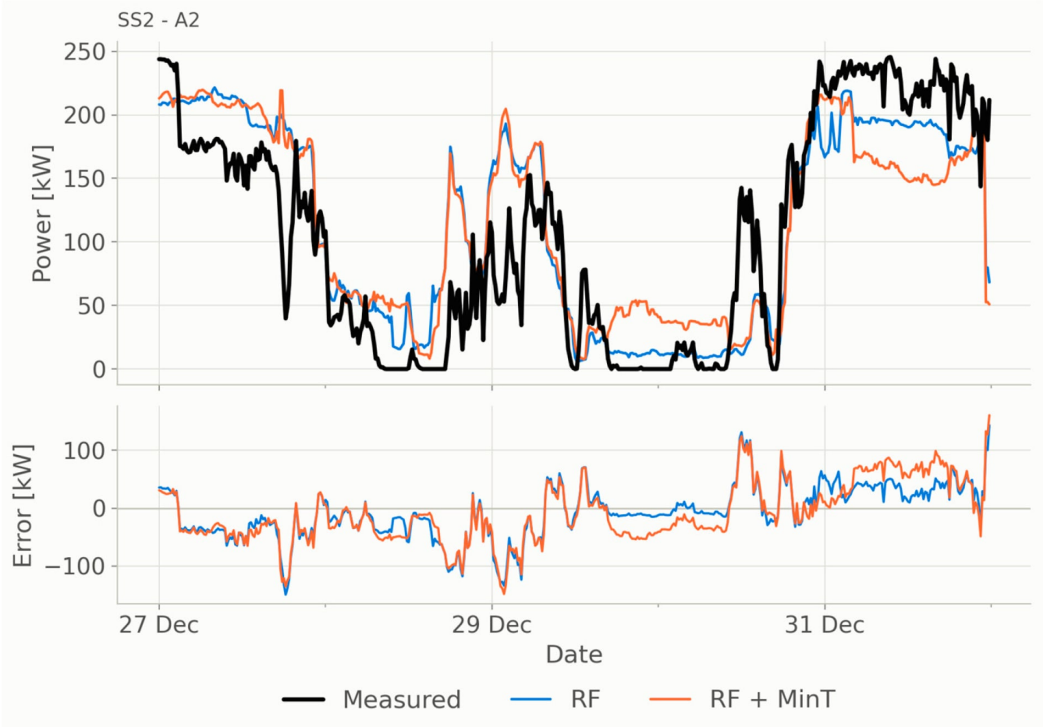


Figure 14. Active substation SS2: comparison of the day-ahead base Random-Forest forecast and the MinT-reconciled forecast for injected power (A2), with corresponding errors.

Table 10 confirms this pattern across the class, with reconciliation reducing MAE for only four of the nine substations and yielding a median increase of 0.1% across all nine. The changes remain within ±2% for six of the eight wind-dominated substations, but the MAE increases by 6.3% and 7.4% for SS2 and SS4 and by 3.7% for SS16. In SS2 and SS4, the error index showed a large noisy component (standard deviation of 12–40 kW) in the first

half of the year, which disappeared from June to July onwards. Enforcing coherence on this class therefore does not improve the substation forecast.

Table 10. Day-ahead forecasting performance before and after MinT reconciliation for active substations.

	MAE RF [kW]	MAE RF+MinT [kW]	ΔMAE [%]	MAPE RF [%]	MAPE RF+MinT [%]	ΔMAPE [p.p.]	R^2 RF	R^2 RF+MinT	ΔR^2
SS2	37.411	39.780	+6.3	81.85	80.73	−1.12	0.478	0.461	−0.017
SS4	27.743	29.794	+7.4	73.58	72.92	−0.66	0.382	0.351	−0.031
SS5	12.047	11.995	−0.4	64.82	64.55	−0.26	0.388	0.397	+0.009
SS8	52.803	52.786	−0.0	76.51	76.59	+0.08	0.318	0.319	+0.000
SS9	15.896	15.918	+0.1	76.57	76.70	+0.13	0.453	0.450	−0.003
SS11	18.353	18.480	+0.7	73.07	73.33	+0.26	0.400	0.397	−0.003
SS14	31.421	31.024	−1.3	80.92	79.01	−1.90	0.299	0.315	+0.016
SS15	33.472	32.828	−1.9	71.20	71.22	+0.02	0.332	0.344	+0.012
SS16	3.489	3.619	+3.7	68.61	67.83	−0.79	0.646	0.661	+0.014

Table 11 extends the comparison to all forecasting horizons using a separate model and covariance matrix for each horizon. MinT reduces MAE in 27 of the 30 passive cases, with median reductions ranging from 2.4% to 4.0% across horizons. Conversely, it increases MAE for photovoltaic injections at every horizon, while its effect on the other active substations remains limited, except for SS2 and SS4.

Table 11. Median change [%] of the test-set MAE with MinT reconciliation and number of substations improved, for horizons D-1 to D-5.

Class	Target	D-1	D-2	D-3	D-4	D-5
Passive	A1	−2.8 (6/6)	−2.4 (5/6)	−2.6 (6/6)	−3.8 (6/6)	−4.0 (4/6)
Hybrid	A1	+1.4 (1/2)	+0.6 (1/2)	−0.2 (1/2)	+1.9 (1/2)	+1.3 (1/2)
Hybrid	A2	+16.8 (0/2)	+7.3 (0/2)	+5.6 (0/2)	+6.7 (0/2)	+10.8 (0/2)
Active	A2	+0.1 (4/9)	+0.5 (3/9)	+0.4 (4/9)	+0.3 (3/9)	−0.1 (5/9)

Reconciliation also entails an additional computational cost reported in Table 12. The hierarchy requires three to five RF models instead of one, so that training takes on average 160 s instead of 37 s per substation, direction, and horizon, while prediction and the MinT projection (about 9 ms) are negligible. The hierarchical forecast therefore costs about 4.4 times the base forecast.

Overall, the results show that the main benefit of hierarchical reconciliation is the enforcement of coherence across aggregation levels, while improvements in forecasting accuracy depend on the quality of the lower-level forecasts. For passive substations, where connection-point aggregates are based on a large number of customers and retain regular temporal patterns, MinT provides a consistent, although moderate, improvement across the entire class. The benefit becomes less systematic for hybrid substations and is negligible for active substations, where the lower-level nodes consist of fewer and more variable generators. These results indicate that the effectiveness of reconciliation is closely related to the predictability of the underlying hierarchy. MinT therefore guarantees coherent forecasts

in all cases, while its contribution to forecasting accuracy is greatest when the lower-level aggregates provide reliable predictive information.

Table 12. Computational cost: RF models per hierarchy and mean wall-clock time per substation, direction, and horizon.

Class	Target	Models [-]	Training Base [s]	Training Hier. [s]	Prediction Base [s]	Prediction Hier. [s]	MinT [ms]	Cost Ratio [-]
Passive	A1	5.0	38.4	226.3	0.26	1.47	14.1	5.9
Hybrid	A1	4.5	36.6	175.7	0.23	1.00	7.9	4.8
Hybrid	A2	4.0	24.6	111.8	0.21	0.86	6.7	4.5
Active	A2	3.1	38.1	123.0	0.25	0.77	6.4	3.2
All		3.9	36.6	160.0	0.25	1.02	9.0	4.4

4. Conclusions

This paper presented a data-driven framework for short-term forecasting at the SS level, combining the characterization of substation operating conditions, horizon-specific RF models, and hierarchical reconciliation through the MinT method. The methodology was assessed on one year (2025) of quarter-hourly measurements from 17 secondary substations and their underlying connection points, considering withdrawn (A1) and injected (A2) power separately.

The results show that the operating characteristics of a substation strongly affect its forecasting performance. Passive substations exhibit regular and highly autocorrelated consumption profiles, resulting in accurate forecasts and a gradual loss of performance from D-1 to D-5. Hybrid substations retain similar characteristics for withdrawn power, while their predominantly photovoltaic injection introduces greater variability, particularly in the magnitude of generation peaks. Active substations represent the most challenging case. Their predominantly wind-driven injection exhibits weaker temporal persistence and substantially higher forecasting errors, which grow with the horizon as the weather forecasts lose accuracy. This indicates that increasing the availability of recent historical measurements alone is insufficient to improve forecasts for highly variable generation profiles. For this class, the framework does not achieve sufficient accuracy for operational use, highlighting the need for DSOs to collect more extensive and detailed data to better characterize generation patterns and improve forecasting performance.

The feature-importance analysis supports this interpretation. Autoregressive information dominates withdrawn-power forecasting for passive and hybrid substations, accounting for 89% and 78% of the total importance, respectively, and remains relevant for photovoltaic-dominated hybrid injection at 69%. Conversely, meteorological variables account for 74% of the importance for active-substation injection, while the autoregressive contribution decreases to 19%. These results show that the most relevant forecasting information varies with the operating characteristics of the substation, supporting the use of differentiated forecasting approaches across the analyzed classes.

Hierarchical reconciliation exhibits a similar dependence on the characteristics of the underlying hierarchy. MinT guarantees coherence between substation and lower-level forecasts in all cases, but its contribution to forecasting accuracy depends on the predictability of the connection point aggregates. For passive substations, where clusters aggregate many connection points with regular consumption patterns, reconciliation consistently improves the day-ahead forecast, reducing MAE for all six substations with a median improvement of 2.8%. Its benefit becomes less systematic for hybrid substations and negligible for active

substations, whose lower-level hierarchy consists of fewer and more variable generation units. Thus, while coherence is ensured independently of forecasting accuracy, the accuracy benefit of reconciliation is greatest when reliable information is available at the lower levels of the hierarchy.

The study has some limitations. The dataset covers a single DSO and one year, which limits the generalizability of the findings. Forecasting performance at generation-dominated substations is strongly affected by errors in the general-purpose weather forecasts used as inputs, which refer to a height of 10 m. The lack of plant-specific information on wind and photovoltaic generation further constrains the representation of local generation. In addition, MinT reconciliation increases computational cost by approximately a factor of four.

Overall, the study shows that forecasting performance at the SS level cannot be considered independently of the operating characteristics of the underlying network. The proposed classification provides a simple means of identifying different forecasting conditions and interpreting the expected performance of both the base and reconciled forecasts. From a DSO perspective, this supports a differentiated forecasting strategy in which conventional autoregressive approaches remain effective for consumption-dominated substations, while generation-dominated substations require more specialized models and information. Future work should therefore focus on generation-aware forecasting approaches incorporating more detailed meteorological forecasts and plant-specific information, particularly for wind-dominated substations, on the comparison with other learning models, such as gradient boosting and recurrent neural networks, on probabilistic forecasts, together with scalable reconciliation strategies suitable for deployment across large substation populations.

Author Contributions: Conceptualization, D.A., M.S., G.R. and M.M.; methodology, D.A., M.S., G.R. and M.M.; software, D.A. and M.S.; validation, D.A., M.S., G.R. and M.M.; formal analysis, M.S. and M.M.; investigation, D.A. and M.S.; resources, G.R. and M.M.; data curation, D.A. and M.S.; writing—original draft preparation, D.A.; writing—review and editing, M.S., G.R. and M.M.; visualization, D.A. and M.S.; supervision, G.R. and M.M.; project administration, G.R. and M.M.; funding acquisition, G.R. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the G.E.M.I.N.I. project, funded by the Italian Ministry of Environment and Energy Security (MASE) under Mission Innovation 2.0–Green Powered Future Mission (GPFM) (MI_DDR_00364 - CUP F53C25001800001), pursuant to Ministerial Decree No. 386 of 17 November 2023.

Data Availability Statement: The datasets presented in this article are not publicly available and were provided to the authors under confidentiality restrictions that prohibit their disclosure.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. EU Market Outlook for Solar Power 2024–2028—SolarPower Europe. Available online: https://www.solarpowereurope.org/insights/outlooks/eu-market-outlook-for-solar-power-2024-2028/detail (accessed on 13 August 2026).
2. Number of EVs, Air Conditioners, and Heat Pumps in Operation in the Stated Policies Scenario, 2023 and 2035—Charts—Data & Statistics—IEA. Available online: https://www.iea.org/data-and-statistics/charts/number-of-evs-air-conditioners-and-heat-pumps-in-operation-in-the-stated-policies-scenario-2023-and-2035 (accessed on 13 August 2026).
3. Kotsonias, A.; Hadjidemetriou, L.; Asprou, M.; Panayiotou, C.G. Operational challenges and solution approaches for low voltage distribution grids—A review. *Electr. Power Syst. Res.* **2025**, *239*, 111258. [CrossRef]
4. Scrocca, A.; Daccò, E.; Andreotti, D.; Rancilio, G.; Bovera, F.; Falabretti, D.; Delfanti, M. Local flexibility markets in Europe: A critical review of market designs, operational maturity and stakeholder perspectives. *Renew. Sustain. Energy Rev.* **2026**, *226*, 116434. [CrossRef]
5. Wang, Y.; Zhang, N.; Chen, Q.; Kirschen, D.S.; Li, P.; Xia, Q. Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV. *IEEE Trans. Power Syst.* **2018**, *33*, 3255–3264. [CrossRef]

6. Dorostkar-Ghamsari, M.R.; Fotuhi-Firuzabad, M.; Lehtonen, M.; Safdarian, A.; Hoshyarzade, A.S. Stochastic Operation Framework for Distribution Networks Hosting High Wind Penetrations. *IEEE Trans. Sustain. Energy* **2019**, *10*, 344–354. [CrossRef]
7. Lusis, P.; Khalilpour, K.R.; Andrew, L.; Liebman, A. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Appl. Energy* **2017**, *205*, 654–669. [CrossRef]
8. Khan, Z.A.; Ullah, A.; Haq, I.U.; Hamdy, M.; Mauro, G.M.; Muhammad, K.; Hijji, M.; Baik, S.W. Efficient Short-Term Electricity Load Forecasting for Effective Energy Management. *Sustain. Energy Technol. Assess.* **2022**, *53*, 102337. [CrossRef]
9. Sun, X.; Luh, P.B.; Cheung, K.W.; Guan, W.; Michel, L.D.; Venkata, S.S.; Miller, M.T. An Efficient Approach to Short-Term Load Forecasting at the Distribution Level. *IEEE Trans. Power Syst.* **2016**, *31*, 2526–2537. [CrossRef]
10. Rouwhorst, G.; Duque, E.M.S.; Nguyen, P.H.; Slootweg, H. Improving Clustering-Based Forecasting of Aggregated Distribution Transformer Loadings with Gradient Boosting and Feature Selection. *IEEE Access* **2022**, *10*, 443–455. [CrossRef]
11. Ding, N.; Benoit, C.; Foggia, G.; Besanger, Y.; Wurtz, F. Neural network-based model design for short-term load forecast in distribution systems. *IEEE Trans. Power Syst.* **2016**, *31*, 72–81. [CrossRef]
12. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [CrossRef]
13. Grabner, M.; Wang, Y.; Wen, Q.; Blazic, B.; Struc, V. A Global Modeling Framework for Load Forecasting in Distribution Networks. *IEEE Trans. Smart Grid* **2023**, *14*, 4927–4941. [CrossRef]
14. Faustine, A.; Nunes, N.J.; Pereira, L. Efficiency Through Simplicity: MLP-Based Approach for Net-Load Forecasting with Uncertainty Estimates in Low-Voltage Distribution Networks. *IEEE Trans. Power Syst.* **2025**, *40*, 46–56. [CrossRef]
15. Caballero-Peña, J.; Cadena-Zarate, C.; Parrado-Duque, A.; Osma-Pinto, G. Distributed energy resources on distribution networks: A systematic review of modelling, simulation, metrics, and impacts. *Int. J. Electr. Power Energy Syst.* **2022**, *138*, 107900. [CrossRef]
16. Wilms, H.; Cupelli, M.; Monti, A. On the Necessity of Exogenous Variables for Load, PV and Wind Day-Ahead Forecasts using Recurrent Neural Networks. In Proceedings of the 2018 IEEE Electrical Power and Energy Conference, EPEC 2018, Toronto, ON, Canada, 10–11 October 2018. [CrossRef]
17. Hayes, B.P.; Gruber, J.K.; Prodanovic, M. Multi-nodal short-term energy forecasting using smart meter data. In *IET Generation, Transmission and Distribution*; John Wiley & Sons: Hoboken, NJ, USA, 2018; Volume 12, pp. 2988–2994. [CrossRef]
18. Shering, T.; Alonso, E.; Apostolopoulou, D. Investigation of Load, Solar and Wind Generation as Target Variables in LSTM Time Series Forecasting, Using Exogenous Weather Variables. *Energies* **2024**, *17*, 1827. [CrossRef]
19. Wang, J.; Zhong, H.; Lai, X.; Xia, Q.; Wang, Y.; Kang, C. Exploring key weather factors from analytical modeling toward improved solar power forecasting. *IEEE Trans. Smart Grid* **2019**, *10*, 1417–1427. [CrossRef]
20. Pinheiro, M.G.; Madeira, S.C.; Francisco, A.P. Short-term electricity load forecasting—A systematic approach from system level to secondary substations. *Appl. Energy* **2023**, *332*, 120493. [CrossRef]
21. Nespoli, L.; Medici, V.; Lopatichki, K.; Sossan, F. Hierarchical demand forecasting benchmark for the distribution grid. *Electr. Power Syst. Res.* **2020**, *189*, 106755. [CrossRef]
22. Wickramasuriya, S.L.; Athanasopoulos, G.; Hyndman, R.J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *J. Am. Stat. Assoc.* **2019**, *114*, 804–819. [CrossRef]
23. Mogos, A.S.; Ansari, O.A.; Liang, X.; Chung, C.Y. Hierarchical Load Forecast Aggregation for Distribution Transformers Using Minimum Trace Optimal Reconciliation and AMI Data. *IEEE Access* **2023**, *11*, 93472–93486. [CrossRef]
24. Bao, Y.; Xiong, T.; Hu, Z. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing* **2014**, *129*, 482–493. [CrossRef]
25. Free Open-Source Weather API|Open-Meteo.com. Available online: https://open-meteo.com/ (accessed on 14 August 2026).
26. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
27. Cerqueira, V.; Torgo, L.; Mozetič, I.; Soldatova, L.; Vanschoren, J.; Mozetič Igormozetic, I. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Mach. Learn.* **2020**, *109*, 1997–2028. [CrossRef]
28. Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. *Energies* **2020**, *13*, 3764. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.