RESEARCH ARTICLE

Hierarchical Load Forecast Aggregation for Distribution Transformers Using Minimum Trace Optimal Reconciliation and AMI Data

AMAN SAMSON MOGOS [1], (Member, IEEE), **OSAMA ASLAM ANSARI** [2], (Member, IEEE), **XIAODONG LIANG** [1], (Senior Member, IEEE), AND **C. Y. CHUNG** [3], (Fellow, IEEE)

[1]Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada
[2]Amazon, Vancouver, BC V6B 0M3, Canada
[3]Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, China

Corresponding author: Xiaodong Liang (xil659@mail.usask.ca)

ABSTRACT Overloading and load imbalance have a significant impact on the health of distribution transformers. The load of a distribution transformer can be considered in a hierarchical way: individual single-phase customers connected directly to the transformer (the bottom level), the load at each phase (the middle level), and the total load among three phases (the top level). Load at each hierarchical level can be predicted individually, known as "base forecast", through a state-of-the-art forecasting method, but this practice often leads to incoherency and bias, i.e., forecasts at a lower hierarchical level are not aggregated correctly to the forecast at a higher-hierarchical-level. In this paper, a novel load aggregation technique based on minimum trace (MinT)-based optimal reconciliation is proposed to improve the accuracy of prediction models. Base forecasts at each hierarchical level are firstly determined using independent autoregressive integrated moving average (ARIMA) models; MinT is then used to optimally reconcile base forecasts to ensure higher accuracy. The proposed method is validated by case studies. Advanced metering infrastructure (AMI) data recorded by Saskatoon Light and Power in Saskatoon; Canada is used as historical data in this study.

INDEX TERMS Distribution transformer, health monitoring, hierarchical load forecast aggregation, minimum trace optimal reconciliation.

NOMENCLATURE

$\hat{P}_T$	Base forecast.
$\tilde{P}_T$	Reconciled forecast.

ACRONYMS

ADMM	Alternating Direction Method of Multipliers.
AMI	Advanced Metering Infrastructure.
ARIMA	Auto-Regressive Integrated Moving Average.
BU	Bottom-Up.
CLC	Closed-Loop Clustering.
DERs	Distributed Energy Resources.
DGA	Dissolved Gas Analysis.
DP	Degree of Polymerization.
EVs	Electric Vehicles.
FA	Furanic Content Analysis.
GLS	Generalized Least Squares.
GTOP	Game-Theoretic Optimal Projection.
HI	Health Index.
HLF	Hierarchical Load Forecasting.
HST	Hot Spot Temperature.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MinT	Minimum Trace.
MinTSa	Minimum Trace Sample.
MinTSh	Minimum Trace Shrinkage.
MSE	Mean Squared Error.
OLS	Ordinary Least Squares.
OQA	Oil quality analysis.

RF	Random Forest.
RMA	Reduced Model Approach.
RMSE	Root Mean Square Error.
TD	Top Down.
VAR	Vector Autoregression.
WLS	Weighted Least Squares.
WLS-SS	WLS Structural Scaling.
WLS-VS	WLS Variance Scaling.
WNN	Wavelet Neural Network.

INDICES

b_t	n-vector representing bottom-level observations.
$\hat{b}_T(h)$	Bottom-level base forecast.
$B_T(h)$	The mean of the bottom-level base forecast.
d	Differencing order
$\hat{e}_T(h)$	Base forecast errors at h-step-ahead.
$\tilde{e}_t(h)$	Reconciled forecast errors at h-step-ahead
$\hat{\varepsilon}_h(h)$	Bottom-level base forecast error.
G	Reconciliation matrix in the order of $m \times n$.
k_h	Proportionality constant.
I_n	An identity matrix representing the lowest hierarchy level.
n	Bottom-level number of observations.
m	Total number of observations.
p	Order of autoregressive (AR) coefficients.
P_i	Power consumption true data.
$\hat{P}_i$	Power consumption predicted data.
P_t	m-vector that has all power consumption observations at the time t.
$\hat{P}_T(h)$	Base forecasts vector at h-step-ahead.
$\tilde{P}_T(h)$	Reconciled forecasts vector at h-step-ahead.
P_{t+h}	True values at $t + h$.
S	Summing matrix.
T	Duration of historical data
q	Order of moving average (MA) coefficients
V_h	Error ($\hat{\varepsilon}_h(h)$) variance-covariance matrix.
W_h	Base forecast variance-covariance matrix.
$\hat{W}_1$	Base forecast errors at $h = 1$
U	Unit Matrix
∂	Disaggregating proportion.
ξ_T	Historical data for time T.
Ω_i	Autoregressive (AR) coefficients.
Ψ_j	Moving average (MA) coefficients.

I. INTRODUCTION

In distribution systems, distribution transformers must be highly reliable to ensure continuous power supply to consumers [1] as transformer failures may result in huge financial losses to utilities and consumers [2]. Although transformers normally have a long lifespan (exceeding 50 years), the health of a distribution transformer may deteriorate due to oil leakage, overloading, unbalanced loading, and harmonics [3], and their failure rate increases due to demand increase and certain load types, for example, electric vehicle charging can dramatically lower a transformer's lifespan by 200% to 300%

[1], [4]. Transformer load forecasts are crucial for health monitoring and failure prevention caused by overloading and phase imbalance.

A distribution transformer's load can be considered in a hierarchical way: the amount of electricity consumed by each customer is aggregated to determine the load of each phase (Phase A, B or C); and the load at each phase is then aggregated to determine the total load among three phases of the transformer. Additional levels may be added similarly for a substation transformer. Load forecast at the top level determines the total load among three phases, which can assist in the evaluation of a transformer's health and potential failures due to overloading and winding overheating. Load forecast at the middle level determines the load at each phase, which assists in evaluating the phase imbalance of a transformer, and corrective actions may be required by distribution system operators. Load forecast at the bottom level determines the consumption of individual customers.

Hierarchical Load Forecasting (HLF) has been applied in the context of power generation to improve prediction accuracy through forecast reconciliation. Several HLF methods were proposed, including the game-theoretic optimal projection (GTOP)-based reconciliation [5], the least squares-based optimal reconciliation [6], and a vector autoregression framework (VAR)-based probabilistic forecasting [7]. These methods were tested on wind and solar power generation data with varying degrees of success. For example, a HLF method using the Alternating Direction Method of Multipliers (ADMM) for wind power is proposed in [8] to reconcile base forecasts and ensure disaggregated forecasts are aggregated correctly, resulting in improved prediction accuracy. A review of cutting-edge techniques and their pertinence to day-ahead wind power generation prediction is conducted in [5]. In [9], hourly generation data in Brazil is utilized to prove the accuracy of the aggregated generation prediction at the top hierarchy level using top-down (TD), bottom-up (BU), and optimal reconciliation methods. Optimal reconciliation as a post-forecast technique produces an unbiased and coherent reconciled forecast [10].

Among very limited HLF research in the literature, most existing HLF methods focus on the top hierarchy level. In [11], a closed-loop clustering (CLC) method for HLF is employed for substation load using smart meters and photovoltaic (PV) power generation data, showing a better performance than traditional BU and TD approaches, but it only focuses on the top hierarchy level. A multiplicative autoregressive integrated moving average (m-ARIMA) method in [12], the ensemble methods in [13], and a multi-scale expert aggregation method based on random forest (RF) for BU in [14] also only focus on the top-level hierarchy and ignore forecasts at the lower hierarchy, which are important for transformer health monitoring as well. A generic short-term load forecast is proposed in [15] using load pattern similarity-based child node classification, and wavelet neural networks (WNNs), and it can be used for distribution substations, feeders, and transformers. To improve HLF, a reduced model approach (RMA) is employed for electricity demand forecasting [16], along with smart-meter data with high resolution and dimension.

Very few research is reported on HLF in transformers. Although independent models have been used for base forecasts at different levels of a transformer's load, if not aggregated properly, this practice can lead to poor forecast accuracy. The literature review on HLF is presented in Fig. 1.

To overcome the limitations of HLF methods, a novel hierarchical load forecast aggregation method is proposed in this paper through the minimum trace sample (MinTSa) and minimum trace shrinkage (MinTSh)-based reconciliation algorithm, where power consumption forecasts at all levels of distribution and substation transformers can be reconciled and aggregated accurately. In this study, the base forecast at each hierarchical level is firstly conducted using an ARIMA model (in fact, any existing forecasting method can be employed for base forecasts); the base forecast is then optimally reconciled using MinT to ensure coherent forecasts with high accuracy. The proposed method incorporates information about the correlation structure of time-series data, resulting in a minimum reconciled forecast error and great computational efficiency in a large time-series dataset. The proposed method is validated through several case studies and shows superior performance compared to four benchmark methods.

The main contributions of this paper include.

- A novel transformer load forecast aggregation technique is proposed by a MinTSa and MinTSh-based post-forecast algorithm to optimally reconcile base forecasts obtained independently at each hierarchy level and to increase the accuracy of reconciled forecasts.
- The proposed method is independent of the base forecast method at each hierarchy level, hierarchical structure and forecast horizon.

The paper is organized as follows: in Section II, the loading impact on a transformer's health and the purpose of transformer load forecast aggregation are introduced; in Section III, the theory of HLF reconciliation is provided; in Section IV, the proposed method is explained; in Section V, simulation results of several case studies are presented; and the conclusion is drawn in Section VI.

II. LOADING IMPACT ON TRANSFORMER HEALTH AND THE PURPOSE OF LOAD FORECAST AGGREGATION

Loading history is one important pointer for the transformer health index (HI) along with dissolved gas analysis (DGA), oil quality analysis (OQA), furanic content analysis (FA), and degree of polymerization (DP) [17], [18]. Overloading and unbalanced loads are major contributors to distribution transformer failures and degradation [2], [19].

The load increase leads to the hotspot temperature (HST) increase, reducing the insulation life of a transformer. A transformer's loading and environmental conditions have considerable impacts on its aging process. Loading in particular is directly related to the current in the winding, overloading raises temperature in the winding and oil and shortens its service life [17], [20].

Unbalanced loads are responsible for unbalanced winding losses, which in turn impact the top-oil temperature and HST per phase [21]. In recent years, the introduction of electric vehicles (EVs), and increasing distributed energy resources (DERs) in distribution systems have exacerbated the
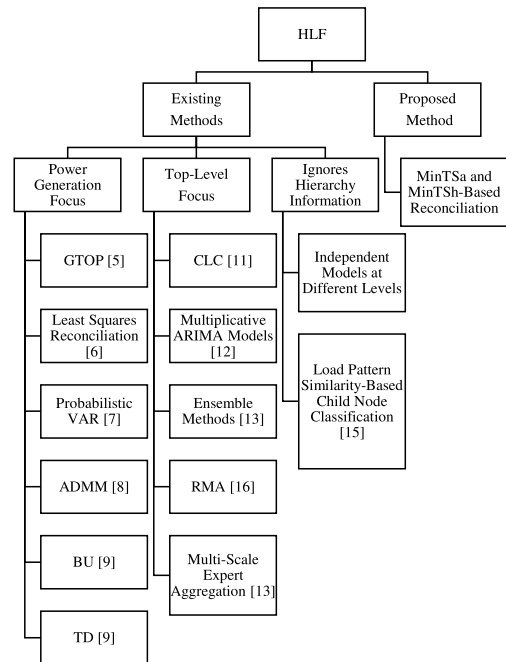

FIGURE 1. Hierarchical load forecasting.

unpredictability of load behaviours in distribution transformers [22]. Therefore, the load forecast of a distribution transformer is essential for its health monitoring. Loading behaviours of domestic consumers directly influence the reliability risk of distribution transformers [23], and thus, load forecast for individual customers is also critical.

Load forecasting is crucial in establishing the health state of a transformer, which allows utilities to proactively address potential problems. It may be used to determine future loadings of a transformer, an important aspect to calculate its health index, HI, which can be modelled as a multivariant regression model as follows:

$$HI = \beta_0 + \beta_1 L + \beta_2 T + \beta_3 IR + \beta_4 t + \beta_5 t^2 + \cdots + \beta_k t^k \tag{1}$$

where L is the loading factor, T is the temperature factor, IR is the insulation resistance factor, t is the time elapsed since the start of the measurement, and β_0, β_1, β_2, β_3, $\ldots$, β_k are regression coefficients estimated from the data.

Additional benefits of load forecasting for distribution transformers include:

- *Help identifying anomalies*: Accurate load forecasting may assist in the transformer anomaly detections, such as unanticipated changes in loading patterns or abrupt spikes in load demands. These anomalies may signal a possible transformer malfunction, such as winding failures or insulation deterioration, and may be utilized as a trigger for additional examination and maintenance.
- *Support optimal maintenance scheduling*: Based on the expected loading patterns of a transformer, accurate load forecasting may be used to plan maintenance operations at optimal times, which can limit the maintenance effect on electrical systems and reduce risks of unplanned outages or equipment failures.

- *Improve reliability and resilience*: Accurate load fore-casting and transformer health monitoring may con-tribute to improved reliability and resilience of power systems by allowing early diagnosis of faults and proac-tive repair efforts. This may lower the likelihood of equipment failure, reduce downtime, and enhance the overall system performance.

III. THEORY OF HIERARCHICAL LOAD FORECAST RECONCILIATION

A. NOTATION

Let P_t be m-dimensional vector representing observations of all power consumption at time t, and b_t be n-dimensional vector representing observations at the bottom hierarchical level, i.e., the advanced metering infrastructure (AMI) data. The general matrix representation of P_t is [10]

$$P_t = Sb_t \qquad (2)$$

where S is the order $m \times n$ "summing matrix" that is used to aggregate the power consumption at the bottom-level hier-archy to the higher level. m represents the overall quantity of observations and n indicates the quantity of observations at the bottom level only. S contains hierarchical information, which can be obtained from the utility's geographical infor-mation system (GIS). Fig. 2 shows a hierarchical time series structure of the transformer loading.

Each higher hierarchy (parent) is the sum of the lower hier-archy (children). Let P_t represent the total transformer power at time t; $P_{A,t}$, $P_{B,t}$ and $P_{C,t}$ be transformer per phase power, and $P_{A1,t}$, $P_{A2,t}$,…$P_{C3,t}$ are power consumption (AMI data) of customers. In Fig. 2, $n = 8$ and $m = 12$.
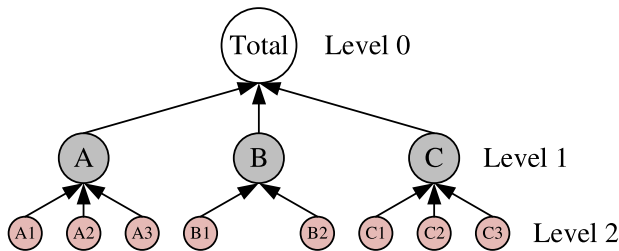

FIGURE 2. A hierarchical tree structure.

The hierarchical structure can be represented in a matrix form below, where I_n is an identity matrix representing the bottom level with a dimension, $n = 8$.

$$P_t = \begin{bmatrix} P_t \\ P_{A,t} \\ P_{B,t} \\ P_{C,t} \\ P_{A1,t} \\ P_{A2,t} \\ P_{A3,t} \\ P_{B1,t} \\ P_{B2,t} \\ P_{C1,t} \\ P_{C2,t} \\ P_{C3,t} \end{bmatrix} ; b_t = \begin{bmatrix} P_{A1,t} \\ P_{A2,t} \\ P_{A3,t} \\ P_{B1,t} \\ P_{B2,t} \\ P_{C1,t} \\ P_{C2,t} \\ P_{C3,t} \end{bmatrix} \text{ and}$$

$$S = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let's define $\hat{P}_T(h)$ as a vector in the same order as P_t, which represents the h-step-ahead base forecast of each time series derived from historical data extended for time T. It should be noted that any base machine learning or statistical forecasting methods can be used to get the base forecast. Linear reconciliation methods can be represented by [10].

$$\tilde{P}_T(h) = SG\hat{P}_T(h) \qquad (3)$$

where G is the carefully selected reconciliation matrix of order $n \times m$, and $\tilde{P}_T(h)$ is a collection of reconciled forecasts that are now constructively coherent. Forecast reconciliation techniques are based on the idea that a set of base forecasts can be mapped linearly to a set of reconciled forecasts. The objective of the G matrix is to convert the base forecasts to bottom-level disaggregated forecasts that are subsequently summed up by the S matrix.

The two commonly used methods for HLF forecasting are BU and TD. For BU, we set $G = [0_{n \times (m-n)} | I_n]$, where $0_{i \times j}$ is $i \times j$ null matrix with $i = n, j = m - n$, and I_n is the identity matrix with $n \times n$ dimensions. The base forecast at the bottom-level hierarchy yielded from $\hat{P}_T(h)$ are aggregated using S to get BU reconciled forecasts. Similarly, we can set $G = [\partial | 0_{n \times (m-1)}]$, where $\boldsymbol{\partial} = [\partial_1, \partial_2 \ldots \partial_n]$ is a proportion vector that disaggregates the base forecast at the top level to the bottom level, and then S is used to obtain TD reconciled forecasts [10]. As an example, the G matrices for BU and TD are illustrated in a matrix form for $n = 4$ and $m = 6$ as follows:

$$G_{BU} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad G_{TD} = \begin{bmatrix} \hat{\partial}_1 & 0 & 0 & 0 & 0 & 0 \\ \hat{\partial}_2 & 0 & 0 & 0 & 0 & 0 \\ \hat{\partial}_3 & 0 & 0 & 0 & 0 & 0 \\ \hat{\partial}_4 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

B. THE THEORY OF MINIMUM TRACE OPTIMAL RECONCILIATION

Let,

$$\hat{e}_T(h) = P_{T+h} - \hat{P}_T(h) \qquad (4)$$

where $\hat{e}_T(h)$ is the base forecast error for h-step-ahead, $\tilde{P}_T(h)$ is the base forecast for h-step-ahead, and P_{T+h} is the current true value.

A reconciled forecast with the minimum error variance is obtained such that $SGS = S$ is satisfied [10]. For every G matrix that satisfies $SGS = S$, the covariance matrix of the reconciled h-step-ahead forecast errors can be obtained by

$$V_h \text{Var} = \left[\left(P_{T+h} - \tilde{P}_h \right) \right] = SGW_hG^TS^T \qquad (5)$$

$$W_h = \text{Var}\left[\left(P_{T+h} - \hat{P}_h\right)\right] \qquad (6)$$

where $\tilde{P}_t(h)$ is determined by (3), and W_h represents base forecast errors variance-covariance matrix for h-step-ahead. Assuming that forecast errors are normally distributed, Eqs. (5) and (6) can be utilized to build forecast intervals for all forecasting methods that generate unbiased and coherent forecasts, assuming that original forecasts are unbiased. The quality of the reconciled forecast depends on the effectiveness of the estimator for W_h.

Residuals of the reconciled forecast are located at the diagonal of the V_h matrix and the summation of values located at the diagonal of a matrix is known as a "trace". Therefore, the final target is to obtain the G matrix that will result in a minimum trace V_h in (5) that satisfies $SGS = S$.

Theorem 1: Let W_h in (6) be the positive definite matrix. The optimal reconciliation matrix G that has the minimum trace of (5) and satisfies $SGS = S$, is given by (7).

$$G = \left(S^T W_h^{-1} S\right)^{-1} S^T W_h^{-1} \qquad (7)$$

Therefore, the reconciled forecast using the MinT method is given by

$$\tilde{P}_T(h) = S\left(S^T W_h^{-1} S\right)^{-1} S^T W_h^{-1} \hat{P}_T(h) \qquad (8)$$

Using MinT optimal reconciliation, the accuracy of reconciled forecasts is equal to or greater than the accuracy of base forecasts.

Let $\hat{\varepsilon}_h(h)$ be the error of the bottom-level base forecasts. Its variance-covariance matrix is W_h as in (6). Assuming that the error is additive, $\hat{e}_h(h) = S\hat{\varepsilon}_h(h)$. This results in $W_h = SV_hS^T$ to be singular. Eq. (5) can be further expressed by [10].

$$\text{Var}\left[\left(P_{T+h} - \tilde{P}_h\right)\right] = SGSV_h(SGS)^T = SV_hS^T \qquad (9)$$

Eq. (9) highlights that V_h is independent of the reconciliation matrix G. Considering the assumption of error-additivity, G that satisfies $SGS = S$ is the solution of MinT with the minimum variance.

Eq. (7) provides the solution for the ordinary least square (OLS) as follows:

$$G^{OLS} = (S^T S)^{-1} S^T, \qquad (10)$$

The state-of-the-art methods to compute the covariance matrix W_h are discussed below. These estimators are used as a benchmark to validate the proposed method.

C. THE BENCHMARK COVARIANCE MATRIX ESTIMATORS
In this section, state-of-the-art methods used to estimate W_h are discussed.

1) ORDINARY LEAST SQUARES (OLS)
Set $W_h = k_h I$, $\forall h$, where $k_h > 0$, W_h is the variance matrix of every base forecast error at the bottom level of the hierarchy, I is the $n \times n$ identity matrix. Referring (3) and (10), the reconciled forecast using OLS is estimated by (11).

$$\tilde{P}_T^{OLS}(h) = S(S^T S)^{-1} S^T \hat{P}_T(h) \qquad (11)$$

This method is only suitable in some situations, such as base forecast errors remain stochastic and equivariant. It has the most straightforward assumption, known as "homoscedasticity", and is applicable only if the identical method is applied to obtain base forecasts for each power consumption observation in all hierarchies of the transformer loading. The OLS estimator does not require assumptions when estimating the covariance matrix.

2) WEIGHTED LEAST SQUARES – VARIANCE SCALING
Set $W_h = k_h \text{diag}\left(\hat{W}_1\right)$, $\forall h$, where $k_h > 0$, and

$$\hat{W}_1 = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_T(1)\hat{e}_T'(1)' \qquad (12)$$

$\hat{W}_1$ is a sample covariance estimator of the base forecast error at $h = 1$. As expressed in (4) [10], $\hat{e}_T$ is a n-dimensional vector of residuals of the base forecast model, and $\hat{e}_T(1) = \hat{P}_{T+1} - \hat{P}_T(1)$. The weighted least squares (WLS) estimator utilizing variance scaling (VS) adjusts the base forecasts based on the variance of errors. The expression to compute $\tilde{P}_T(h)$ in (8) can be modified as follows to represent WLS-VS estimator:

$$\tilde{P}_T^{WLS-VS}(h) = S\left(S^T \hat{W}_{1,D}^{-1} S\right)^{-1} S^T \hat{W}_{1,D}^{-1} \hat{P}_T(h) \qquad (13)$$

WLS-VS can handle heteroscedasticity, when there is a different variance of base forecast errors at different hierarchical time series levels caused by using different methods to obtain base forecasts. This is an important feature of transformer load forecast aggregation as it is not always possible to use the same forecasting method to get the best base forecasts due to the variety of loads, such as residential, industrial, EVs, and DERs. For this reason, the WLS-VS estimator provides better accuracy compared to OLS.

3) WEIGHTED LEAST SQUARES – STRUCTURAL SCALING
Set $W_h = k_h \Lambda$, $\forall h$; $k_h > 0$, $\Lambda = \text{diag}(S1)$, and 1 is a n-dimensional unit vector. The expression to compute $\tilde{P}_T(h)$ in (8) can be adapted as follows to represent the WLS-SS estimator:

$$\tilde{P}_T^{WLS-SS}(h) = S\left(S^T \Lambda^{-1} S\right)^{-1} S^T \Lambda^{-1} \hat{P}_T(h) \qquad (14)$$

It is assumed that every bottom-level hierarchy base forecast error has a variance of k_h and that the errors are not attributed to each other. As a result, each element of Λ encompasses the quantity of forecast error variations attributable to the higher hierarchy level [24]. This estimator, known as the WLS estimator that uses structural scaling (SS), depends on the hierarchical structure. In contrast to OLS, this method is more realistic as it only requires that the variance of forecast errors is equal at the bottom level of the hierarchy. In addition, it is independent of historical data and can handle heteroscedasticity in hierarchical load forecast aggregation. It is mostly helpful when the base forecast is unavailable, for instance, at a very early stage of smart meters deployment by utilities, no sufficient historical samples are available. Table 1 shows the comparison among state-of-the-art benchmark covariance matrix estimators.

TABLE 1. Comparison of benchmark covariance matrix W_h estimators.

Estimator	Covariance Matrix	Number of Parameters Used	Variance*	Correlation
OLS (10)	$k_h I$	0	Homo	No
WLS-VS (12)	$k_h \widehat{W}_{1,D}$	n	Hetero	No
WLS-SS (13)	$k_h \Lambda$	0	Hetero	No

*Homo and Hetero represent Homoscedasticity and Heteroscedasticity, respectively

IV. THE PROPOSED METHOD

In the proposed method, MinTSa for smaller distribution networks and MinTSh for large distribution networks are used by considering information on the hierarchical structure and covariance of errors at the base forecast; and ARIMA [25] is chosen to simulate base forecasts due to its simplicity (Note: base forecasts can be done through any existing forecasting method, and it is not the focus of this study; the optimal reconciliation for base forecasts is the main focus).

A. THE PROPOSED COVARIANCE MATRIX ESTIMATORS

1) MIN-T SAMPLE (MINTSA)

The covariance matrix estimated by the MinT estimator is a $n \times n$ without zero elements. For MinTSa, we set $W_h = k_h \widehat{W}_1$, $\forall h$, where $k_h > 0$, and $\widehat{W}_1$ is a sample covariance estimator of the base forecast error at $h = 1$. Given that only a sample of $\widehat{W}_h$ at $h = 1$ is utilized, MinTSa is relatively simple compared to MinTSh. It works best when $m < T$, i.e., the total quantity of observations is lower than the quantity of historical data used. Practically, the number of customers served by a particular distribution transformer is not greater than 50 in most cases. For T to be greater than 50, historical data must at least cover 12.5 hours in a 15-minute resolution. Similarly, 250 hours (10.42 days) of data is required for a substation transformer with 1,000 customers.

The expression to compute $\tilde{P}_T(h)$ in (8) can be further expanded as in (15) to reflect the MinTSa estimator.

$$\tilde{P}_T^{\text{MinTSa}}(h) = S \left(S^T \widehat{W}_1^{-1} S \right)^{-1} S^T \widehat{W}_1^{-1} \hat{P}_T(h), \quad (15)$$

2) MIN-T SHRINKAGE (MINTSH)

Set $W_h = k_h \widehat{W}_{1,D}^*$, $\forall h$, where $k_h > 0$, and $\widehat{W}_{1,D}^*$ is the diagonal of the shrinkage estimator.

$$\widehat{W}_{1,D}^* = \lambda_D \widehat{W}_{1,D} + (1 - \lambda_D) \widehat{W}_1 \quad (16)$$

where $\widehat{W}_{1,D}$ is a diagonal of the matrix $\widehat{W}_1$, and λ_D is the shrinkage parameter. The off-diagonal components of $\widehat{W}_1$ are substantially reduced toward zero, but the diagonal values (variances) remain unchanged. Assuming variances are constant, λ_D can be computed by

$$\lambda_D = \frac{\sum_{i \neq j} Var(\hat{r}_{ij})}{\sum_{i \neq j} (\hat{r}_{ij}^2)} \quad (17)$$

where $\hat{r}_{ij}$ is the $(i, j)^{th}$ component of the correlation matrix when $h = 1$ (one step ahead) shrinks it to become an identity matrix. In the case of MinTSa, $\lambda_D = 0$, leading to $W_h = k_h \widehat{W}_1$. The MinTSh estimator is used when a large bottom-level

series is present. For a higher hierarchy level at a substation transformer, MinTSh has a better performance than MinTSa.

The expression to compute $\tilde{P}_T(h)$ in (8) can be further extended as follows to represent the MinTSh estimator:

$$\tilde{P}_T^{\text{MinTSh}}(h) = S \left(S^T \widehat{W}_{1,D}^{*-1} S \right)^{-1} S^T \widehat{W}_{1,D}^{*-1} \hat{P}_T(h) \quad (18)$$

The main innovation of the proposed method compared to existing hierarchical load forecast methods is the new and efficient approach to compute the covariance matrix W_h and to consider the covariance of the base forecast errors of two time series data in the hierarchy.

B. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

ARIMA is a linear forecasting approach that works well with static time series data. In the first step, a static time series is built through differencing d times and performing different nonlinear operations like logging [26]. The resulting data is characterized as a linear transformation of previous p datasets and q errors as indicated in (19).

$$P_t = \Omega_1 P_{t-1} + \Omega_2 P_{t-2} + \ldots + \Omega_p P_{t-p} + \Psi_1 \varepsilon_{t-1}$$
$$+ \Psi_2 \varepsilon_{t-2} + \ldots \Psi_q \varepsilon_{q-1} \quad (19)$$

where P_t is the true value at a given time t, ε_t is an error, which is Gaussian distributed $(0, \sigma^2)$ characterized white noise, Ω_i is autoregressive (AR) coefficients for $(i = 1, 2, \ldots, p)$, and Ψ_j are moving average (MA) coefficients for $(j = 1, 2, \ldots, q)$. The integers p and q represent orders of AR and MA, respectively. The important parameters to model ARIMA are p, q and d.

C. ALGORITHM OF THE PROPOSED METHOD

The proposed optimal reconciliation method is MinT, which has two variations: MinTSa and MinTSh. The base forecast and real values are used to calculate errors, then the hierarchical tree information is used along with the reconciliation algorithm to obtain the reconciled forecast, and a lower error is ensured by comparing them to the base forecast. The ultimate task of the reconciliation algorithm is to reduce the error of the reconciled forecast, in turn improving its performance. The base forecast errors of transformer loading in the hierarchy are correlated, resulting in the elements outside the diagonal of the covariance matrix W_h to have non-zero values. The proposed method can better handle this situation than the benchmark covariance matrix estimators. The proposed method is shown in Fig. 3 and Table 2.

D. FORECAST EVALUATION METRICS

Error metrics to evaluate the proposed method include the mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and root-mean-square error (RMSE). MAE measures the average absolute differences, MAPE measures the average percentage differences, and MSE calculates the average squared differences between predicted and actual values. RMSE is the square root of MSE, providing a more interpretable measure. Lower values of these metrics indicate better model performance, and thus, they offer a quantitative comparison and evaluation
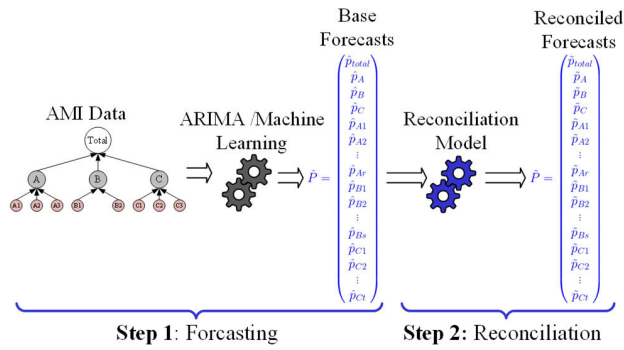
FIGURE 3. Graphical representation of the proposed method.

TABLE 2. Algorithm 1.

MinT Optimal Reconciliation Algorithm
Input: Power consumption $P_{Cx} = P_{Cx}(t-h) \dots P_{Cx}(t-1), P_{Cx}(t)$, C is the number of customers, h is the lagging horizon, and x is the index
Output: Reconciled Load Forecast $\widetilde{P}_h$
1:      **For** h = 1 to h = H **do**: Start Base Forecast $\widehat{P}_h$
2*:   **Set** p, d, q: ARIMA Coefficients
3*: $P_h = \Omega_1 P_{t-1} + \Omega_2 P_{t-2} + \dots + \Omega_p P_{t-p} + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots \Psi_q \varepsilon_{q-1}$ (19)
4:      End the **for** loop
5:      **For** h = 1 to h = H **do**: Start Reconciliation
6:      **If** W_h the estimator is MinTSh, Set $W_1 = \lambda_D \widehat{W}_{1,D} + (1-\lambda_D)\widehat{W}_1$; $\widehat{W}_{1,D} = \text{diag}(\widehat{W}_1)$; $\lambda_D = \frac{\sum_{i \neq j} Var(\hat{r}_{ij})}{\sum_{i \neq j}(\hat{r}_{ij}^2)}$; Compute (15)
else if W_h the estimator is MinTSa, Set $W_h = k_h \widehat{W}_1$; $\lambda_D = 0$; Compute (17)
7:      Calculate $\widetilde{P}_h = S(S^T W_h^{-1} S)^{-1} S^T W_h^{-1} \widehat{P}_h$: (8)
8:      End the **for** loop

* Steps 2 and 3 can be changed according to the chosen base forecasting algorithm.

of different models' accuracy. They can be mathematically expressed by (20) - (23).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| P_i - \hat{P}_i \right|, \tag{20}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{P_i - \hat{P}_i}{P_i} \right|, \tag{21}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(P_i - \hat{P}_i \right)^2, \tag{22}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(P_i - \hat{P}_i \right)^2}{n}}, \tag{23}$$

where P_i is the true value, $\hat{P}_i$ is the forecasted value, n refers to the number of samples, and i is the index.

V. CASE STUDIES
To validate the proposed method, five case studies are conducted using three test grids.

A. TEST GRIDS
Test systems A, B and C are used to validate the proposed method.
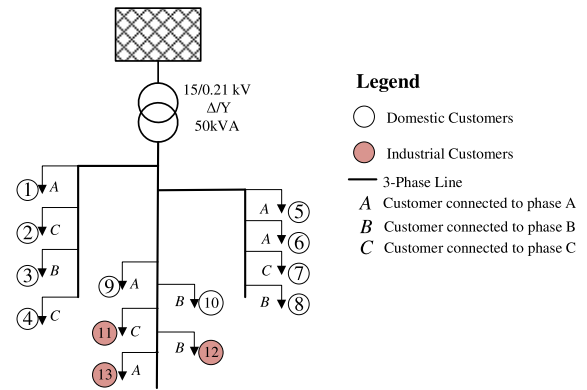

FIGURE 4. Test system A.

Test system A is composed of one 50 kVA-15/ 0.21 kV-Delta/Wye distribution transformer, 10 domestic customers and 3 industrial customers. The grid parameters are obtained from [27]. The details of Test System A are illustrated in Fig. 4.
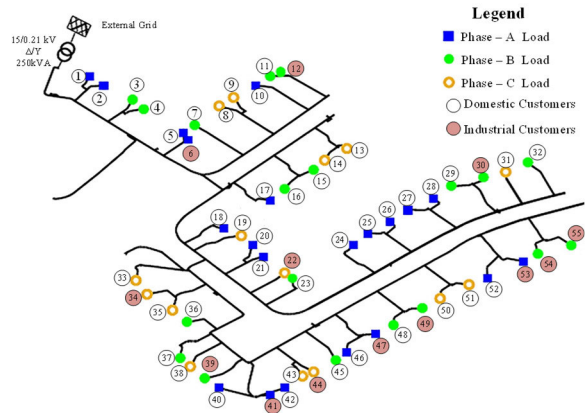
FIGURE 5. Test system B.

Test system B is the IEEE European network [28], [29] and is modified to represent a North American grid. It operates as an unbalanced grid, delivering single-phase connections to 55 domestic and industrial customers, with 21, 19 and 15 connected to phases A, B and C, respectively. Fig. 5 shows a detailed geographical single-phase diagram of Test System B.

Fig. 6 shows Test system C, a more complicated distribution grid composed of one 2 MVA 132/14.4 kV substation transformer, twenty 14.4/0.21 kV distribution transformers, and 1,000 customers (18% are industrial customers and the rest are domestic customers). Power ratings of distribution transformers range from 25 to 200 kVA, and the number of customers connected to each transformer ranges from 13 to 62. Test system C has one more hierarchy than Test systems A and B.

FIGURE 6. Test system C.

B. AMI DATASET DESCRIPTION

The electricity power consumption data is provided by Saskatoon Light and Power, a distribution utility in Saskatoon, SK, Canada. The data has a resolution of 15 minutes for over 65,000 customers in the industrial, commercial, and residential sectors. A dataset for 1,000 customers is selected for this paper, 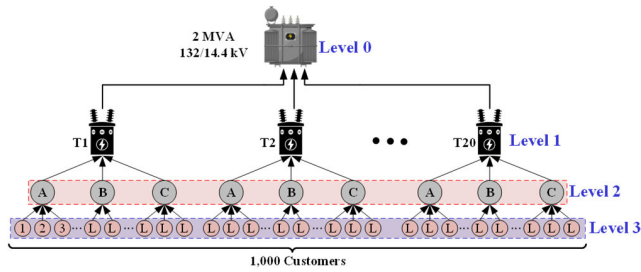which was recorded from 0:15 AM on March 1, 2017, to 11:45 PM on March 1, 2018. With a time resolution of 15 minutes, there are 35,136 steps in the dataset and a one-time horizon $h = 15$ minutes. In the proposed method, to get base forecasts using ARIMA, 80% of the data is used for training the model, and 20% is used for testing the performance of the model.

C. NUMERICAL RESULTS OF CASE STUDIES

To validate the proposed method (MinTSh and MinTSa), five case studies are conducted, in which six prediction horizons are considered, ranging from 15 minutes to eight hours ($h = 1, 2, 4, 8, 16, 32$). Four state-of-the-art benchmark methods (BU, OLS, WLS-SS, and WLS-VS) are chosen to compare with the proposed method.

In Cases 1-4, optimal reconciliation is performed at all three hierarchical levels of the transformer loading: 1) the total load – top level (Level 0), 2) the three-phase load – middle level (Level 1), and 3) customer consumption – bottom level (Level 2). In Case 5, one higher hierarchy level is added to incorporate a load for a substation transformer.

1) CASE 1 – SMALL NETWORK SIMULATION USING ONE MONTH DATA

In Case 1, one-month historical data is used for base forecasts, optimal reconciliation is then conducted using the proposed method with such limited data using Test System A. In this case, the accuracy improvement of the reconciled forecast vs. base forecasts evaluated by MAE is shown in Table 3 for all hierarchical levels and six forecast horizons. MinTSa outperforms other methods, followed by MinTSh. At the forecast horizon $h = 2$ and hierarchy level 1, there is a 7.41% improved accuracy. The proposed method (MinTSa and MinTSh) performs well for load forecasts at the middle level, but MinTSa performs better than MinTSh for a lower number of customer load forecasts at the bottom level. Among the four benchmark methods, BU performs the worst with negative values, indicating lower accuracy than base forecasts. BU provides 0% improved accuracy at the bottom level, so it is only suitable for a higher-level load forecast.

Table 4 shows improved accuracy evaluated by MAPE. MinTSa performs better than others in Levels 0 and 1 for all

six horizons, but MinTSh performs better at Level 2. Bold numbers in the table show the best accuracy improvement. Although benchmark methods (except BU) have acceptable results, they are less accurate than the proposed method.

In Table 5, MinTSa has the highest improved accuracy of 9.82% for the forecast horizon $h = 2$ at level 1 evaluated by MSE. As the forecast horizon increases, the performance of MinT tends to decrease at all levels. However, the proposed method is still suitable for short/medium horizon prediction in distribution systems.

TABLE 3. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by MAE (Case 1).

h	1	2	4	8	16	32	Ave.
Hierarchical level 0							
BU	-5.88	-6.25	-6.65	0.32	-5.88	-6.98	-5.22
OLS	-0.01	-0.09	0.14	0.39	-0.01	-0.01	0.07
WLS-SS	-0.06	-0.34	-0.09	1.53	-0.06	-0.03	0.16
WLS-VS	0.09	-0.18	0.05	1.46	0.09	0.09	0.27
MinTSh	0.26	0.01	0.14	1.67	0.26	0.23	0.43
MinTSa	1.47	1.14	0.78	2.82	1.47	1.34	1.50
Hierarchical level 1							
BU	1.40	1.75	0.97	3.38	1.30	1.46	1.71
OLS	3.20	3.36	3.05	2.05	2.85	2.96	2.91
WLS-SS	4.90	5.11	4.75	4.33	3.86	4.17	4.52
WLS-VS	5.38	5.60	5.15	4.53	4.66	4.96	5.05
MinTSh	5.58	5.82	5.33	4.85	5.16	5.26	5.33
MinTSa	7.36	7.41	6.56	6.37	6.86	6.95	6.92
Hierarchical level 2							
BU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	-3.14	-3.24	-3.11	-2.23	-3.14	-2.86	-2.95
WLS-SS	-1.00	-1.05	-0.98	-0.70	-1.00	-0.75	-0.91
WLS-VS	1.46	1.38	1.41	0.42	1.46	1.20	1.22
MinTSh	1.94	1.84	1.84	0.87	1.94	1.13	1.59
MinTSa	2.89	3.00	2.75	2.20	2.89	2.46	2.70

Table 6 shows the accuracy improvement of the proposed reconciled forecast vs. base forecasts evaluated by RMSE. MinTSa shows the best performance, followed by MinTSh. MinTSa has the highest improved accuracy among all linear reconciliation methods, including OLS and WLS, with the most accurate and coherent forecasts at all hierarchy levels.

The accuracy improvement of the overall hierarchical tree structure for Case 1 is shown in Fig. 7. The proposed reconciliation models based on trace minimization yield accurate and consistent predictions at all hierarchical levels.

Fig. 8 shows a comparison between the true, predicted, and reconciled values for the first 100 samples. The reconciled lines (black and blue solid lines) are closest to the true line (dashed light blue line) with the lowest errors. MinTSa and MinTSh show consistent performance across all prediction horizons.

2) CASE 2 – SMALL NETWORK RANDOM-ORIGIN SIMULATION USING ONE MONTH DATA

In Case 2, one-month historical data is used for base forecasts in Test System A. The simulation for the forecast horizon, $h = 1$, is conducted for one particular time snap; and again for 30 random time snaps, and their average is taken. The objective of Case 2 is to show that the proposed method for
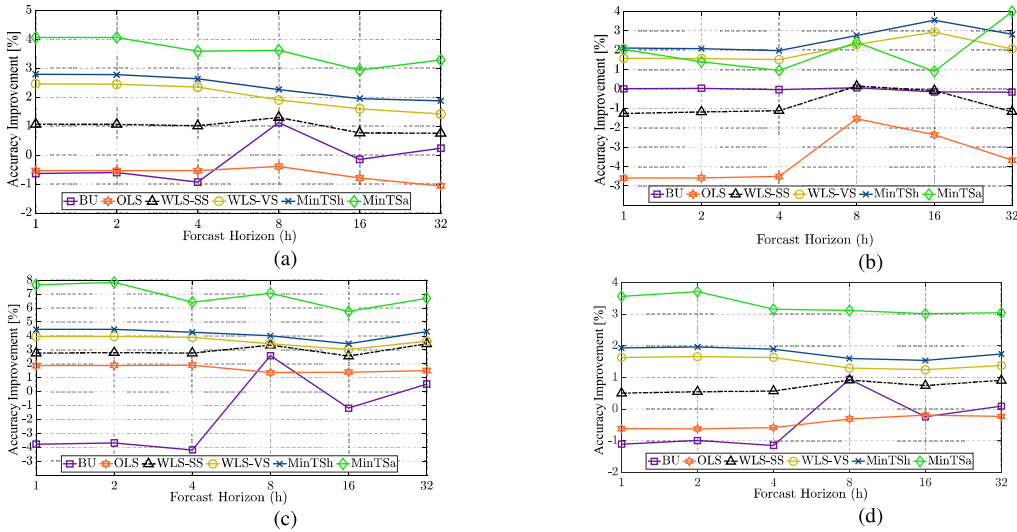
FIGURE 7. The accuracy improvement in % of the overall hierarchical tree structure for Case 1: (a) MAE, (b) MAPE, (c) MSE and (d) RMSE.
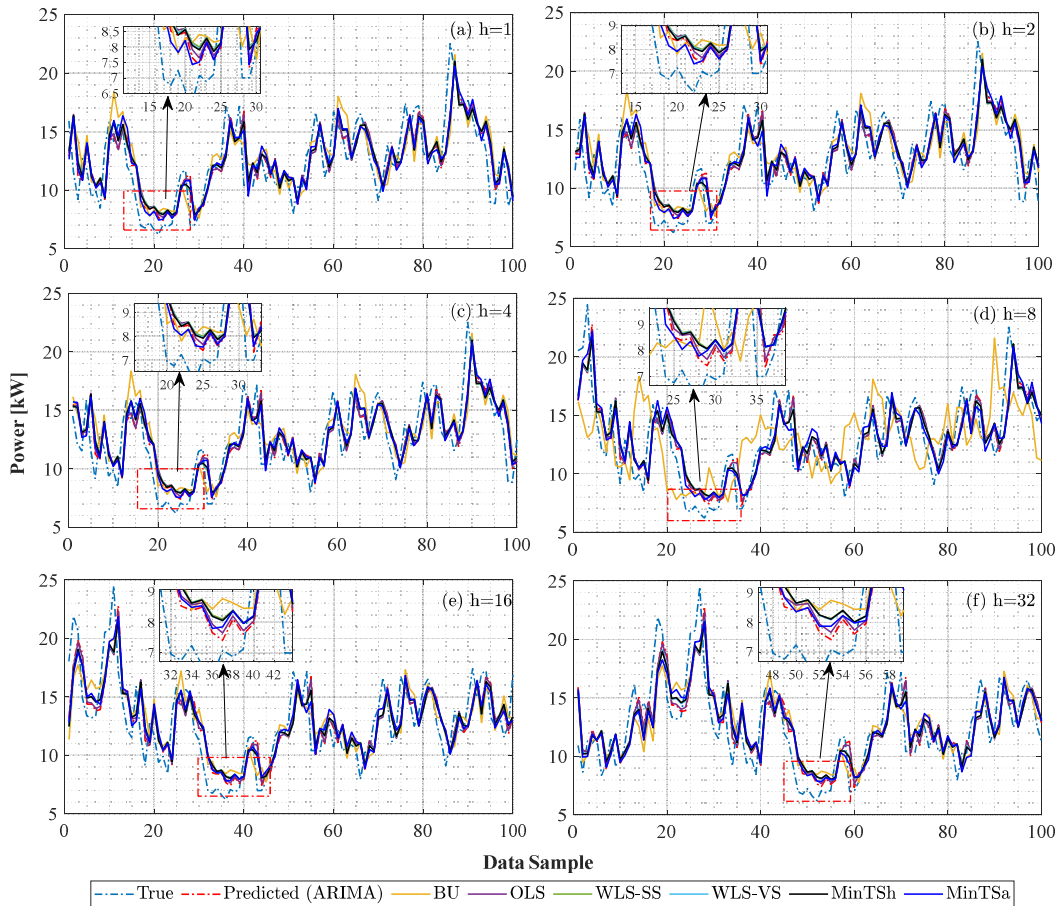
FIGURE 8. Comparison of true, predicted and reconciled forecasts – Case 1: (a) h=1, (b) h=2, (c) h=4, (d) h=8, (e) h=16, (f) h=32.

load forecast performs well at any time of the year. The two sets of simulations are compared in Table 7. "h= 1(30)" means the average of 30 simulations, and "h = 1" means

the result of one simulation. In fact, "h= 1(30)" shows even better accuracy than "h = 1" for most hierarchy levels and accuracy metrics for both MinTSa and MinTSh.

FIGURE 9. Comparison of true, predicted and reconciled forecast – Case 2: $h = 1(30)$: (a) Phase A, (b) Phase B, (c) Phase C, (d) Total Power.

TABLE 4. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by MAPE (Case 1).

Hierarchical level 0						
h	1	2	4	8	16	32	Ave.
BU	-3.69	-4.01	-4.49	-0.44	-2.96	0.06	-2.59
OLS	-0.49	-0.57	-0.38	-0.18	-0.12	0.29	-0.24
WLS-SS	-0.26	-0.51	-0.33	0.51	0.03	1.22	0.11
WLS-VS	-0.08	-0.33	-0.17	0.52	0.12	1.17	0.20
MinTSh	0.21	-0.03	-0.01	0.80	0.16	1.24	0.39
MinTSa	**2.80**	**2.50**	**1.76**	**3.62**	**0.83**	**1.61**	**2.19**
Hierarchical level 1						
h	1	2	4	8	16	32	Ave.
BU	0.89	1.20	0.59	0.78	-0.90	-1.63	0.16
OLS	3.16	3.33	3.07	1.97	1.44	1.27	2.37
WLS-SS	4.10	4.30	4.04	3.12	2.37	1.93	3.31
WLS-VS	4.69	4.87	4.50	3.52	2.73	2.26	3.76
MinTSh	4.84	5.03	4.63	3.70	2.83	2.28	3.89
MinTSa	**6.28**	**6.31**	**5.39**	**4.77**	**2.76**	**2.19**	**4.62**
Hierarchical level 2						
h	1	2	4	8	16	32	Ave.
BU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	-5.60	-5.61	-5.49	-1.95	-2.84	-4.28	-4.29
WLS-SS	-1.92	-1.85	-1.74	-0.19	-0.36	-1.55	-1.27
WLS-VS	1.26	1.23	1.21	2.20	3.03	2.06	1.83
MinTSh	**1.84**	**1.77**	**1.72**	**2.70**	**3.71**	2.91	**2.44**
MinTSa	1.53	0.80	0.42	2.13	0.70	**4.24**	1.64

A comparison of the true, predicted, and reconciled forecasts for phases A, B, C and the total power of three phases for $h = 1(30)$ is presented in Fig. 9. The superiority of the proposed method is confirmed as the MinTSa and MinTSh reconciled lines are closest towards the true line, showing the tendency of the proposed reconciliation method to minimize the error.
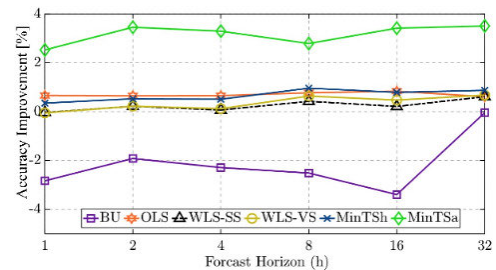
FIGURE 10. The Accuracy Improvement of the Overall Hierarchical Tree Structure for Case 3 using RMSE.

3) CASE 3 – SMALL NETWORK SIMULATION USING ONE YEAR DATA

In Case 3, one-year historical data is used for base forecasts in Test system A, and the proposed method is then applied for optimal reconciliation to determine if it can handle high data variance in one year due to seasonal changes, which causes dramatic changes of electricity consumption behaviours of customers using Test system A. In Case 3, RMSE is used for evaluation. In Table 8, the average accuracy improvement of MinTSa is the highest at all levels followed by MinTSh, indicating the robustness of the proposed method subjected to large variances in the dataset. The accuracy improvement of the overall hierarchical tree structure of the proposed method vs. benchmark methods for Case 3 is shown in Fig. 10, and the proposed method shows superior performance over benchmark methods. MinTSa (the green line) is always on the top, mostly followed by MinTSh (the blue dark line).

The comparison of the true, predicted, and reconciled forecasts using the proposed method and four benchmark methods for the first 100 samples in Case 3 is shown in Fig. 11.
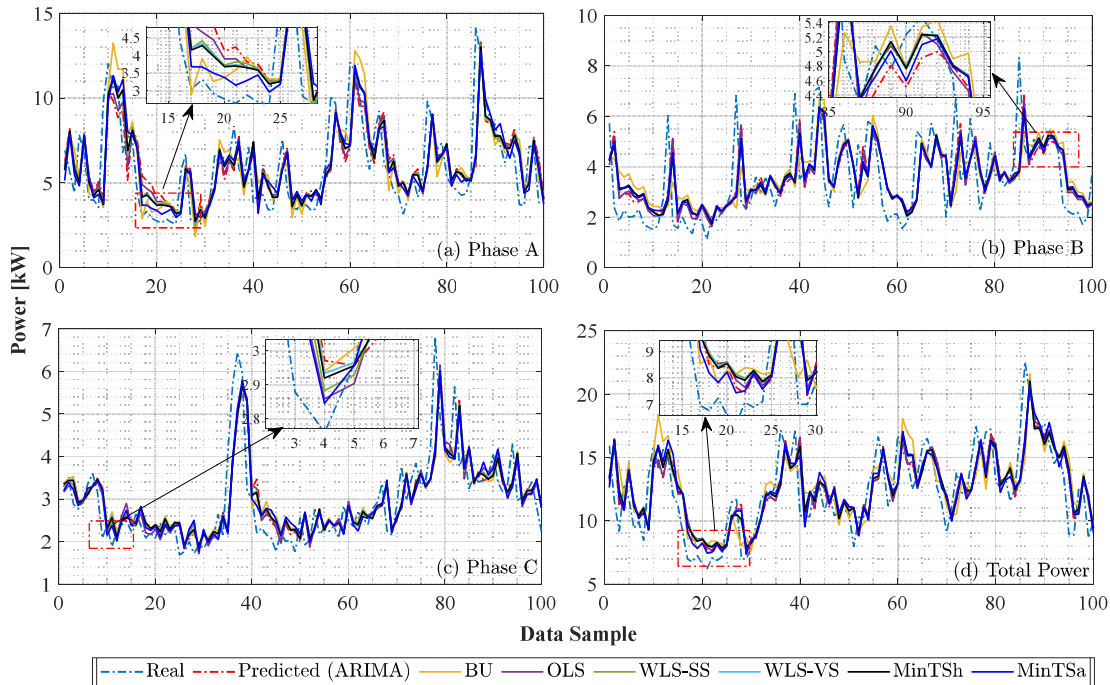
FIGURE 11. Comparison of true, predicted and reconciled forecasts – Case 3: (a) h = 1, (b) h = 2, (c) h = 4, (d) h = 16, (e) h = 32.

TABLE 5. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by MSE (Case 1).

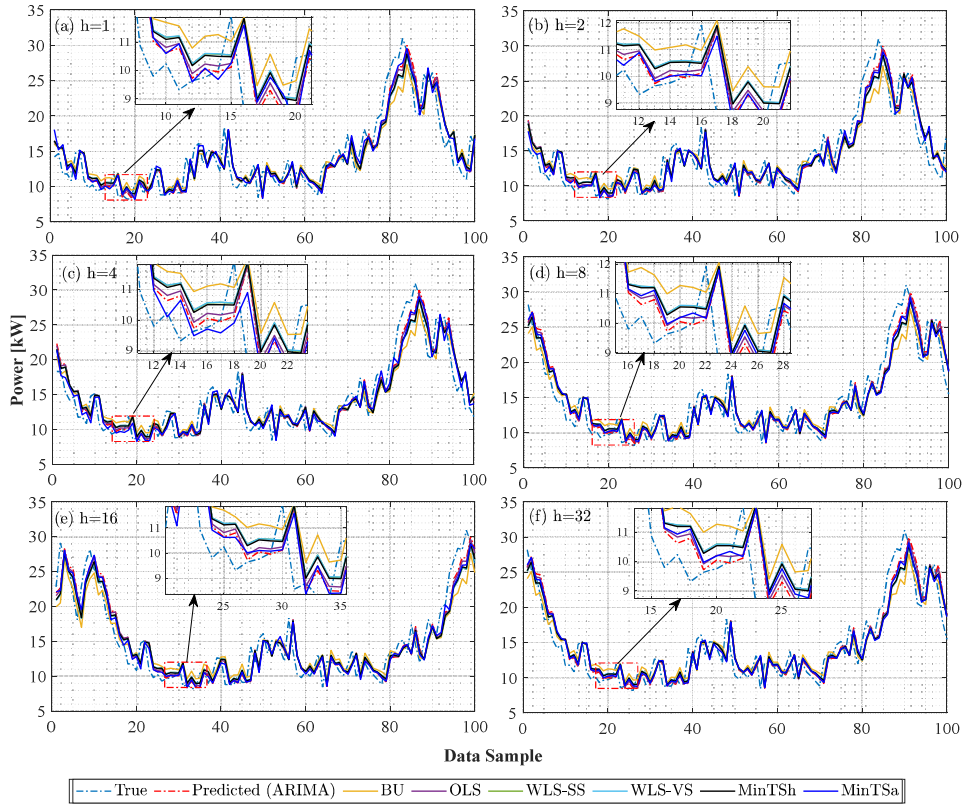| Hierarchical level 0							
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave.**
BU	-13.9	-13.8	-14.1	3.2	-3.9	3.33	-6.52
OLS	-0.10	-0.16	0.31	1.18	1.23	2.31	0.79
WLS-SS	-0.13	-0.25	0.35	4.25	3.00	6.48	2.28
WLS-VS	0.40	0.26	0.78	4.28	3.07	6.28	2.51
MinTSh	0.86	0.74	1.03	4.87	3.27	6.75	2.92
MinTSa	**4.68**	**5.03**	**3.14**	**9.21**	**5.83**	**8.89**	**6.13**
Hierarchical level 1							
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave.**
BU	-0.05	0.28	-0.88	4.62	-0.2	-1.5	0.39
OLS	4.25	4.39	3.96	3.14	2.36	1.66	3.29
WLS-SS	6.27	6.49	5.88	5.62	3.82	3.21	5.21
WLS-VS	6.67	6.92	6.24	5.50	3.62	2.76	5.29
MinTSh	6.97	7.24	6.46	5.93	3.90	3.41	5.65
MinTSa	**9.23**	**9.82**	**7.92**	**8.34**	**5.60**	**5.54**	**7.74**
Hierarchical level 2							
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave.**
BU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	0.95	0.90	1.04	-0.3	0.59	0.69	0.64
WLS-SS	1.38	1.36	1.46	0.26	0.92	1.00	1.06
WLS-VS	3.88	3.79	3.87	0.64	2.38	2.10	2.78
MinTSh	4.62	4.49	4.53	1.34	3.14	3.05	3.53
MinTSa	**8.34**	**8.00**	**7.43**	**4.03**	**5.86**	**5.90**	**6.59**

TABLE 6. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by RMSE (Case 1).

Hierarchical level 0							
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave**
BU	-6.71	-6.69	-6.80	1.59	-1.91	1.68	-3.1
OLS	-0.05	-0.08	0.15	0.59	0.62	1.16	0.40
WLS-SS	-0.07	-0.12	0.17	2.15	1.51	3.29	1.16
WLS-VS	0.20	0.13	0.39	2.16	1.55	3.19	1.27
MinTSh	0.43	0.37	0.52	2.46	1.65	3.43	1.48
MinTSa	**2.37**	**2.55**	**1.58**	**4.71**	**2.96**	**4.55**	**3.12**
Hierarchical level 1							
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave**
BU	-0.12	0.28	2.26	3.38	0.24	-0.68	0.89
OLS	1.91	2.02	1.52	2.05	1.33	0.95	1.63
WLS-SS	2.87	3.09	2.82	4.33	2.14	1.68	2.82
WLS-VS	2.95	3.18	2.68	4.53	1.92	1.37	2.77
MinTSh	3.11	3.36	2.93	4.85	1.65	1.68	3.01
MinTSa	**4.21**	**4.74**	**4.30**	**6.37**	**3.46**	**2.76**	**4.31**
Hierarchical level 2							
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave**
BU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	-2.21	-2.28	-2.16	-1.61	-1.27	-1.32	-1.8
WLS-SS	-0.65	-0.68	-0.59	-0.53	-0.26	-0.27	-0.5
WLS-VS	1.32	1.26	1.30	0.25	0.77	0.80	0.95
MinTSh	1.72	1.65	1.66	0.59	1.17	1.23	1.34
MinTSa	**3.57**	**3.49**	**3.18**	**1.97**	**2.79**	**2.71**	**2.95**

Reconciliation algorithms push the predicted line, which is the result of ARIMA, toward the true line to minimize the error. The error is a vertical distance between the true and reconciled lines. The proposed method has the lowest errors for all prediction horizons as reconciled lines of MinTSh and MinTSa are closest to the true line.

TABLE 7. Single and random-origin accuracy improvement (%) comparison of MinTSa and MinTSh for $h = 1$ (Case 2).

			MAE	MAPE	MSE	RMSE
MinTSa	Level 0	h+1 (30)	**2.574**	**3.310**	**9.110**	**4.674**
		h+1	1.467	2.804	4.685	2.371
	Level 1	h+1 (30)	**8.118**	**6.834**	**14.164**	**7.348**
		h+1	7.364	6.276	9.232	4.205
	Level 2	h+1 (30)	**3.332**	**2.467**	**8.336**	2.779
		h+1	2.894	1.535	5.852	**3.568**
	Overall	h+1 (30)	**4.747**	**2.923**	**9.903**	**4.499**
		h+1	4.068	2.052	7.676	3.568
MinTSh	Level 0	h+1 (30)	**0.656**	**0.290**	2.825	1.424
		h+1	0.260	0.211	**0.863**	**0.432**
	Level 1	h+1 (30)	**5.888**	**5.294**	**9.062**	**4.758**
		h+1	5.576	4.842	6.967	3.109
	Level 2	h+1 (30)	1.413	**2.632**	1.840	0.728
		h+1	**1.940**	1.839	**4.618**	**1.720**
	Overall	h+1 (30)	2.727	**2.777**	**4.842**	**2.088**
		h+1	**2.799**	2.118	4.462	1.932

TABLE 8. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by RMSE (Case 3).

	Hierarchical level 0						
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave**
BU	-6.79	-5.32	-6.04	-5.09	-9.34	0.55	-5.34
OLS	0.45	0.23	0.22	1.01	0.27	0.58	0.46
WLS-SS	-0.74	-0.49	-0.76	0.79	-1.00	1.21	-0.16
WLS-VS	-0.81	-0.55	-0.84	0.80	-1.01	1.18	-0.20
MinTSh	-0.32	-0.29	-0.39	0.99	-0.60	1.41	0.13
MinTSa	2.21	1.47	2.68	1.63	1.51	6.29	2.63
	Hierarchical level 1						
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave**
BU	-0.62	0.63	-0.34	-0.51	-2.23	1.22	-0.31
OLS	0.72	1.12	0.98	0.34	0.64	0.72	0.75
WLS-SS	0.97	1.66	1.19	0.93	0.53	1.33	1.10
WLS-VS	0.92	1.61	1.17	1.09	0.70	1.21	1.12
MinTSh	1.21	1.86	1.43	1.37	0.89	1.39	1.36
MinTSa	4.18	4.85	3.78	4.70	2.74	3.86	4.02
	Hierarchical level 2						
h	**1**	**2**	**4**	**8**	**16**	**32**	**Ave**
BU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OLS	0.36	0.02	0.22	-0.03	0.44	-0.33	0.11
WLS-SS	0.55	0.33	0.44	0.37	0.67	-0.03	0.39
WLS-VS	0.25	0.10	0.31	0.39	0.81	-0.02	0.31
MinTSh	0.66	0.45	0.68	0.77	1.01	0.13	0.61
MinTSa	3.63	2.69	2.88	3.90	2.69	2.13	2.99

4) CASE 4 – UNBALANCED MEDIUM NETWORK SIMULATION

Test system B is used in Case 4 to simulate load forecast aggregation using the proposed method for an unbalanced medium size network with 55 customers. The simulation is done for two forecast horizons, $h = 1$, and 32, and one-month historical data is used to generate base forecasts. Table 9 shows the accuracy improvement of reconciled forecasts compared to base forecasts according to error indices, MAE, MAPE, MSE, and RMSE, for all levels of the hierarchy. MinTSa outperforms other methods, followed by MinTSh. The proposed methods (MinTSa and MinTSh) have the highest improved forecast accuracy at the middle level. The bold numbers in Table 9 indicate the best accuracy improvement.

Although the benchmark methods (except BU) show some improvements, they are less accurate than the proposed method. The highest accuracy improvements scored by the proposed methods are 8.51%, 7.27%, 13.47%, and 6.88% for MAE, MAPE, MSE, and RMSE, respectively. Among the benchmark methods, BU performs the worst with negative values, indicating lower accuracy than base forecasts. BU provides 0% improved accuracy at the bottom level, so it is only suitable for higher-level load forecast aggregation. MinTSa and MinTSh show consistent performance across both horizons.

Fig. 12 shows the comparison of the true, predicted, and reconciled forecasts for the first 100 samples of Case 4 ($h = 1$ and 32 forecast horizons for phases A, B, C, and for the total power among three phases). MinTSa (blue solid line) and MinTSh (black solid line) are the closest to the true line (dashed blue line) compared to the lines of the benchmark methods. The superiority of the proposed method is confirmed as the MinTSa reconciled line is the closest to the true line, showing the tendency of the proposed reconciliation method to minimize the error (a vertical distance between the true and reconciled line). Case 4 demonstrates that the proposed method performs well in a medium-sized distribution grid.

TABLE 9. Accuracy improvement (%) of the proposed method vs benchmark methods (Case 4).

h	1 (15 min)			32 (8 hrs)		
Level	**0**	**1**	**2**	**0**	**1**	**2**
	MAE					
BU	-2.38	3.30	0.00	0.51	1.69	0.00
OLS	0.44	0.92	-2.03	0.72	0.43	-2.61
WLS-SS	1.71	4.58	-0.31	3.29	3.79	-0.61
WLS-VS	1.81	4.81	0.33	3.39	3.85	0.53
MinTSh	1.81	4.92	**2.66**	**3.39**	3.88	2.10
MinTSa	**1.97**	**8.06**	2.37	3.06	**8.51**	**2.65**
	MAPE					
BU	-5.15	0.92	0.00	-3.01	-0.38	0.00
OLS	0.39	0.96	-3.79	0.65	0.49	-5.92
WLS-SS	1.05	3.92	0.62	2.46	3.48	-0.71
WLS-VS	1.08	4.00	1.30	2.46	3.48	2.30
MinTSh	1.09	4.12	1.48	**2.45**	3.52	2.42
MinTSa	**1.29**	**7.27**	**2.65**	1.98	**4.45**	**3.41**
	MSE					
BU	-8.80	5.60	0.00	-2.61	5.29	0.98
OLS	0.16	2.05	-0.24	1.32	1.02	-1.23
WLS-SS	1.97	7.80	0.15	6.37	7.77	2.01
WLS-VS	1.97	8.22	-0.31	6.52	8.15	2.55
MinTSh	2.01	8.37	2.56	6.55	8.25	3.98
MinTSa	**4.84**	**13.47**	**4.34**	**8.23**	**11.2**	**4.07**
	RMSE					
BU	-4.30	2.44	0.00	-1.30	2.87	0.00
OLS	0.08	0.81	-1.50	0.66	0.35	-1.77
WLS-SS	0.99	3.68	-0.46	3.24	3.92	-0.61
WLS-VS	0.99	3.87	0.02	3.31	4.11	0.01
MinTSh	1.01	3.96	0.12	3.33	4.17	1.25
MinTSa	**2.45**	**6.88**	**2.44**	**4.21**	**5.80**	**1.96**

5) CASE 5 – LARGE NETWORK SIMULATION

Case 5 aims to validate the proposed method in a real-world distribution grid using Test System C with a substation transformer and distribution transformers
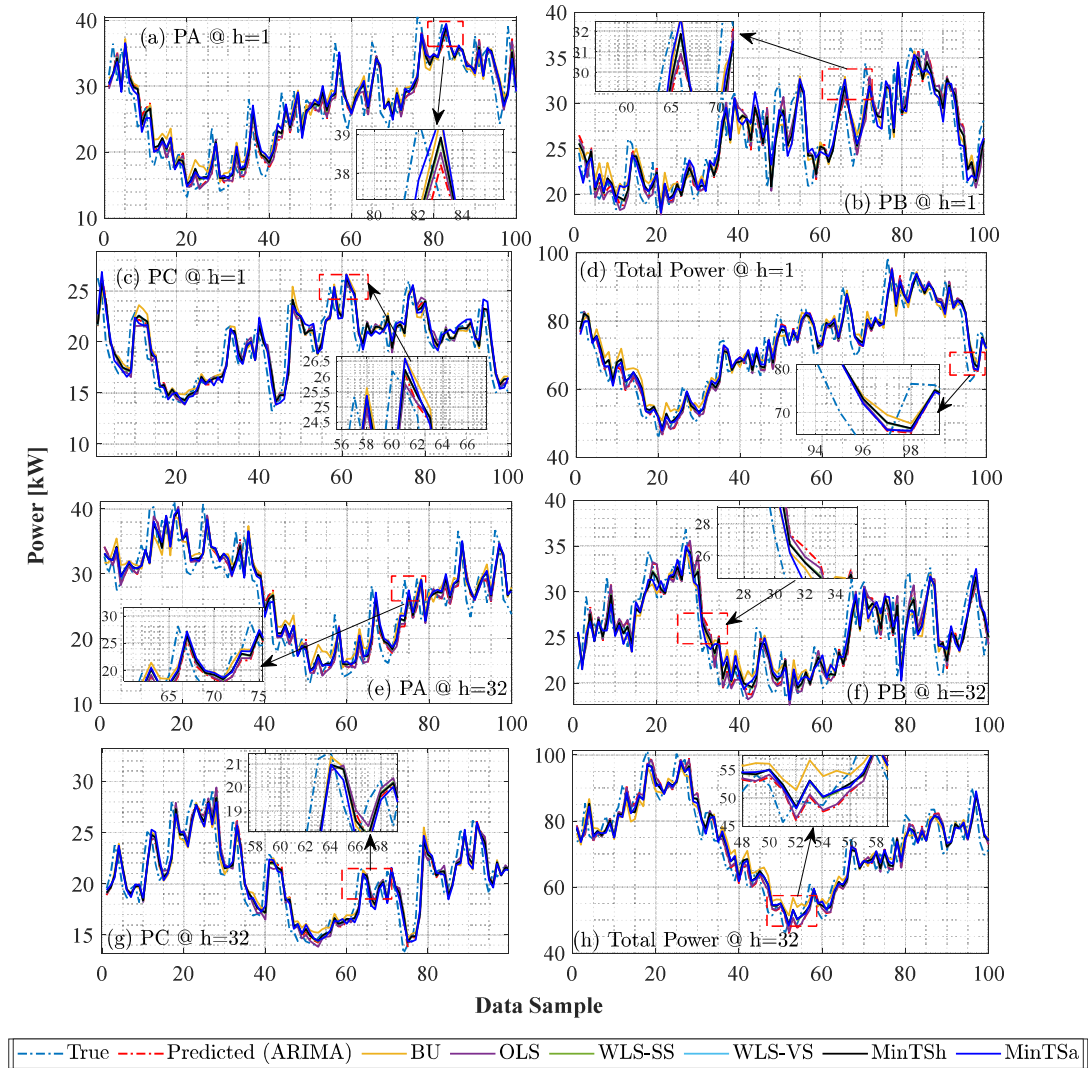
FIGURE 12. Comparison of true, predicted and reconciled forecasts – Case 4: (a) Phase A @ h=1, (b) Phase B @ h=1, (c) Phase C @ h=1, (d) Total Power @ h=1, (e) Phase A @ h=32, (f) Phase B @ h=32, (g) Phase C @ h=32, (h) Total Power @ h=32.

(a four-hierarchical-level system). The simulation is conducted for two forecast horizons, $h = 1$ and 32, and one-month historical data is used to generate base forecasts. The average accuracy improvements using the proposed method for the two forecast horizons at each hierarchical level are evaluated by RMSE and MAPE as shown in Tables 10 and 11, respectively.

In Case 5, MinTSh has the highest MAPE average accuracy improvement of 9.65% at the top level (Level 0) for $h = 1$. MinTSh shows a similar or better performance than MinTSa mainly because MinTSh has a better strategy to deal with data variance when the number of the bottom level's observations (b_t) is high. MinTSa and MinTSh have the highest accuracy improvement compared to the benchmark methods.

OLS and WLS-SS do not require parameters to estimate the covariance matrix, W_h, and only use the information contained in the hierarchy structure; while WLS-VS, MinTSa

TABLE 10. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by RMSE (Case 5).

h	1 (15 min)				32 (8 hrs.)			
Level	0	1	2	3	0	1	2	3
BU	-1.75	2.71	0.99	0.00	-1.86	2.69	0.98	0.00
OLS	1.05	0.66	-0.98	-2.65	0.78	0.46	-1.23	-2.13
WLS-SS	3.22	3.19	2.65	-0.01	2.98	3.06	2.01	-0.65
WLS-VS	2.84	4.11	3.28	0.91	2.65	4.09	2.55	-0.03
MinTSh	**6.48**	**7.52**	4.09	2.01	**5.21**	**6.98**	3.98	2.44
MinTSa	4.53	6.32	**4.68**	**4.14**	4.02	5.25	**4.07**	**3.12**

and MinTSh depend on errors in the historical data. Only MinTSa and MinTSh consider the covariance of the base forecast errors of two time series data in the hierarchy [6]; while OLS, WLS-VS, and WLS-SS are based on the assumption that the base forecast errors between two time series are stochastic. This is the main reason that the proposed

TABLE 11. Accuracy improvement (%) of the proposed method vs benchmark methods evaluated by MAPE (Case 5).

h	1 (15 min)				32 (8 hrs.)			
Level	0	1	2	3	0	1	2	3
BU	-2.66	-0.55	1.32	0.00	-3.96	-0.67	0.93	0.00
OLS	1.56	0.65	-0.12	-6.91	0.72	0.23	-0.66	-6.91
WLS-SS	3.89	3.14	3.20	-1.32	2.67	2.66	2.05	-1.32
WLS-VS	4.11	3.26	2.69	2.51	3.22	3.11	2.54	2.51
MinTSh	**9.65**	**8.22**	**5.27**	3.26	**9.22**	**7.33**	**4.99**	3.12
MinTSa	7.22	6.51	4.36	**3.28**	6.95	5.33	4.23	**3.78**

MinT-based hierarchical load forecasting performs better than the benchmark methods.

VI. CONCLUSION

In this paper, a novel transformer load forecast aggregation method through the minimum trace optimal reconciliation and AMI data is proposed, along with an independent state-of-the-art base load forecast at each hierarchical level. The proposed method aims to exploit the information contained in the hierarchical structure and improve the accuracy of traditional load forecasting practices for distribution and sub-station transformers. It is proven through several case studies that the proposed method significantly improves the load forecast accuracy and has superior performance compared to the four chosen benchmark methods (BU, OLS, WLS-VS, and WLS-SS) in simple and complex distribution networks. Although the primary purpose of this work is for transformer health monitoring, the proposed method is also applicable for distribution grid rebalancing and renewable power generation aggregation.

The recommended future work includes:

- *Incorporating the Grid Topology Information*: Integration of the grid topology information into the hierarchical load forecasting process can be done by considering spatial relationships among different grid components, such as transformers, substations and feeders, to enhance the forecasting accuracy.
- *Integrating Demand-Side Management*: Integration of demand-side management strategies into the hierarchical load forecasting model allows for load shedding, demand response, and peak shaving to be incorporated as part of the forecasting process.
- *Uncertainties Propagation in Hierarchical Forecasts*: Techniques can be developed to propagate uncertainties across different levels of the hierarchy. The uncertainty at a higher level may affect forecasts at lower levels, and accounting for these uncertainties can improve the overall reliability of load forecasting.
- *Dynamic Hierarchical Structure:* Dynamic adjustment of the hierarchical structure based on changing load characteristics or grid conditions involves automatically restructuring the hierarchy to adapt to load variations and system changes.

ACKNOWLEDGMENT

REFERENCES

[1] K. Ashok, D. Li, N. Gebraeel, and D. Divan, "Online detection of inter-turn winding faults in single-phase distribution transformers using smart meter data," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5073–5083, Nov. 2021.

[2] J. Singh, S. Singh, and A. Singh, "Distribution transformer failure modes, effects and criticality analysis (FMECA)," *Eng. Failure Anal.*, vol. 99, pp. 180–191, May 2019.

[3] R. Murugan and R. Ramasamy, "Understanding the power transformer component failures for health index-based maintenance planning in electric utilities," *Eng. Failure Anal.*, vol. 96, pp. 274–288, Feb. 2019.

[4] H. Kikusato, K. Mori, S. Yoshizawa, Y. Fujimoto, H. Asano, Y. Hayashi, A. Kawashima, S. Inagaki, and T. Suzuki, "Electric vehicle charge–discharge management for utilization of photovoltaic by coordination between home and grid energy management systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3186–3197, May 2019.

[5] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3276–3284, May 2018.

[6] Y. Zhang and J. Dong, "Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation," *IEEE Trans. Power Syst.*, early access, Aug. 31, 2018, doi: 10.1109/TPWRS.2018.2868175.

[7] R. J. Bessa, A. Trindade, C. S. P. Silva, and V. Miranda, "Probabilistic solar power forecasting in smart grids using distributed information," *Int. J. Electr. Power Energy Syst.*, vol. 72, pp. 16–23, Nov. 2015.

[8] L. Bai and P. Pinson, "Distributed reconciliation in day-ahead wind power forecasting," *Energies*, vol. 12, no. 6, p. 1112, 2019.

[9] T. Silveira Gontijo and M. Azevedo Costa, "Forecasting hierarchical time series in power generation," *Energies*, vol. 13, no. 14, p. 3722, Jul. 2020.

[10] S. L. Wickramasuriya, G. Athanasopoulos, and R. J. Hyndman, "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization," *J. Amer. Stat. Assoc.*, vol. 114, no. 526, pp. 804–819, Apr. 2019.

[11] C. Zhang and R. Li, "A novel closed-loop clustering algorithm for hierarchical load forecasting," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 432–441, Jan. 2021.

[12] J. C. López, M. J. Rider, and Q. Wu, "Parsimonious short-term load forecasting for optimal operation planning of electrical distribution systems," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1427–1437, Mar. 2019.

[13] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3906–3908, Jul. 2018.

[14] B. Goehry, Y. Goude, P. Massart, and J.-M. Poggi, "Aggregation of multi-scale experts for bottom-up load forecasting," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 1895–1904, May 2020.

[15] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, and M. T. Miller, "An efficient approach to short-term load forecasting at the distribution level," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2526–2537, Jul. 2016.

[16] N. Alemazkoor, M. Tootkaboni, R. Nateghi, and A. Louhghalam, "Smart-meter big data for load forecasting: An alternative approach to clustering," *IEEE Access*, vol. 10, pp. 8377–8387, 2022.

[17] M. Arshad, S. M. Islam, and A. Khaliq, "Fuzzy logic approach in power transformers management and decision making," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 21, no. 5, pp. 2343–2354, Oct. 2014.

[18] J. A. Jardini, H. P. Schmidt, C. M. V. Tahan, C. C. B. De Oliveira, and S. U. Ahn, "Distribution transformer loss of life evaluation: A novel approach based on daily load profiles," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 361–366, 2000.

[19] A. Tariku and G. Bekele, "Distribution transformer failure study and solution proposal in Ethiopia," in *Proc. IEEE PES/IAS PowerAfrica*, Nairobi, Kenya, Aug. 2020, pp. 1–5.

[20] M. Dong, A. B. Nassif, and B. Li, "A data-driven residential transformer overloading risk assessment method," *IEEE Trans. Power Del.*, vol. 34, no. 1, pp. 387–396, Feb. 2019.

[21] L. Zhou, L. Wang, X. Zhang, Q. Liu, L. Guo, and D. Wang, "An estimation method for real-time thermal capacity of traction transformers under unbalanced loads," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11438–11446, Nov. 2021.

[22] M. A. Awadallah, B. N. Singh, and B. Venkatesh, "Impact of EV charger load on distribution network capacity: A case study in Toronto," *Can. J. Electr. Comput. Eng.*, vol. 39, no. 4, pp. 268–273, Fall 2016.

[23] B. D. Lahoti and D. E. Flowers, "Evaluation of transformer loading above nameplate rating," *IEEE Trans. Power App. Syst*, vol. PAS-100, no. 4, pp. 1989–1998, Apr. 1981.

[24] G. Athanasopoulos, R. A. Ahmed, and R. J. Hyndman, "Hierarchical forecasts for Australian domestic tourism," *Int. J. Forecasting*, vol. 25, no. 1, pp. 146–166, Jan. 2009.

[25] R. H. Shumway and D. S. Stoffer, "ARIMA Models," in *Time Series Analysis and Its Applications*. Cham, Switzerland: Springer, 2017.

[26] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and F. Petropoulos, "Forecasting with temporal hierarchies," *Eur. J. Oper. Res.*, vol. 262, no. 1, pp. 60–74, Oct. 2017.

[27] L. Navarro-espinosa and O. Alejandro. (2015). *Dissemination Document: Low Voltage Networks Models and Low Carbon Technology Profiles*. [Online]. Available: https://www.researchgate.net/publication/283569482_Dissemination_Document_Low_Voltage_Networks_Models_and_Low_Carbon_Technology_Profiles

[28] *IEEE PES Test Feeder*. Accessed: Mar. 1, 2023. [Online]. Available: https://cmte.ieee.org/pes-testfeeders/resources/

[29] K. P. Schneider, B. A. Mather, B. C. Pal, C.-W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, May 2018.

[30] S. Panigrahi and H. S. Behera, "A hybrid ETS–ANN model for time series forecasting," *Eng. Appl. Artif. Intell.*, vol. 66, pp. 49–59, Nov. 2017.

AMAN SAMSON MOGOS (Member, IEEE) was born in Wejeba, Eritrea. He received the B.Sc. degree in electrical and electronics engineering from the Eritrea Institute of Technology, Mainefhi, Eritrea, in 2013, the M.Sc. degree in electrical engineering from the Polytechnic University of Milan (Politecnico di Milano), Milan, Italy, in 2020, the M.B.A. degree from Unicaf University/South Wales University (Online), Larnaca, Cyprus, in 2020, and the Ph.D. degree in electrical engineering from the University of Saskatchewan, Saskatoon, Canada, in 2023.

From 2013 to 2018, he worked as a Power Systems Engineer with the Eritrean Electric Corporation, Asmara, Eritrea. From 2018 to 2020, he was a Project Engineer with FRIEM S.p.A, Milan, Italy. Currently, he is working as a Research Engineer with the Saskatchewan Research Council (SRC), Saskatoon, Canada. His research interests include power systems operation, application of artificial intelligence (AI) in power systems, distribution grid modernization, and renewable energy integration.

OSAMA ASLAM ANSARI (Member, IEEE) received the bachelor's degree in electrical engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2015, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Canada, in 2017 and 2021, respectively.

His current research interests include power systems reliability, the application of artificial intelligence techniques to power systems, and electricity markets.

XIAODONG LIANG (Senior Member, IEEE) was born in Lingyuan, Liaoning, China. She received the B.Eng. and M.Eng. degrees in electrical engineering from Shenyang Polytechnic University, Shenyang, China, in 1992 and 1995, respectively, the M.Sc. degree in electrical engineering from the University of Saskatchewan, Saskatoon, Canada, in 2004, and the Ph.D. degree in electrical engineering from the University of Alberta, Edmonton, Canada, in 2013.

From 1995 to 1999, she was a Lecturer with Northeastern University, Shenyang, China. In October 2001, she joined Schlumberger (SLB), Edmonton, and was promoted to the Principal Power Systems Engineer with this world's leading oil field service company, in 2009. She worked with Schlumberger for almost 12 years, until August 2013. From 2013 to 2019, she was with Washington State University, Vancouver, WA, USA, and the Memorial University of Newfoundland, St. John's, NL, Canada, as an Assistant Professor and later an Associate Professor. In July 2019, she joined the University of Saskatchewan, Saskatoon, where she is currently a Professor and the Canada Research Chair in Technology Solutions for Energy Security in Remote, Northern, and Indigenous Communities. She was an Adjunct Professor with the Memorial University of Newfoundland from 2019 to 2022. Her current research interests include power systems, renewable energy, and electric machines.

Dr. Liang is the Registered Professional Engineer in the Province of Saskatchewan, Canada, a fellow of IET, and the Deputy Editor-in-Chief of IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS.

C. Y. CHUNG (Fellow, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, China, in 1995 and 1999, respectively.

He is currently the Head and the Chair Professor of Power Systems Engineering with the Department of Electrical Engineering, The Hong Kong Polytechnic University. His current research interests include smart grid technologies, renewable energy, power system stability/control, planning and operation, computational intelligence applications, power markets, and electric vehicle charging.

Dr. Chung is an IEEE PES Distinguished Lecturer. He was a recipient of the 2021 IEEE Canada P. Ziogas Electric Power Award. He is a Senior Editor of IEEE TRANSACTIONS ON POWER SYSTEMS, a Consulting Editor of IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, and the Vice Editor-in-Chief of *Journal of Modern Power Systems and Clean Energy*.

• • •