STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, May 2026, pp 4371–4398.
Published online in International Academic Press (www.IAPress.org)

Leveraging Hierarchical Time Series to Enhance Short - Term Electricity Consumption Forecasts: A Case Study of South Africa

Mantombi Bashe[1], Claris Shoko[2], Thakhani Ravele[3,*], Caston Sigauke[3]

[1]*Eskom Holdings SOC Ltd, South Africa*
[2]*Department of Statistics, University of Botswana*
[3]*Department of Mathematical and Computational Sciences, University of Venda, South Africa*

Abstract Accurate short-term electricity consumption forecasts are essential for maintaining grid stability and for developing effective energy policies. In this study, we propose a novel multilevel forecasting methodology that enhances forecast accuracy by carefully combining and reconciling the outputs of multiple machine learning models. Using high-frequency data from South Africa, we apply Stochastic Gradient Boosting (SGB) and XGBoost to model the consumption attributed to key energy sources: solar (PV and CSP), coal, diesel, nuclear, and wind. One of the important implications is that forecast performance varies across energy sources. The performance of XGBoost outperforms that of the SGB model for solar, coal, and diesel consumption, while the SGB model performs best for nuclear and wind energy. This split points to the importance of source-specific models. We also demonstrate that a reconciled ensemble of these models yields the most accurate and robust overall forecasts, particularly for complex, non-renewable sources. Beyond point estimates, we evaluate probabilistic prediction intervals and demonstrate that linear quantile regression provides a significant improvement over the benchmark linear regression, resulting in sharper intervals with narrower widths and better scores, without sacrificing coverage. This study fills a gap in the literature by providing a validated, scalable framework that utilises a reconciled ensemble of methods to enhance forecast quality. The results provide clear, actionable guidance for energy planners and policymakers seeking to optimise the generation mix to provide a reliable electricity supply.

Keywords Forecast combinations, hierarchical time series, electricity consumption, energy sources, renewable energy.

DOI: 10.19139/soic-2310-5070-3194

1. Introduction

1.1. Overview

Electricity generation refers to the process of generating electricity power from primary sources of energy, such as coal, wind, solar, and nuclear energy; whereas electricity consumption is the amount of electricity utilised by residential, commercial and industrial users [1]. The electricity generation sector in South Africa (SA) faces increased variability and uncertainty due to dependence on traditional (non-renewable) and renewable energy sources. The traditional energy sources are coal, nuclear and diesel, which Eskom, the major player in electricity generation in SA, mainly utilises. At the same time, renewable energy sources include wind and solar energy, as well as other energy sources from independent power producers, such as biomass, landfill gas, and small hydropower technologies. The South African economic development and quality of life depend on the stability and efficiency of its electricity generation systems. The energy sector faces significant challenges, including fluctuating energy demand, integrating renewable energy sources, and managing loads efficiently. Short-term forecasts are

*Correspondence to: Thakhani Ravele (Email: thakhani.ravele@univen.ac.za). Department of Mathematical and Computational Sciences, University of Venda, South Africa.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

crucial in the electricity consumption sector for maintaining supply and demand balance, optimising power plant performance, and minimising outage risks.

Eskom supplies the electricity used to meet the consumption needs (86%), IPPs (10%) and imports (4%) [2]. Total electricity supplied by both Eskom and IPPs excluding pumping and wheeling come from the following energy sources: coal (80%), wind (5.6%), hydro (4.5%), nuclear (3.9%), solar (3.1%), diesel (2.5%), and other (0.5%) [2].

Accurate short-term forecasting alone is insufficient in a system with multiple abstraction levels, which requires hierarchical forecasting across all layers and horizons. Therefore, forecasts must be coherent across all dimensional hierarchies for optimal decision-making. Hierarchical forecasting involves coherent forecasts within a set hierarchical structure [3].

Hierarchical forecasting requires models that capture dependencies and interactions to ensure accurate forecasts are consistent across levels. The dependencies reflect how the national electricity consumption forecast depends on the sum of the forecasts of electricity consumption per energy sources utilised to generate electricity. For example, the country-wide electricity consumption forecast is based on the collective forecasts of individual energy sources, so the aggregate electricity consumption forecast is subject to the forecast of each energy source. On the other hand, the interactions describe how changes at one level in the hierarchy affect the forecasts at other levels. For example, a change in national electricity consumption can affect individual energy source's forecasts, requiring adjustments for factors like maintenance issues to ensure adequate electricity generation.

This study focuses on electricity consumption. It aims to construct a reliable and efficient short-term forecasting modelling framework that will guide the decision-makers of the national power grid operators. Forecasting electrical load at various scales, from individual consumers to the aggregate level, is an important component in many smart grid applications [4].

Table 1. Summary of different publications that discuss hierarchical forecasting.

Authors	**Brief summary of contribution**
Silveira & Gontijo [7]	Applied hierarchical forecasting to Brazilian hourly power generation data grouped by electrical subsystems. Compared ARIMA and ETS models using bottom-up, top-down, and optimal reconciliation approaches (OLS, WLS, MinT). Found MinT models performed best.
Hyndman et al. [6]	Compared bottom-up, top-down, and optimal combination approaches for forecasting Australian domestic tourism demand. Found optimal combination outperformed bottom-up and top-down approaches.
Hansen et al. [5]	Applied three-level spatial hierarchical forecasting to Western Denmark's onshore wind turbine production for day-ahead markets using hourly settlement data. Showed spatial hierarchical forecast improved commercial forecasts.
Wen et al. [9]	Applied reconciliation methods (including bottom-up) to average profits of four wind power producers in electricity markets using Danish transmission system operator's two-year hourly price data. Found reconciled average profits outperformed individual average profits.
Rombouts et al. [11]	Utilized random forest, XGBoost and LightGBM reconciliation methods for delivery demand in London, Manchester and New York. Compared against five state-of-the-art linear reconciliation approaches. Found ML-based forecast reconciliation improved accuracy over linear methods.
Kafa et al. [12]	Used ETS, ARIMA and LightGBM to predict monthly mail flow for a French postal organization. Compared top-down, bottom-up and optimal combination approaches. Found optimal combination-based ETS was most accurate, while top-down-based LightGBM was best for social KPIs.
Athanasopoulos et al. [8]	Provided a detailed discussion and review of forecast reconciliation methods.

1.2. Hierarchical forecasting: Literature review

Table 1 provides a summative overview for a better understanding of different publications that will be used as part of the literature review for this research work.

In a hierarchy, information is organised into different levels, and each level represents an aggregation of the information from a level below [5]. Hierarchical time series is where multiple time series can be hierarchically organised and aggregated in several different groups based on products, geography or other features [6]. There are different approaches for forecasting hierarchical time series. However, this section only discusses the ones relevant to this study.

In [7], hierarchical forecasting was applied to Brazilian hourly power generation data. The objective of [7] was to measure the accuracy of the main measures of aggregating and disaggregating the forecasts using the Autoregressive Integrated Moving Average (ARIMA) and Error, Trend, Seasonal (ETS) models. The data in [7] was grouped by electrical subsystems. In [7], bottom-up (BU), top-down (TD) and optimal reconciliation approaches, ordinary least squares (OLS), weighted least squares (WLS) and minimum trace (MinT) were utilised to aggregate and dis-aggregate predictions made for grouped time series. In BU approaches, the forecast is performed at the bottom level and then aggregated up [8, 10]. In the TD approach, the forecast is based on the aggregate series and is disaggregated based on historical proportions [10, 16]. In [17], the reconciliation approach is defined as the one that ensures that the hierarchical time series forecast adheres to aggregation constraints. However, in [5], reconciliation is defined as adjusting or combining forecasts to ensure that forecasts from different levels of the hierarchy are consistent with each other across all levels of the hierarchy. The reconciliation methods ensure that the reconciled forecasts are unbiased if the base forecasts are unbiased [18].

The result in [7] showed that MinT models performed better than OLS and WLS. The hierarchical BU and TD forecasting approaches were also utilised in [6] and compared with the optimal combination approach, where the forecasting data was based on Australian domestic tourism demand. The results showed that the optimal combination outperforms the BU and TD approaches. In [5], a three-level spatial hierarchical forecast was applied to Western Denmark's onshore wind turbine production. The forecast in [5] was intended for a day-ahead market, where hourly settlement data was used. The results showed that the spatial hierarchical forecast improved the commercial forecasts.

Recently, [9] applied different reconciliation methods, including the BU, to the average profits of four wind power producers (WPPs) who participated in electricity markets. The data utilised in [9] was a two-year hourly price data from the Danish transmission system operator. The reconciliation average profits outperformed the individual average profits for all WPPs.

In certain situations, the demand values are volatile and correlated over time; conventional parametric forecasting techniques, which include ARIMA and other models, do not yield good forecasting results [10]. The solution to these challenges is a class of algorithms called universal approximators, some of them are Multi-Layer perceptron (MLP), Random Forest (RF), Gradient Boosting (GB) and eXtreme Gradient Boosting (XGBoost), which are based on machine learning techniques [10].

In the recent past [11] utilised random forest, XGBoost and light gradient boosting machine (LightGBM) reconciliation methods and compared these reconciliation methods to a base forecast, which was based on preferred forecast models and further benchmarked by five state-of-the-art linear reconciliation approaches. The data used in [11] was the demand for deliveries in London, Manchester and New York. The results showed that the ML-based forecast reconciliation results improved forecast accuracy compared to existing linear reconciliation methods.

In a recent study [12] utilised ETS, ARIMA and LightGBM to predict monthly mail flow for a major French postal organisation. [12], computed the hierarchical forecasts based on the forecast combinations using the main approaches: TD, BU and optimal combination (OC). The results showed that the OC-based ETS method was the best in terms of accuracy, while the superior method was TD-based LightGBM for social KPIs. A detailed discussion and review of forecast reconciliation methods is presented in [8].

This study utilises the XGBoost and the SGB for hierarchical time series forecasting, benchmarked with the traditional Exponential Smoothing. The XGBoost and SGB are powerful machine learning algorithms, similar to LightGBM and CatGBM, or other deep learning approaches for hierarchical forecasting, due to several factors that include high predictive accuracy and the ability to capture complex, nonlinear patterns [13]. The regularisation

techniques (L1 (Lasso) and L2 (Ridge)) for the XGBoost model help prevent overfitting [14], handle missing data internally, and result in a more robust model. Although LightGBM is generally faster in terms of computational efficiency and scalability, XGBoost offers a balance of speed and performance, particularly with optimised implementations and parallel processing capabilities [15]. The XGBoost and the SGBM algorithms are tree-based, thus offering better interpretability compared to deep learning models. This allows for feature importance analysis, giving an understanding of which factors drive energy consumption predictions. While models like LightGBM and CatBoost offer advantages in specific scenarios (e.g., speed for very large datasets, handling categorical features), XGBoost and SGB often provide a compelling combination of accuracy, robustness, interpretability, and ease of use for many energy forecasting problems with structured data.

1.3. Contributions and research highlights

This study investigates whether the XGBoost approach outperforms the Stochastic Gradient Boosting (SGB) framework in predicting electricity consumption generated from from non-renewable and renewable energy sources. It includes a comprehensive comparison of various prediction measures across several energy sources. The study also examines the prediction interval estimation using two regression models: linear regression and linear quantile regression.

The key findings of the study are:

- Among all the non-renewable energy sources, the XGBoost technique demonstrated higher capability in forecasting electricity consumption generated from coal and diesel energy sources. In contrast, the SGB technique was more capable of forecasting electricty consumption generated from nuclear energy source.
- The research also contrasted the forecasts produced by ETS models and established that the optimal combination reconciliation method performed better than the bottom-up method for overall and non-renewable energy sources. However, the XGBoost method outperformed these two traditional methods.
- Linear quantile regression outperforms linear regression in accuracy (coverage probability) and precision (interval score and average width), although linear regression achieves a higher coverage probability.
- The linear quantile regression technique preserves reasonable coverage probabilities, but provides smaller interval scores and narrower mean widths, enhancing its performance for intentional predictions.
- Total non-renewable energy prediction intervals are much wider than total renewable energy prediction intervals because they have a higher interval score, average width, and prediction interval normalised average width (PINAW), indicating lower predictability in forecasts for non-renewable energy sources.

The rest of the paper is organised as follows: Section 2 presents the models, and the empirical results are presented and discussed in Section 3. Lastly, Section 4 concludes.

2. Models

The modelling framework is given in Figure 1. This flowchart explains a detailed four-step forecasting process. In phase 1, historical data is gathered, cleaned, and separated for model training and testing. In Phase 2, the XGBoost and Stochastic Gradient Boosting machine learning models are utilised for the first contribution modelling. In Phase 3, forecasting reconciliation and hierarchy level forecasting consistency address the combination of the bottom-up approaches. In Phase 4, the finalisation of forecasting assesses performance using the MAE, MAPE, and MRAE metrics, and includes the quantification of associated uncertainty.

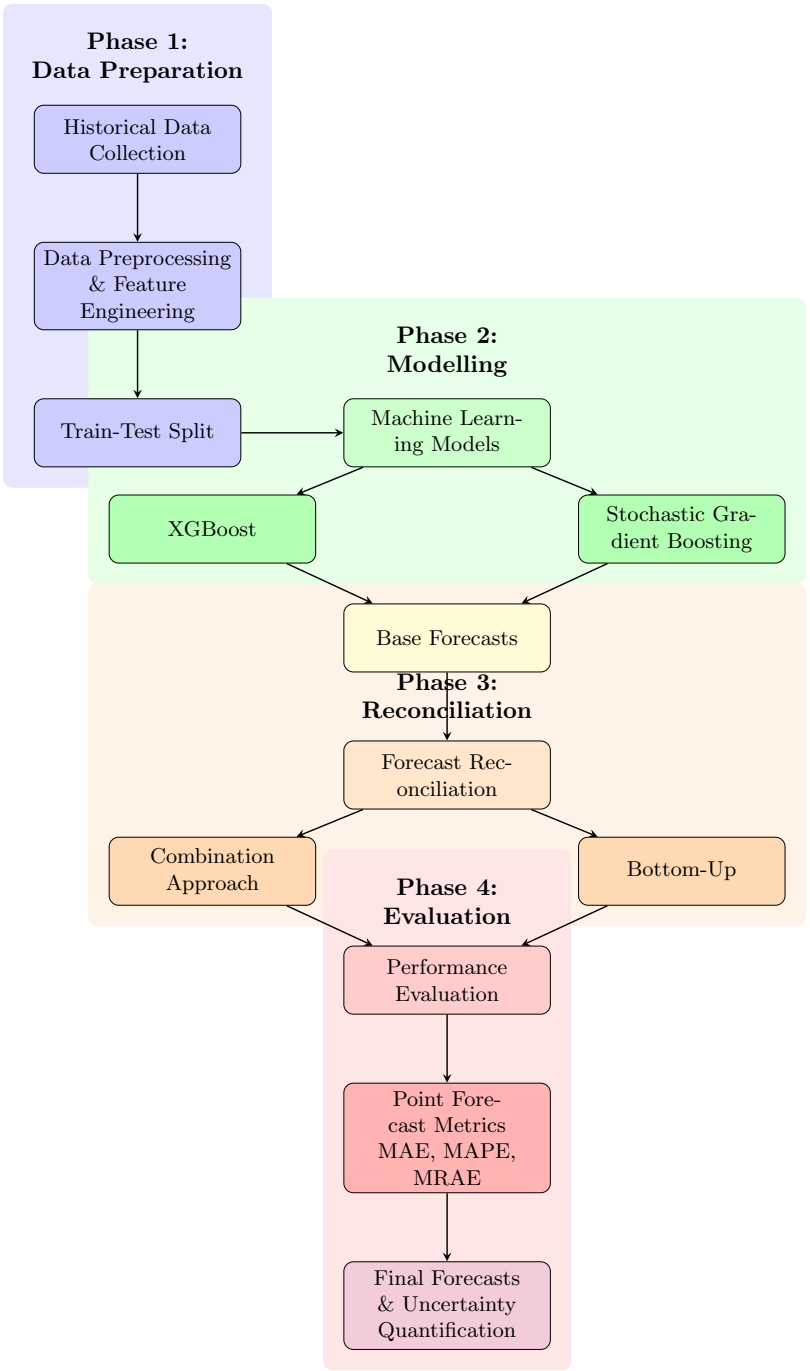


Figure 1. Forecasting modelling flowchart.

2.1. Hierarchical structure

A hierarchical time series is an n-dimensional multivariate series arranged to satisfy certain linear constraints. Let $Y_t \in \mathbb{R}^n$ be the vector of observations of all time series in the hierarchy at time t, and let $b_t \in \mathbb{R}^{n_b}$ be the vector of observations of the most disaggregated time series at time t. The entire hierarchy at time t is represented by:

$$\boldsymbol{Y_t} = \boldsymbol{S b_t}, \text{ for } t = 1, 2, ..., T,$$

where T is the length of the time series and $\boldsymbol{S}$ is the summing matrix that defines the aggregation constraints.

$$\boldsymbol{Y_t} = \begin{bmatrix} Y_t \\ Y_{R,t} \\ Y_{N,t} \\ Y_{RW,t} \\ Y_{RP,t} \\ Y_{RC,t} \\ Y_{RO,t} \\ Y_{NG,t} \\ Y_{NT,t} \\ Y_{NN,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_{RW,t} \\ Y_{RP,t} \\ Y_{RC,t} \\ Y_{RO,t} \\ Y_{NG,t} \\ Y_{NT,t} \\ Y_{NN,t} \end{bmatrix} = \boldsymbol{S b_t},$$

where $\boldsymbol{Y_t}$ is the total electricity consumption, $Y_{R,t}$ total electricity consumption generated from renewable sources, $Y_{N,t}$ total electricity consumption generated from non-renewable sources, $Y_{RW,t}$ electricity consumption generated from wind, $Y_{RP,t}$ electricity consumption generated from photovoltaic (PV), $Y_{RC,t}$ electricity consumption generated from concentrated solar power (CSP), $Y_{RO,t}$ electricity consumption generated from other renewable energy sources, $Y_{NG,t}$ electricity consumption generated from diesel, $Y_{NT,t}$ electricity consumption generated from coal and $Y_{NN,t}$ is electricity consumption generated from nuclear.

The hierarchical structure of the electricity consumption generated from different energy sources is shown in Figure 2.

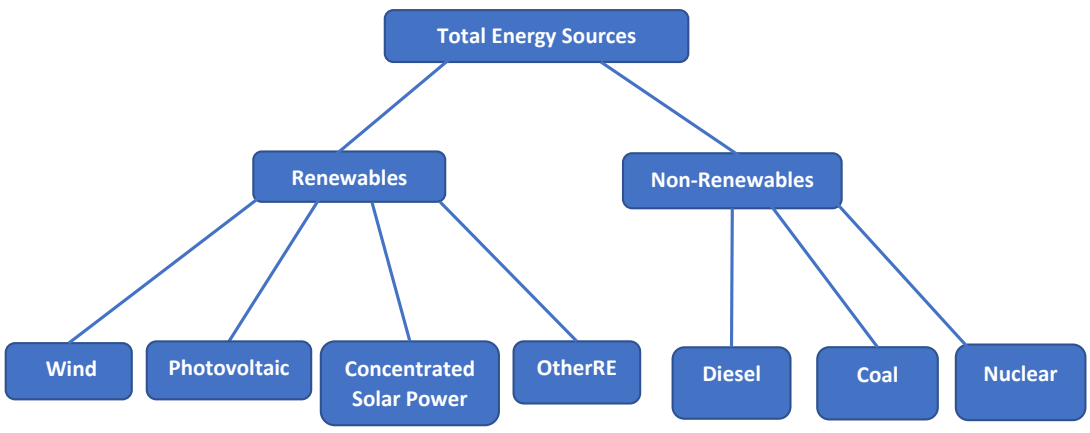


Figure 2. Hierarchical structure of the different energy sources. Data source: https://www.eskom.co.za/dataportal/.

In Figure 2 the total energy sources are made up of renewables and non-renewable energy sources utilised in SA. Then the renewables consists of a mixture of energy sources (wind and otherRE) and technologies (photovoltaic and concentrated solar power). However, in this study both the renewable energy sources and technologies are loosely referred to as renewable energy sources. The non-renewables consists of coal, nuclear and diesel.

2.2. Machine learning algorithms for forecasting

2.2.1. Extreme Gradient Boosting (XGBoost) The XGBoost is a powerful machine learning algorithm known for its high performance in various predictive tasks [19], and as discussed briefly in section 1.2. XGBoost is an implementation of gradient-boosted decision trees designed for speed and performance. The core objective of XGBoost is to minimise a regularised objective function that combines a convex loss function and a penalty term to prevent overfitting. The XGBoost is formulated as shown in Equation (1):

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F}, \tag{1}$$

where $\hat{y}_i$ is the predicted value for instance i, K is the number of trees, f_k is a function in the set of regression trees $\mathcal{F}$, x_i represents the feature vector for instance i.

The objective function to be minimised is in Equation (2):

$$\mathcal{L} = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{2}$$

where l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i, Ω is a regularisation term that penalises the complexity of the model.

The regularisation term $\Omega(f_k)$ is defined in Equation (3):

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{3}$$

where T is the number of leaves in the tree, w_j is the weight of leaf j, γ and λ are regularisation parameters.

2.2.2. Stochastic gradient boosting The GB can be identified as a machine learning technique generally applied to classification and regression problems [21]. One of the main disadvantages of the GB is its greedy nature, which leads to overfitting on training data [22, 23]. One of the most popular variants, stochastic gradient boosting (SGB), alleviates this problem by randomly selecting, without replacement, a subset of the training data at each iteration [20, 22]. The general formulation of the SGB is given by Equation (4):

$$F(x) = \sum_{m=1}^{M} \beta_m h(x; \gamma_m), \tag{4}$$

where $h(x; \gamma_m) \in \mathbb{R}$ are functions of x with parameters γ_m and β_m that limit over fitting [22, 23].

2.3. Forecast Reconciliation Methods

This section discusses the forecast reconciliation methods used in this study.

2.3.1. Combination approach According to [24], the optimal combination produces forecasts for all series across all hierarchical levels. Then it combines these forecasts with a linear model to obtain reconciled forecasts. The procedure is outlined as follows [8]: Suppose that the h-step-ahead reconciled forecasts are depicted by equation (5).

Table 2. Model comparisons.

Models	Strengths	Weaknesses
M1 (XGBoost) [19]	1. Is known for its high predictive accuracy. 2. Manages missing data effectively. 3. Provides insight into feature importance. 4. Includes L1 and L2 regularization to prevent overfitting.	1. Computationally expensive. 2. Requires careful tuning of hyperparameters. 3. Less interpretable than simpler models.
M2 (SGB) [22, 20, 23]	1. Can model temporal dependencies by feature engineering with lags. 2. Reduces overfitting risk using ensemble methods. 3. Provides measures for feature importance.	1. Sensitive to hyperparameters like learning rate and number of estimators. 2. Training is time-consuming due to hyperparameter exploration. 3. Handles missing values poorly in most implementations.

$$\hat{\mathbf{Y}}_n(h) = \mathbf{SG}\hat{\mathbf{Y}}_n(h), \tag{5}$$

Then, the covariance of the errors of these forecasts is given by equation (6).

$$\mathbf{V}_h = Var\left[y_{n+h} - \hat{\mathbf{Y}}_n(h)\right] = \mathbf{SGW}_h\mathbf{G}'\mathbf{S}', \tag{6}$$

where $\mathbf{W}_h$ is the variance-covariance matrix of the h-step-ahead base forecast errors [26]. The matrix $\mathbf{G}$ minimises the trace of $\mathbf{V}_h$ such that it generates unbiased reconciled forecasts, that is, **SGS=S** given by equation (7).

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^*\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^*, \tag{7}$$

where $\mathbf{W}_h^*$ is the generalised inverse of $\mathbf{W}_h$.

2.3.2. Bottom Up The Bottom-up (BU) forecasts model is given in equation (8).

$$\hat{\mathbf{Y}}_n(h) = \mathbf{SG}\hat{\mathbf{Y}}_n(h) \tag{8}$$

The BU forecasts are obtained using $\mathbf{G} = [\mathbf{O}|\mathbf{I}]$, where $\mathbf{O}$ is a null matrix and $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{G}$ extracts only bottom-level forecasts from $\hat{Y}_n(h)$ and $\mathbf{S}$ adds up to give the BU forecasts.

In essence, a Bottom-Up method creates trust from the ground up, demanding flawless item-by-item alignment before the ultimate totals can be believed. By using the overall numbers as inputs and employing sophisticated techniques to determine the most effective set of modifications to make those totals match, the Optimal Combination methodology operates top-down without requiring going over each transaction in detail. The primary practical differences between the BU and OC are summarised in Table 3 below:

2.4. Performance evaluation metrics

To assess the precision of the model, the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Relative Absolute Error (MRAE) are utilised since they are widely recommended in the statistics field. However, it is worth noting that the proposed algorithm does not depend on the metrics; in other words, the algorithm can use either MAE, MAPE, Mean Absolute Scaled Error (MASE) or any other performance metric to choose the best ML forecasting technique and then move forward to the next phase. This is the beauty of the proposed framework, in that it can run with different ML methods and metrics.

2.4.1. Prediction interval evaluation metrics Prediction intervals (PIs) are the key instruments utilised in probabilistic prediction, and these are a set of potential values within which the predicted result exists based on a specified degree of confidence [25]. The PIs convey the intrinsic uncertainty inherent in the prediction model due

Table 3. Comparison of Bottom-Up Reconciliation and Optimal Combination Reconciliation Methods

Feature	Bottom-Up Reconciliation	Optimal Combination Reconciliation Method
Starts with	Individual, detailed transactions.	Overall, aggregated balances.
Focus	Matching every single line item.	Finding the single best combination of high-level adjustments.
Goal	Achieve balance by proving every granular piece matches.	Achieve balance by adjusting top-level figures until they align optimally.
Complexity	High detail required, can be time-consuming.	Less detail needed initially, often faster.
Best For	High-volume operational systems (e.g., bank statements, inventory counts).	Strategic financial consolidation or large-scale reporting.

to variability in data and the complexity of the systems being investigated. The PIs can measure the uncertainty for better decision and risk management systems.

In this study the following measures are used: coverage probability (CP), interval score (IS), average width (AW) and prediction interval normalised average width (PINAW).

Coverage probability The CP is the proportion of observed values that fall within the prediction intervals. CP is calculated using equation (9).

$$\text{Coverage} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_i \in [L_i, U_i]), \tag{9}$$

where y_i is the observed value, L_i and U_i are the lower and upper bounds of the prediction interval, and $\mathbb{I}$ is the indicator function.

Interval score The IS is a scoring rule that evaluates both the coverage and the width of the prediction intervals and is given by equation (10)

$$\begin{aligned}\text{IS} = \frac{1}{n}\sum_{i=1}^{n}\Bigg((U_i - L_i) + \frac{2}{\alpha}(L_i - y_i)\,\mathbb{I}(y_i < L_i) \\ + \frac{2}{\alpha}(y_i - U_i)\,\mathbb{I}(y_i > U_i)\Bigg).\end{aligned} \tag{10}$$

where α is the level of significance, e.g., 0.01 for the 99% intervals, smaller score values indicate more desirable prediction intervals. The scoring method penalises extremely wide intervals and those not covering the observed values.

Average width The average width of the prediction intervals across all forecasts is given by equation (11).

$$\text{Average Width} = \frac{1}{n}\sum_{i=1}^{n}(U_i - L_i) \tag{11}$$

Narrower intervals are preferred, but should not be so narrow that they fail to achieve the desired coverage probability.

Algorithm 1 Energy Forecasting with XGBoost and Bottom-Up Reconciliation.

Input: Time series data $D = \{PV, CSP, Wind, OtherRE, Thermal, Nuclear, OCGT\}$
Output: Reconciled forecasts R
procedure MAIN(D)
    $data \leftarrow$ DataFrame$(PV, CSP, Wind, OtherRE, Thermal, Nuclear, OCGT)$
    $data$.Total_energy $\leftarrow \sum_{i=1}^{7} data_i$ ▷ Calculate top-level total
    Split data into training and test sets:
    $train\_data \leftarrow data[1:39456]$
    $test\_data \leftarrow data[39457:40872]$
    $base\_forecasts \leftarrow train\_data[:, 1:7]$ ▷ Base forecasts
    $model \leftarrow$ TrainXGBoostModel$(train_data, base_forecasts)$
    $reconciled\_forecasts \leftarrow$ ReconcileForecasts$(model, test_data)$
    Evaluate forecasts using MASE, MAE, RMSE, Bias metrics
    Visualize results
    return $reconciled\_forecasts$
end procedure
procedure TRAINXGBOOSTMODEL($train\_data, base\_forecasts$)
    $X \leftarrow$ matrix$(base_forecasts)$
    $y \leftarrow train\_data$.Total_energy
    $dtrain \leftarrow$ xgb.DMatrix$(X,$ label $= y)$
    Set XGBoost parameters:
    booster $\leftarrow$ "gbtree"
    eta $\leftarrow 0.1$
    max_depth $\leftarrow 10$
    subsample $\leftarrow 0.8$
    colsample_bytree $\leftarrow 0.85$
    lambda $\leftarrow 1$
    alpha $\leftarrow 0$
    objective $\leftarrow$ "reg:squarederror"
    $model \leftarrow$ xgb.train(params, $dtrain$, nrounds $= 100)$
    return $model$
end procedure
procedure RECONCILEFORECASTS($model, test\_data$)
    $X\_test \leftarrow$ matrix$(test_data[:, 1:7])$
    Predict top level using XGBoost:
    $top\_level\_pred \leftarrow$ predict$(model, X_test)$
    $bottom\_level\_total \leftarrow \sum_{i=1}^{7} X\_test_i$
    $adjustment\_factor \leftarrow top\_level\_pred / bottom\_level\_total$
    Apply bottom-up reconciliation:
    $reconciled \leftarrow X\_test \times adjustment\_factor$ ▷ Element-wise multiplication
    return $reconciled$
end procedure
procedure EVALUATEFORECASTS($actual, forecast$)
    MASE $\leftarrow \frac{1}{n} \sum_{t=1}^{n} \frac{|forecast_t - actual_t|}{\frac{1}{n-1} \sum_{i=2}^{n} |actual_i - actual_{i-1}|}$
    MAE $\leftarrow \frac{1}{n} \sum_{t=1}^{n} |forecast_t - actual_t|$
    RMSE $\leftarrow \sqrt{\frac{1}{n} \sum_{t=1}^{n} (forecast_t - actual_t)^2}$
    Bias $\leftarrow \frac{1}{n} \sum_{t=1}^{n} (forecast_t - actual_t)$
    return {MASE, MAE, RMSE, Bias}
end procedure

Prediction interval normalised average width The average width of the prediction intervals normalised by the range of the observed data is given by equation (12).

$$\text{PINAW} = \frac{\frac{1}{n}\sum_{i=1}^{n}(U_i - L_i)}{\max(y) - \min(y)} \tag{12}$$

PINAW provides a normalised measure of the prediction intervals' width relative to the observed data's variability.

2.5. Implementation Details

All the analyses was done using the R programming language version 4.4.1. Package xgboost version 1.7.8.1 was used for the Extreme Gradient Boosting model, whereas package 'gbm version 2.2.2' was used for the SGB method. The hierarchical time series analysis was done using package 'hts' version 6.0.3 with the help of package 'thief' version 0.3 for reconciling temporal hierarchical forecasting. Algorithm 1 shows a detailed outline of energy forecasting with XGBoost and Bottom-Up reconciliation. A similar algorithm using the SGB method is given in algorithm 2 in the appendix.

3. Empirical results and Discussions

3.1. Data

The data used in this study is a sample from both Eskom, South Africa's power utility company and Independent Power Producers (IPPs) energy sources and can be accessed from https://www.eskom.co.za/dataportal/ (accessed on 25 April 2024) upon submission of a request form to Eskom. The data energy sources utilised are coal, nuclear and diesel, solar, wind, and other renewables. All the data is measured hourly from 1 April 2019 to 28 November 2023. When the power plants are off, the data consumed by the power plants are excluded.

The covariates used are hours of the day (hour), days of the week (day), nonlinear trend variable (no trend), which was produced by fitting the cubic regression spline model, and four differenced lagged data for each of the energy sources at lags 1, 2, 12 and 24, respectively. The data is split as follows: 95% of the data is used for training the models, and 5% is reserved for testing. For each training instance, the lagged features were computed only from prior data within the training set. The covariates were used in the XGBoost and SGB models.

Mathematical definition of lag features

A lag feature $L_k(y_t)$ is mathematically defined as:

$$L_k(y_t) = y_{t-k},$$

where:

- y_t is the value of the time series at time t
- k is the lag order (the number of time steps to look back)
- For the hourly data used in the study:
  - Lag1 is the value from 1 hour ago.
  - Lag2 is the value from 2 hours ago.
  - Lag12 is the value from 12 hours ago.
  - Lag24 is the value from 24 hours ago (previous day same hour).

These chosen lags (1, 2, 12, and 24) for hourly energy are both physically and statistically valid. Lags 1 to 2 represent short-term inertia in the weather and other influential meteorological factors. The 24-hour lag is fundamental for the daily cycles of solar generation and human activity. Lag 12 reflects half-day patterns such as solar's 12-hour symmetry or dual load peaks. Statistically, these lags consistently reflect the placement of

significant peaks within both the autocorrelation and spectral analyses, confirming that they capture the main cyclical behaviour within the data and are not arbitrarily selected.

Figure 3 shows the time series graphs for electricity consumption generated from the seven energy sources, including a graph of the total electricity consumption for the sampling period April 2019 to November 2023. Figure 3 (a), (c) and (d) show an increase in electricity consumption generated from PV, Wind and other renewable energy sources, respectively. However, over the sampling period, there is a decrease in the total electricity consumption generated from the coal energy sources.

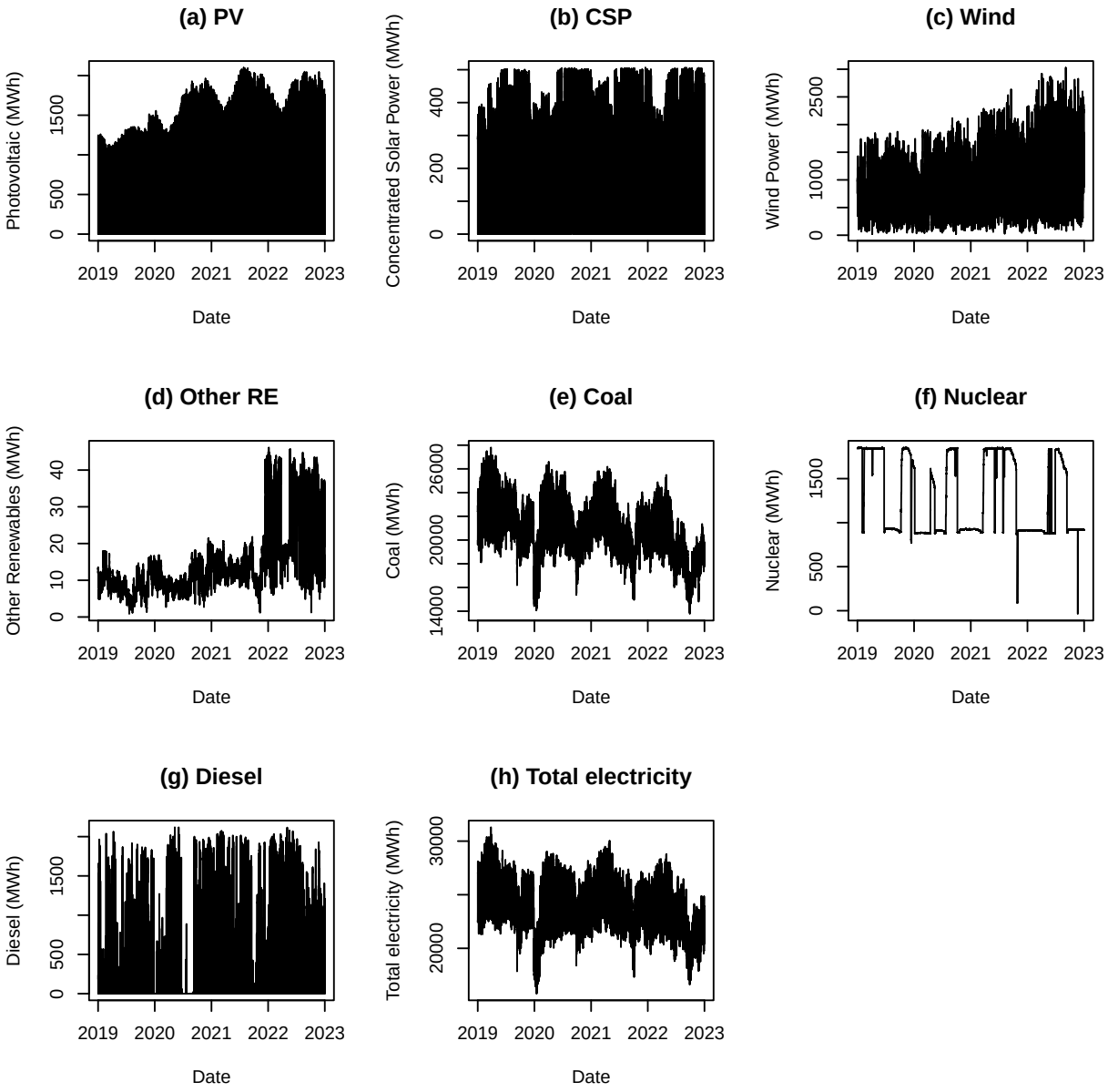


Figure 3. Time series graphs of energy consumption from the seven energy sources, with Total electricity representing the total energy consumption of all the seven energy sources.

3.2. Exploratory data analysis

In order to summarise and give a meaningful description of the key characteristics of the data, Table 4 presents the five-number summary statistics together with the mean for the disaggregated data that represents the bottom level of the hierarchy. We present the Box and Whiskers plots and histograms for each bottom-level variable. The Box and Whiskers plots help to understand the underlying distribution, whether normally distributed or not. This is presented in Figure 4.

Table 4. Descriptive Statistics of electricity consumption generated from different energy sources.

Statistic	PV	CSP	Wind	OtherRE	Coal	Nuclear	Diesel	Total
Min.	0.00	0.0	19.8	0.849	13774	64	0.0	15769
1st Qu.	0.00	0.0	568.6	9.591	19216	911	0.0	22363
Median	20.67	123.5	903.3	13.627	20698	923	0.0	23905
Mean	512.81	172.8	982.5	17.544	20693	1238	252.6	23869
3rd Qu.	1098.88	326.8	1306.5	20.210	22230	1828	356.8	25515
Max.	2099.49	506.2	3102.2	46.153	27807	1854	2120.0	31297

Key: OtherRE:electricity consumption generated from other renewable energy sources.
Total: represents the total electricity consumption generated from all the renewable and non-renewable energy sources.

Results from Table 4 for all renewables, PV, CSP, Wind and OtherRE, the median value is less than the mean value. This means that the electricity consumption generated from renewables is positively skewed. Similarly, electricity consumption generated from nuclear and diesel is positively skewed. However, for the electricity consumption from coal, the mean and median are slightly the same, which means that this electricity consumption is somewhat normally distributed. However, overall, the total electricity consumption is approximately normal, with a median slightly higher than the mean.

The results of Figure 4 show the box and whiskers plot which confirm the findings from Table 4.

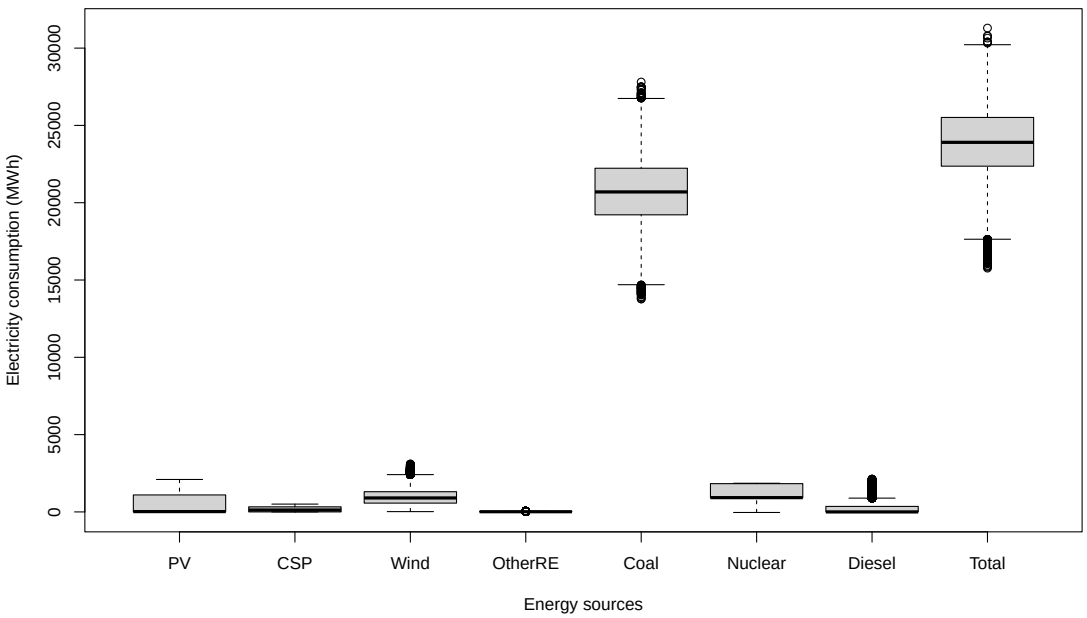


Figure 4. Box and whiskers of electricity consumption generated from different energy sources.

Figure 5 shows the distribution of electricity consumption generated from different energy sources. The coal and total electricity distributions are approximately symmetric, suggesting that most of the total consumption is met by coal energy source. The rest show a departure from the Gaussian distribution except the distribution from other renewables, suggesting a tendency towards unusually high or low electricity consumption periods. The distribution from other renewables is bimodal, signifying the presence of distinct consumption regimes or periods of transition. The histograms can help identify outliers, representing exceptional events that warrant further investigation.

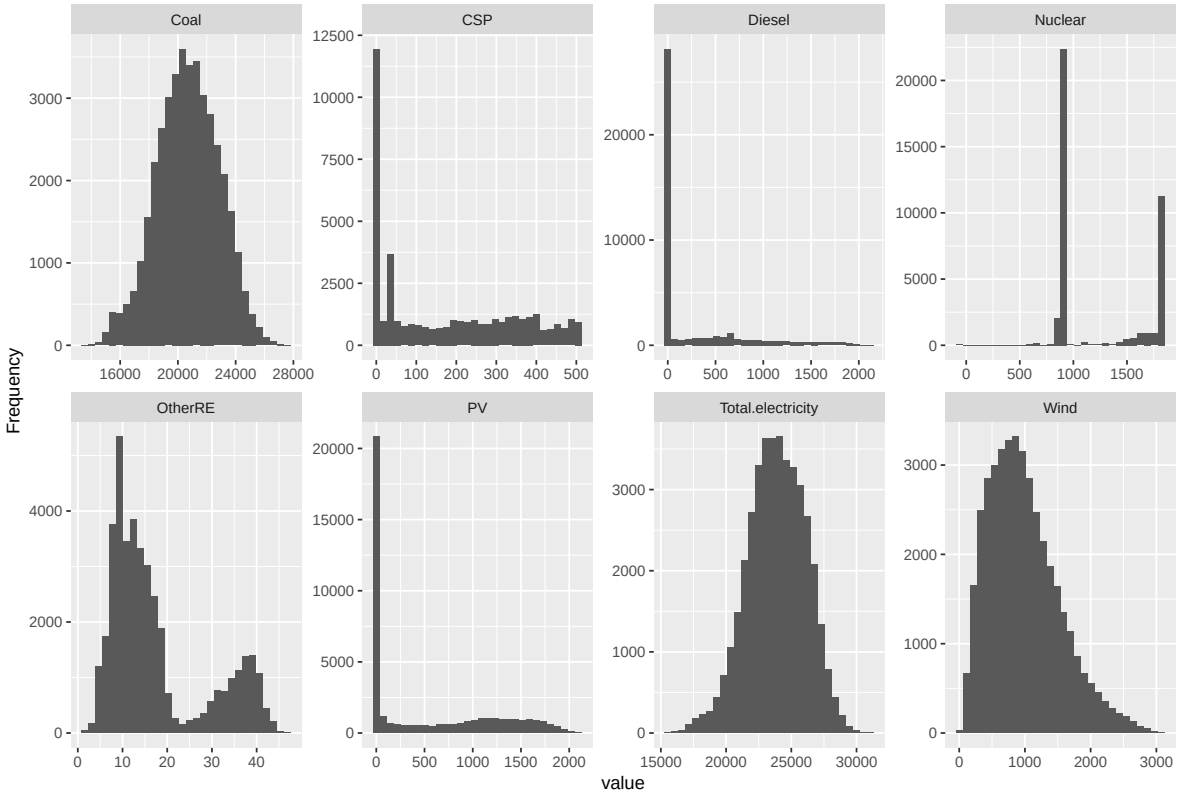


Figure 5. Histograms of different energy sources.

Figure 6 shows box and whiskers plot of the hourly distribution for electricity consumption from different energy sources from 2019 to 2023. The electricity consumption from most energy sources during the daytime hours is characterised by high median values, with a decline observed during the 16th hour. The electricity consumption generated from nuclear and other renewable energy sources is constant throughout. Extremely high values characterise electricity consumption generated from wind and diesel. This is shown by the outliers that are captured on these datasets. Most hourly box plots of electricity consumption generated from the different energy sources have long whiskers and boxes, indicating high variability in the distribution of hourly energy sources. Electricity consumption generated from PV shows high variability during the day, the one from CSP shows high variability from the 7th to the 21st hour, whilst the one from wind has high variability throughout. Electricity consumption from other renewable has high variability throughout, and so does electricity consumption from coal and nuclear energy.

Before training and validation of the models, we created lag variables for each disaggregated time series to improve the performance of our models. In addition to the lag variables, a smoothed time series was also fitted for each energy source. The smoothing parameter, λ, for each energy source, is presented in Table 5. The lag variables and the smoothed time series were incorporated in each of the fitted models as features.

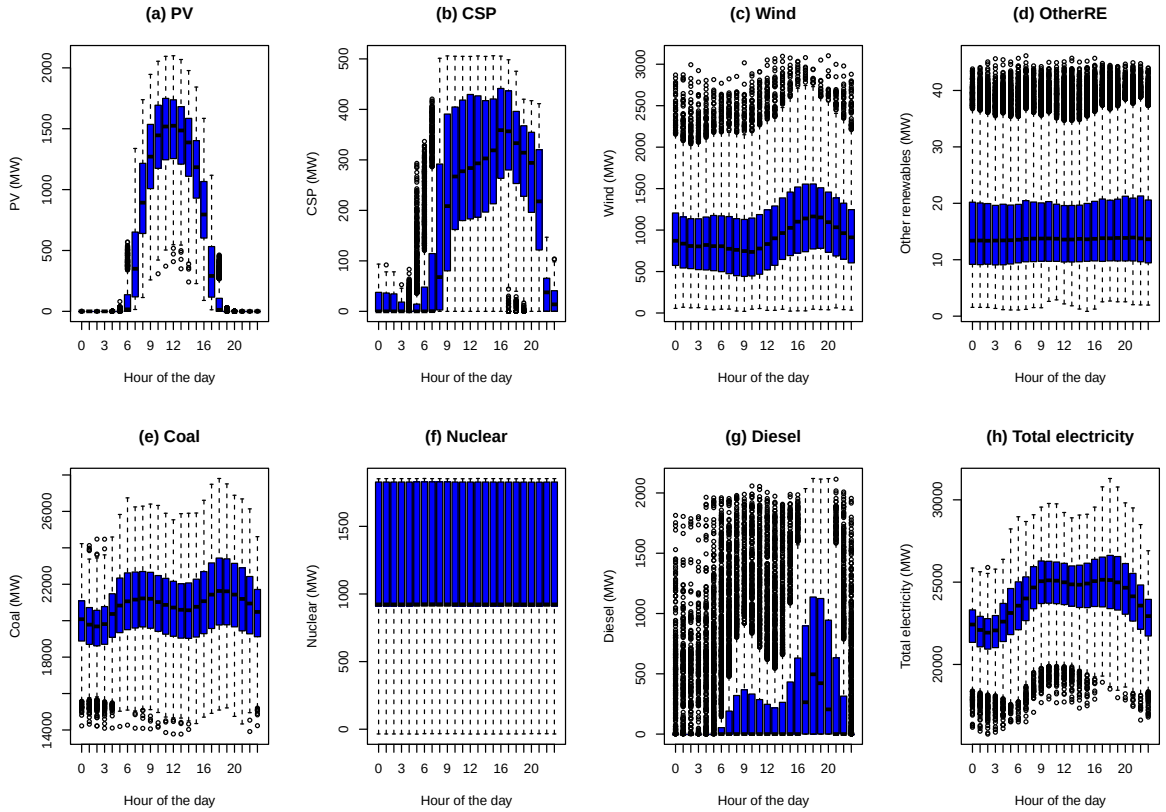


Figure 6. Distribution of hourly electricity consumption generated from different energy sources data.

Table 5. Smoothing parameters for each of the energy sources.

Energy	PV	CSP	Wind	Other RE	Coal	Nuclear	Diesel	Totalnon	TotalALL
λ	0.6726	0.0523	-0.1188	-0.0856	-0.0332	-0.1970	-0.1970	-0.0190	2.09e-08

3.3. Phase I: Forecasting results

In Phase I the analysis compares the forecasts from the XGBoost method with those from the SGB approach. The data is divided into two sets: the training and test sets. The training set constitutes about 97% of the entire dataset (i.e from 2019/04/01 00:00 to 2023/10/04 21:00), and the test set constitutes 3% (i.e. from 2023/10/04 22:00 to 2023/11/28 23:00). Each model is trained, and then sample forecasting is performed using the test set. To assess the ability of each model to predict future values from the test set, different metrics that include the MAE, root mean square error (RMSE), MASE and Bias are used. This analysis is done at all levels: level 2 (disaggregated electricity consumption), level 1 (electricity consumption generated from total renewable and total non-renewable energy sources), and level 0 (total electricity consumption generated from different energy sources).

Table 6 displays response variables as column headers (PV, CSP, Wind, Other RE, Coal, Nuclear, Diesel), and covariates are listed in order of importance for each response variable. Table 6 illustrates how a successful model for a complex system, like a power grid, needs to be multivariate and source-specific. One cannot use the same set of prioritised predictors to forecast solar output and schedule a nuclear plant. The differences in ranking are the key insight, revealing the ”fingerprint” of each energy source’s behaviour on the grid.

The hyperparameters were selected from the search ranges provided in Table 7.

Table 6. Covariate importance ranking for different response variables.

Number	PV	CSP	Wind	OtherRE	Coal	Nuclear	Diesel
1	Lag12	Lag12	dpdfits	dpdfits	dpdfits	dpdfits	Lag12
2	Lag2	Lag2	Lag24	Lag24	hour	Lag24	Lag24
3	hour	dpdfits	Lag12	Lag12	Lag12	day	Lag2
4	Lag1	Lag24	hour	Lag2	day	Lag12	dpdfits
5	dpdfits	Lag1	Lag2	day	Lag24	Lag1	Lag1
6	Lag24	hour	Lag1	Lag1	Lag2	hour	hour
7	day	day	day	hour	Lag1	Lag2	day

Table 7. Hyperparameter search space and final values.

Hyperparameter	Search Range / Values	Description	Final Value
nrounds	{50, 100, 150}	Number of boosting iterations	From bestTune
max_depth	{1, 2, 3, 4, 5, 6, 7,8,9,10 }	Maximum tree depth	From bestTune
eta	{0.01, 0.05, 0.1, 0.2, 0.3}	Learning rate	From bestTune
lambda	{0, 0.1, 0.5, 1, 5}	Minimum loss reduction for split	From bestTune
colsample_bytree	{0.4, 0.6, 0.8, 1.0}	Column sampling ratio	From bestTune
min_child_weight	{1, 2, 3, 5, 10}	Minimum sum of instance weight	From bestTune
subsample	{0.5, 0.6, 0.7, 0.8, 0.9, 1.0}	Row sampling ratio	From bestTune

The final parameters for the XGBoost were tuned using booster = "gbtree", eta = 0.1, maximum depth = 10, subsample = 0.8, colsample bytree = 0.85, lambda = 1, alpha = 0, objective = "reg:squarederror".

Table 8 and Table 9 show the performance of the fitted models in forecasting each of the disaggregated energy sources in level 2 (renewable and non-renewable energy sources, respectively). MASE, MAE, RMSE and Bias are reported for each forecasting method (after performing k-fold cross-validation). The forecasting method with the minimum performance metrics was selected for phase II.

Table 8. Forecast evaluation metrics of electricity consumption generated from renewable energy sources.

(a) Photovoltaic

Model	MASE	MAE	RMSE	Bias
XGBoost	**0.0951**	**14.25**	**63.100**	**-0.552**
SGB	0.105	15.811	93.704	-2.427

(b) Concentrated solar power

Model	MASE	MAE	RMSE	Bias
XGBoost	**0.666**	**19.652**	**32.571**	**4.746**
SGB	1.007	29.704	46.025	9.407

(c) Wind energy

Model	MASE	MAE	RMSE	Bias
XGBoost	2.779	326.522	413.829	-29.669
SGB	**2.604**	**306.038**	**389.389**	**-24.975**

(d) Other renewable energy sources

Model	MASE	MAE	RMSE	Bias
XGBoost	4.185	2.0816	2.892	**0.149**
SGB	**3.382**	**1.682**	**2.158**	-0.264

The results presented in Table 8 show the XGBoost model records lower values of the MASE, MAE, RMSE, and Bias for the electricity consumption generated from PV and CSP energy sources than the SGB approach. This means that XGBoost is more accurate in forecasting PV and CSP than the SGB model. On the other hand, the

SGB model performs better in predicting electricity consumption generated from wind and other renewable energy sources than the XGBoost approach.

Table 9. Forecast evaluation metrics of electricity consumption generated from non-renewable energy sources.

(e) Coal				
Model	MASE	MAE	RMSE	Bias
XGBoost	**1.992**	**369.483**	**468.095**	-40.918
SGB	2.054	380.887	477.683	**-12.788**
(f) Nuclear				
Model	MASE	MAE	RMSE	Bias
XGBoost	19.026	44.883	82.141	4.337
SGB	**10.905**	**25.725**	**50.237**	**1.616**
(g) Diesel				
Model	MASE	MAE	RMSE	Bias
XGBoost	**1.210**	**132.791**	**221.064**	**27.404**
SGB	1.577	172.966	266.903	55.539

The results of electricity consumption generated from non-renewable energy sources show that the XGBoost approach performs better in predicting electricity consumption generated from coal and diesel than the SGB method. The SGB is better at predicting electricity consumption generated from nuclear. Thus, the above results show that XGBoost gives better forecasts for the electricity consumption generated from the four energy sources (PV, CSP, coal, and diesel) out of seven. Based on Figure 4, the electricity consumption generated from these four energy sources varies with the hours of the day. Most electricity consumption is met by coal. This is the case because most electricity in SA is generated from coal [2]. On the other hand, the electricity consumption generated from nuclear and wind do not show any variations in the hour distribution. Table 10 presents the metrics for the XGBoost and the SGB on level 1, total electricity consumption generated from renewable and total non-renewables and level 0, total electricity consumption generated from different energy sources.

Table 10. Forecast evaluation metrics for total renewable and non-renewable energy sources.

(h) Total Renewable (Level 1)				
Model	MASE	MAE	RMSE	Bias
XGBoost	1.579	335.844	431.610	2.931
SGB	**1.537**	**326.968**	**420.327**	**-1.178**
(i) Total Non-renewable (Level 1)				
Model	MASE	MAE	RMSE	Bias
XGBoost	1.412	355.613	449.934	28.758
SGB	**1.363**	**343.339**	**438.851**	**20.469**
(h) Grand Total (Level 0)				
Model	MASE	MAE	RMSE	Bias
XGBoost	**1.217**	**363.516**	**475.350**	**-5.552**
SGB	1.271	379.718	492.167	-23.517

The results from Table 10 show that the SGB model performs better than the XGBoost model in predicting the total electricity consumption generated from renewables and non-renewable energy sources. However, when all the electricity consumption is aggregated (Total), the XGBoost model is a better forecasting approach than the SGB model.

To visualise the performances of these models at various levels, the graph of actual values versus forecasted values for each of the fitted models for level 0, level 1 and level 2 is created. The results are presented in Figure 7(a) (forecasts from the XGBoost), Figure 7(b) (forecasts from the SGB model).

The graphs for the predicted PV, CSP, coal, and diesel are closer to the observed actual electricity consumption, for the test period 2023/10/01 00:00:00 to 2023/11/28 23:00:00, when the XGBoost approach is used in Figure 7 compared to when the SGB method is used (Figure 7(b)). This confirms the findings from the performance metrics in Table 8 and 10.

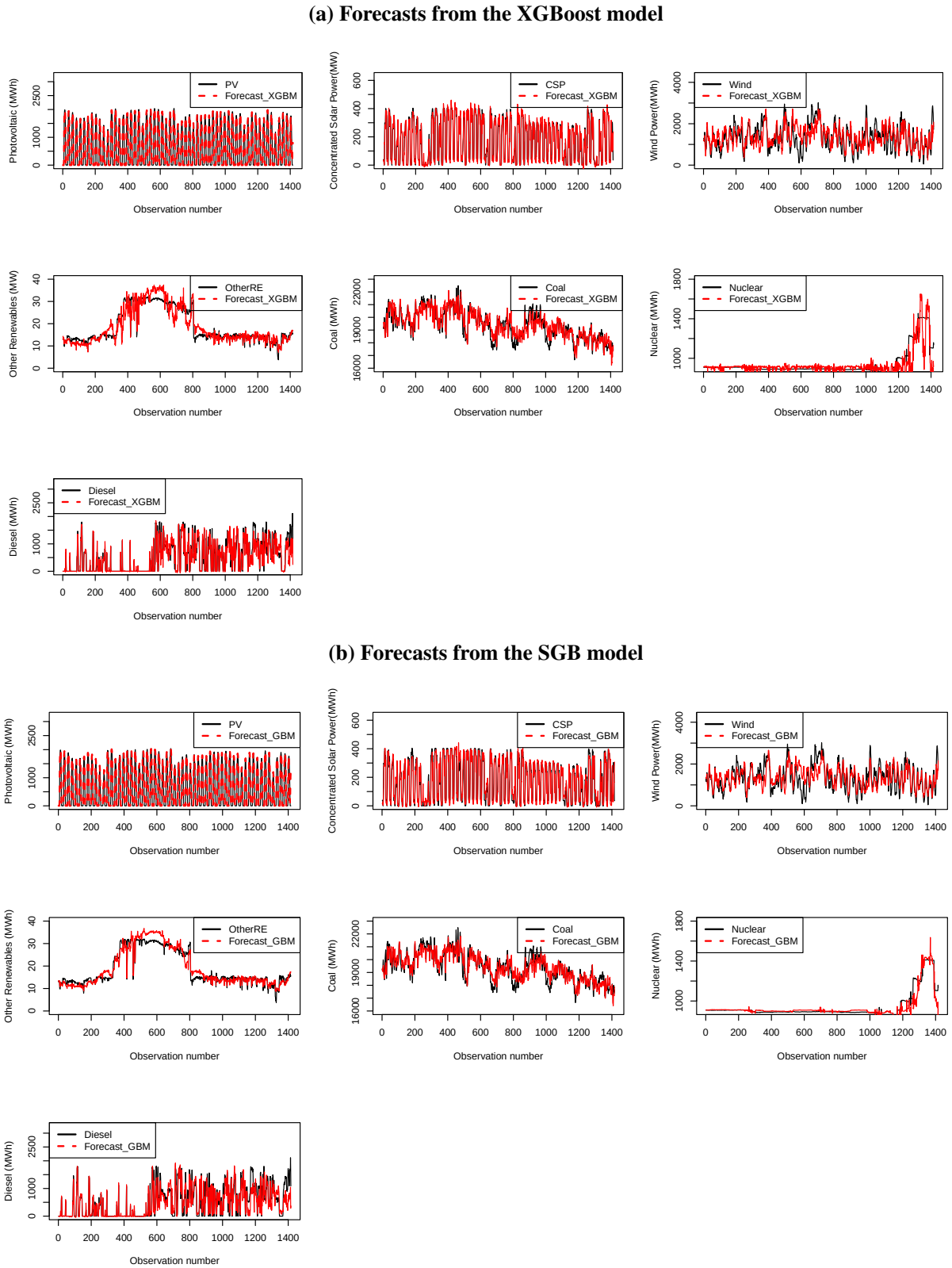


Figure 7. **(a)** Time series graphs with forecasts from the XGBoost model for level 2. **(b)** Time series graph with forecasts from the SGB model for level 2.

Figure 8 presents graphs of the forecasts from the Total renewable, Total non-renewable and Total energy sources superimposed on the observed Total renewable, Total non-renewable and Total energy sources, respectively. These graphs are from the fitted XGBoost model and the SGB model. Results from these graphs show that the XGBoost model perfectly predicts the observed total energy.

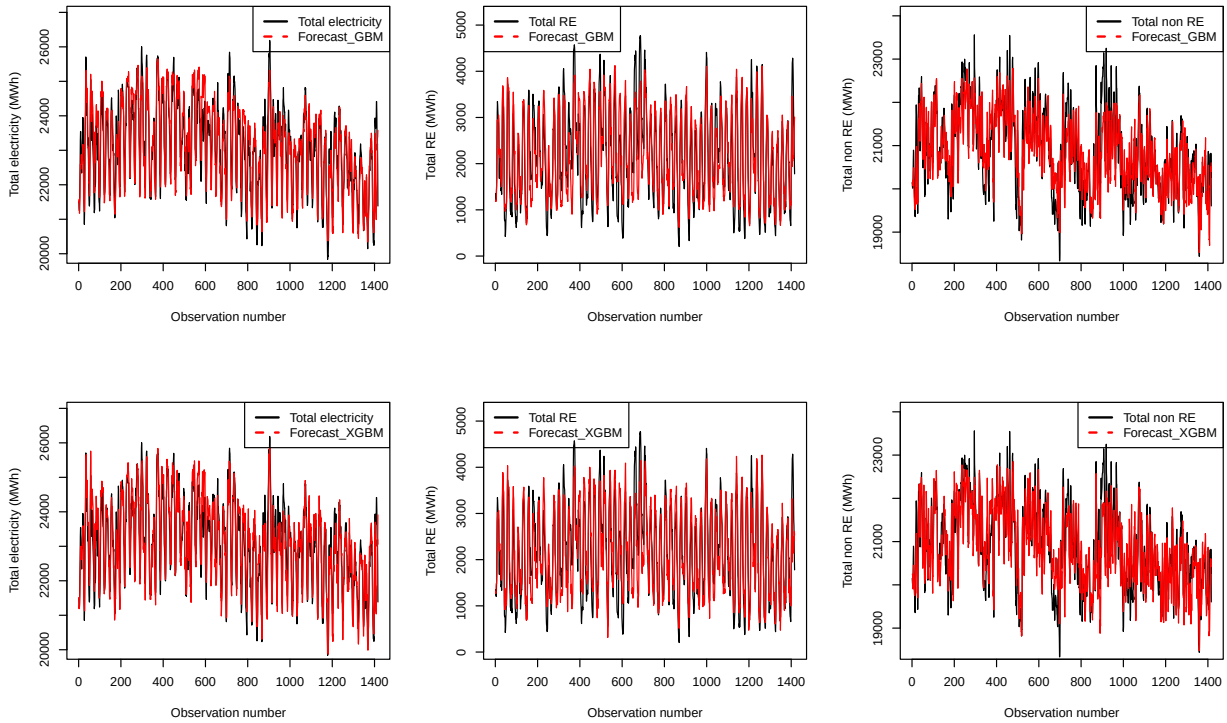


Figure 8. Time series graph with forecasts from the XGBoost model for level 0 (Grand Total) and level 1 (Total renewable and Total non-renewable).

Figure 9 presents the distribution of the residuals from the XGBoost model. This gives further insight into the relative accuracy of the hierarchical forecast models, the XGBoostMost, for the aggregated and disaggregated electricity consumption. The Box plots for each electricity consumption generated from the energy source show normal distributions, although the fitted model failed to capture the outlier in most cases. Overall, the XGBoost model is better than the SGB model. These findings are used in the next phase of hierarchical time series forecasting.

3.4. Phase II: Reconciliation results

To ensure more accurate, coherent forecasts, in this section, the forecasts from the exponential time series (ETS) model and the XGBoost model are compared and also the performances from different reconciliation methods. These reconciliation methods include the BU and the combination approach. Figure10 presents graphs of time series for each hierarchical level. As shown in Figure 10, level 0 has one dataset, which is the aggregated data from all the electricity consumption, level 1 has electricity consumption generated from renewable energy sources category A and non-renewable energy sources category B, and level 2 has electricity consumption generated from seven disaggregated energy sources. From all the level 2 energy sources, coal energy has the highest frequencies throughout the period, meaning that it is the most commonly used energy source because it generates more electricity when compared to other energy sources in SA. Hence, the time series graphs for total electricity

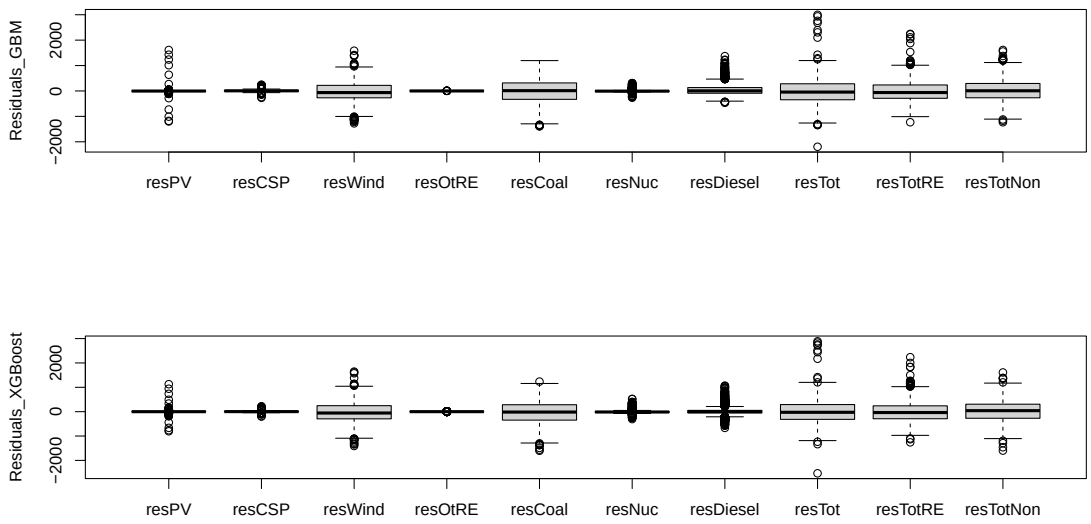


Figure 9. Residuals from the fitted XGBoost and SGB models.

consumption generated from non-renewable in level 1 and the total electricity consumption generated in level 0 depict the same trend as that of the coal energy source.

3.5. Reconciliation using the bottom-up method (XGBoost)

Initially the bottom series are forecasted using the XGBoost model. The reconciliation was then performed through an adjustment factor (combination weight) as shown in equation (13).

$$adjustment.factor = top.level.pred/bottom.level.total, \tag{13}$$

where $top.level.pred$ is the predictions of the total electricity consumption at level 0 and $bottom.level.total$ is the total of the predictions of the disaggregated electricity consumption at level 2 of the hierarchical structure given in Figure 2. In forecast reconciliation, the adjustment factor is important in aligning bottom-level forecasts with the aggregated top-level forecast derived from the XGBoost model. This ensures consistency and accuracy across the hierarchical forecasting structure. This ensures that forecasts on various granularities are consistent and could be used economically for decision-making.

Table 11 presents the performance of the XGBoost, ETS BU and ETS OC reconciliation approaches based on three evaluation metrics: RMSE, MAE and MASE.

From Table 11, the XGBoost reconciliation approach outperforms the ETS BU and ETS OC methods in all evaluation measures considered (RMSE, MAE, and MASE) except for one category in most instances. To be specific, RMSE and MAE values of XGBoost are considerably lower for Total, A (Total renewable energy sources), B (Total non-renewable energy sources), PV, Wind, OtherRE, coal, and OCG, reflecting XGBoost as more accurate with better predictive abilities. However, it is noted that for CSP and Nuclear, the ETS approaches (Bottom-Up and Optimal Combination) are slightly better in RMSE and MAE. This suggests that the XGBoost approach may have problems with forecasting electricity consumption generated from these specific energy sources. MASE also shows that XGBoost is generally better, as it tends to do better than the ETS approaches predicting electricity consumption generated from most energy sources. The only exception is nuclear, where the MASE of XGBoost is far higher. Hence, in conclusion, the overall top performer is XGBoost. However, method choice depends on whether one is

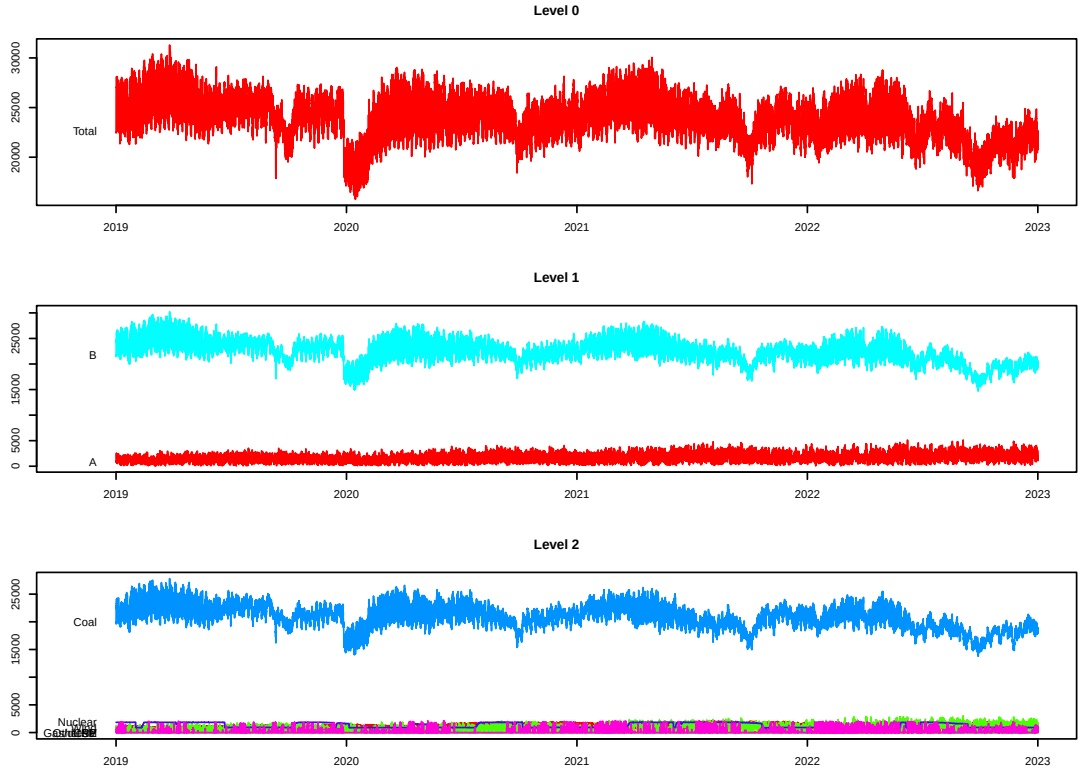


Figure 10. Hierarchical time series graphs using the exponential time series model.
NB: A represents total renewable energy sources, and B represents total non-renewable energy sources and the lower part of level 2 is PV, CSP, Wind, Other renewables, Coal, Nuclear and Diesel energy sources.

Table 11. Hierarchical forecast reconciliation evaluation metrics.

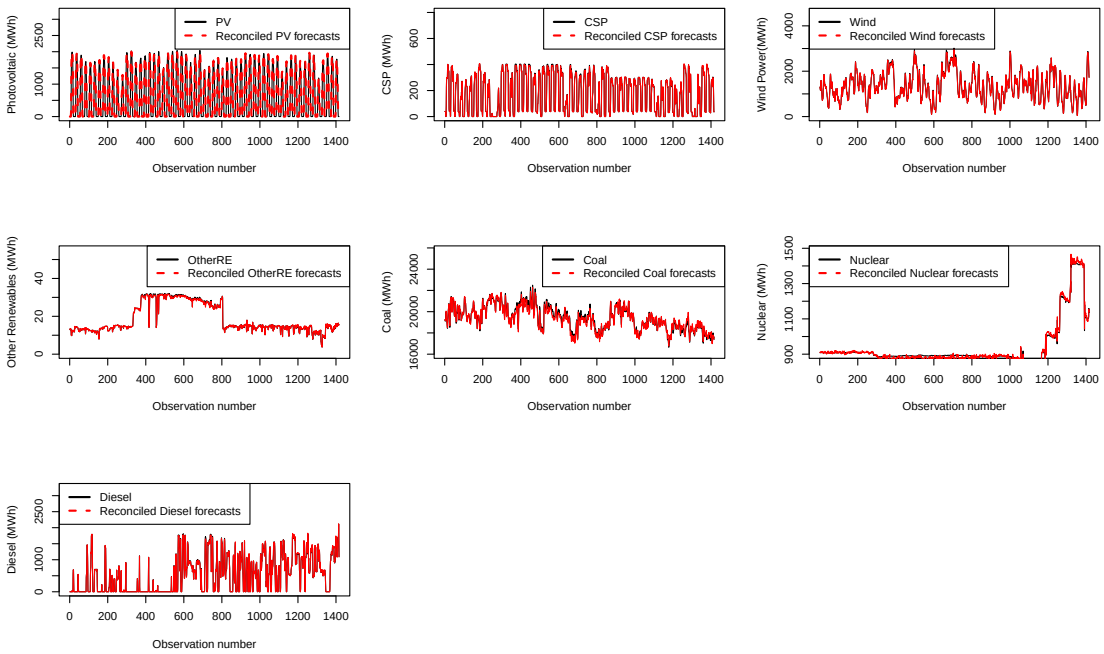
	Method	Total	A	B	PV	CSP	Wind	OtherRE	Coal	Nuclear	Diesel
(a) ETS BU	RMSE	671.67	137.14	590.89	273.48	1.64	163.17	3.23	626.66	0.997	51.15
	MAE	636.54	126.83	584.37	199.85	1.555	152.14	2.907	613.23	0.997	41.54
	MASE	0.36	0.24	0.31	1.50	**0.02**	0.31	**0.36**	0.34	**0.002**	0.13
(b) ETS OC	RMSE	603.08	138.26	521.14	274.41	1.83	162.65	3.23	567.82	**0.93**	60.46
	MAE	570.49	127.35	516.66	200.84	1.73	151.64	2.91	556.49	**0.93**	44.65
	MASE	**0.32**	0.24	**0.28**	1.51	**0.02**	0.31	**0.36**	**0.31**	**0.002**	0.14
(c) XGBoost	RMSE	**270.40**	**32.07**	**240.68**	**13.56**	3.17	**18.51**	**0.28**	**222.72**	11.04	**9.85**
	MAE	**204.78**	**21.10**	**183.67**	**6.84**	**1.85**	**12.22**	**0.19**	**170.17**	8.34	**5.17**
	MASE	0.69	**0.10**	0.73	**0.05**	0.06	**0.10**	0.38	0.92	3.53	**0.05**

Key: A: Total renewable energy sources; B: Total non-renewable energy sources.

modelling electricity consumption from a certain type of energy source. The ETS-based methods remain valid for certain energy sources, such as CSP and nuclear.

Figure 11 presents the reconciled forecasts generated by the XGBoost model, representing the final, adjusted predictions for the test period. These forecasts reflect the total reconciled output from the XGBoost model, offering a consolidated view of the predictions following reconciliation.

(a) Reconciled forecasts from the XGBoost model



(b) Total reconciled forecasts from the XGBoost model

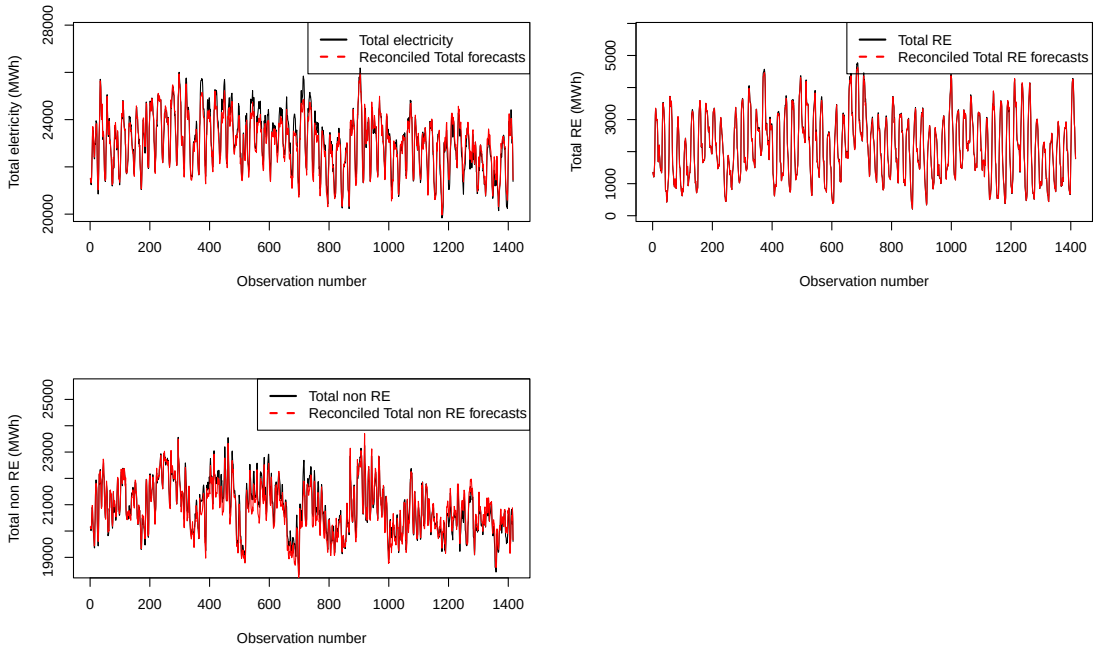


Figure 11. Forecasts from XGBoost reconciliation for the test period.

3.5.1. Results on the evaluation of prediction intervals Table 12 presents evaluation metrics that compare prediction intervals based on two different regression methods: linear and quantile regression. The prediction intervals are evaluated on the reconciliation forecasts obtained from the XGBoost method.

It can be observed in Table12 that the linear quantile regression method performs better than the linear regression method based on accuracy (coverage probability) and precision (interval score and average width). It offers a good compromise with moderate coverage probabilities, yielding lower interval scores and narrower average widths. However, the linear regression has slightly better coverage probability, which is vital to the reliability of the prediction intervals. Based on these results, linear quantile regression is recommended because it maintains a competitive coverage probability while demonstrating superior precision with a lower average width and interval score, providing a more focused range of credible predictions.

Table 12. Evaluation metrics for the prediction intervals.

(a) Prediction intervals based on linear regression			
	Total energy	Total renewable energy sources	Total non-renewable energy sources
Coverage probability	0.958	0.946	0.955
Interval score	2117.72	1259.32	2268.57
Average width	1873.99	1065.67	1980.90
PINAW	0.306	0.241	0.362
(b) Prediction intervals based on linear quantile regression			
	Total energy	Total renewable energy sources	Total non-renewable energy sources
Coverage probability	0.950	0.951	0.951
Interval score	2112.10	1227.47	2219.49
Average width	1803.41	1063.99	1943.66
PINAW	0.300	0.241	0.356

Figures 12 and 13 show graphs of the prediction intervals for total electricity consumption, electricity consumption from total renewable and total non-renewable energy sources from the linear quantile approach. The PIs for the non-renewables are very wide. This is explained by the high scores of their interval score, average width and PINAW from Table 12.

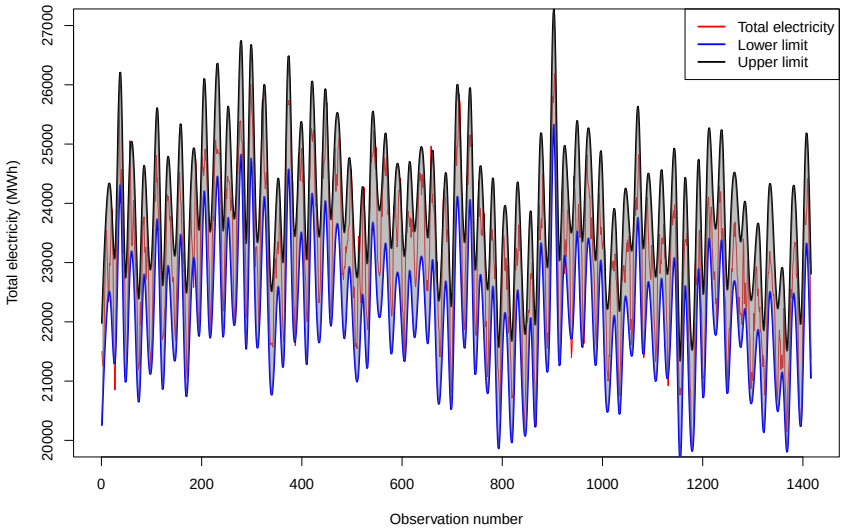


Figure 12. Prediction intervals from the linear quantile regression approach: Total energy for the test period.

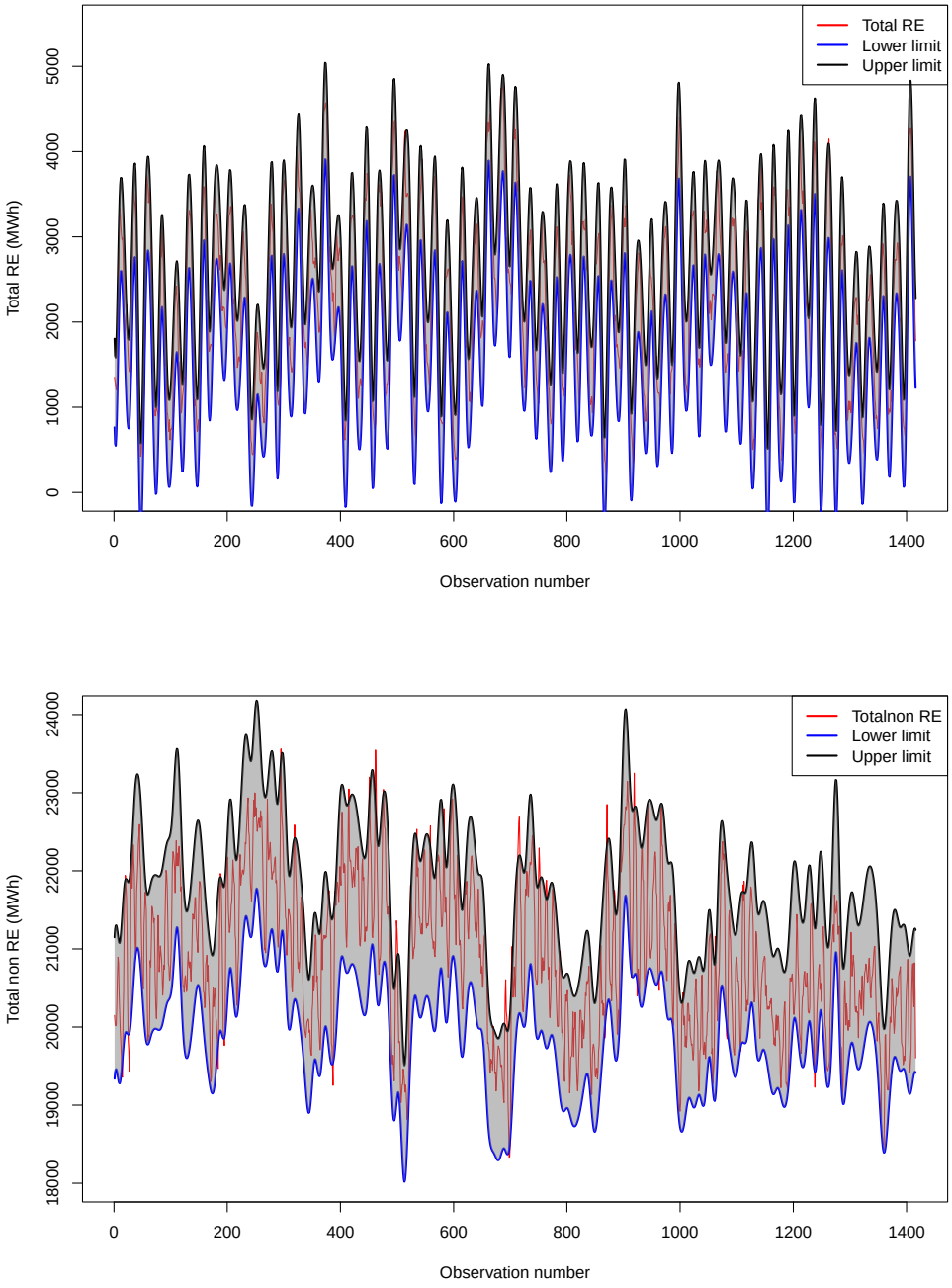


Figure 13. Prediction intervals from the linear quantile regression approach: **Top panel:** Renewable energy. **Bottom panel:** Non renewable energy for the test period.

4. Conclusion

This paper presents a robust framework for short-term forecasting of hierarchical time series in electricity consumption using South African data. Integrating XGBoost and SGB with reconciliation methods has proven effective in improving forecasting accuracy and coherence. The study confirms that the XGBoost model effectively predicts electricity consumption generated from all energy sources. However, it can also accurately predict electricity consumption met by solar (photovoltaic, concentrated solar power), coal, and Diesel energy sources. The SGB method yielded superior performance for electricity consumption met by wind and nuclear energy sources. This indicates the importance of selecting prediction models for electricity consumption specific to certain energy sources rather than using a single model for all sources, thus highlighting the complex nature of electricity consumption prediction.

The study recommends the application of more than one method of forecasting so that the accuracy of prediction for aggregate, non-renewable, and coal energy source is increased. The study further recognises that these hybrid approaches significantly improve predictions for electricity consumption met by more challenging energy sources like wind and nuclear. This study explores different forecasting models and identifies a huge gap in the literature. It presents useful findings for decision-makers in the electricity sector.

The study also examined prediction interval estimation using two regression models: linear regression and linear quantile regression. The findings indicate that linear quantile regression provides improved accuracy with narrower interval widths and lower interval scores while maintaining a competitive level of coverage probability. This approach is therefore recommended as a viable method for reliable predictions.

This work identifies some key directions for future research in hierarchical electricity consumption forecasting. These include the development of more adaptive, source-specific hybrid models, with an emphasis on volatile sources, such as wind. Exogenous variables, including weather and economic data, may lead to higher levels of predictive accuracy. Testing the transferability of the framework to other regions or integrated energy systems would certainly validate its broader applicability.

Data Availability Statement

The data that support the findings of this study are available at https://github.com/csigauke/. The data is analytic data which was used for developing the models used in this study.

Authorship Contribution

The authors contributed equally to this paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to the numerous people for helpful comments on this paper.

Abbreviations

The following abbreviations are used in this manuscript.

ARIMA	Autoregressive Integrated Moving Average
AW	Average Width
BU	Bottom Up
CP	Coverage Probability
CSP	Concentrated Solar Power
ETS	Exponential Smoothing
IS	Interval Score
PI	Prediction Interval
PINAW	Prediction Interval Normalised Average Width
PV	Photovoltaic
OC	Optimal Combination
OtherRE	Other Renewable
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
RMSE	Root Mean Squared Error
SGB	Stochastic Gradient Boosting
XGBoost	Extreme Gradient Boosting

Appendix

Algorithm 2 Energy Forecasting with SGB and Bottom-Up Reconciliation

Input: Time series data $D = \{PV, CSP, Wind, OtherRE, Thermal, Nuclear, OCGT\}$
Output: Reconciled forecasts R
procedure MAIN(D)
    $data \leftarrow$ DataFrame$(PV, CSP, Wind, OtherRE, Thermal, Nuclear, OCGT)$
    $data$.Total_energy $\leftarrow \sum_{i=1}^{7} data_i$
    Split data:
        $train\_data \leftarrow data[1:39456]$
        $test\_data \leftarrow data[39457:40872]$
    $base\_forecasts \leftarrow train\_data[:, 1:7]$
    $model \leftarrow$ TrainSGBModel$(train_data, base_forecasts)$
    $reconciled\_forecasts \leftarrow$ ReconcileForecasts$(model, test_data)$
    metrics $\leftarrow$ EvaluateForecasts$(test_data, reconciled_forecasts)$
    return $(reconciled\_forecasts,$ metrics$)$
end procedure
procedure TRAINSGBMODEL($train\_data, base\_forecasts$)
    $X \leftarrow$ matrix$(base_forecasts)$
    $y \leftarrow train\_data$.Total_energy
    Initialize: $F_0(x) = \bar{y}$
    Set SGB parameters:
        n_trees $\leftarrow 100$; learning_rate $\leftarrow 0.1$; tree_depth $\leftarrow 10$
        subsample $\leftarrow 0.8$; min_obs $\leftarrow 10$

for $m = 1$ to n_trees **do**
  Compute pseudo-residuals:
    $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{m-1}}$
  Randomly select subset:
    $I_m \leftarrow$ random sample of size $\lfloor \text{subsample} \cdot n \rfloor$
  Fit regression tree to residuals:
    $h_m(x) \leftarrow \text{RegressionTree}(X[I_m], r[I_m], \text{depth}, \text{min\_obs})$
  Update model:
    $F_m(x) = F_{m-1}(x) + \nu \cdot h_m(x); \quad \nu = \text{learning\_rate}$
end for
return $F_{\text{n\_trees}}(x)$
end procedure
procedure RECONCILEFORECASTS($model, test\_data$)
  $X\_test \leftarrow \text{matrix}(test\_data[:, 1:7])$
  $top\_level\_pred \leftarrow \text{predict}(model, X\_test)$
  $bottom\_level\_total \leftarrow \sum_{i=1}^{7} X\_test_i$
  Ensure non-negative adjustment:
    $adjustment\_factor \leftarrow \max\left(0.1, \frac{top\_level\_pred}{bottom\_level\_total}\right)$
  $reconciled \leftarrow X\_test \times adjustment\_factor$
  return $reconciled$
end procedure
procedure EVALUATEFORECASTS($actual, forecast$)
  $n \leftarrow \text{length}(actual)$
  $\text{MASE} \leftarrow \frac{\frac{1}{n}\sum_{t=1}^{n} |forecast_t - actual_t|}{\frac{1}{n-1}\sum_{i=2}^{n} |actual_i - actual_{i-1}|}$
  $\text{MAE} \leftarrow \frac{1}{n}\sum_{t=1}^{n} |forecast_t - actual_t|$
  $\text{RMSE} \leftarrow \sqrt{\frac{1}{n}\sum_{t=1}^{n} (forecast_t - actual_t)^2}$
  $\text{Bias} \leftarrow \frac{1}{n}\sum_{t=1}^{n} (forecast_t - actual_t)$
  Additional SGB-specific diagnostics:
    OOB_error $\leftarrow$ Compute out-of-bag error from SGB
    Variable_importance $\leftarrow$ Compute feature importance
  return {MASE, MAE, RMSE, Bias, OOB_error, Variable_importance}
end procedure
procedure REGRESSIONTREE($X, r, depth, min\_obs$)
  if stop condition met **then**
    return $\bar{r}$
  end if
  Find best split:
    $(j^*, s^*) = \arg\min_{j,s}\left[\sum_{x_i \in R_1(j,s)} (r_i - \bar{r}_1)^2 + \sum_{x_i \in R_2(j,s)} (r_i - \bar{r}_2)^2\right]$
  $R_1(j^*, s^*) = \{X | X_j \leq s^*\}; \quad R_2(j^*, s^*) = \{X | X_j > s^*\}$
  $\text{left} \leftarrow \text{RegressionTree}(R_1, r, depth - 1, min\_obs)$
  $\text{right} \leftarrow \text{RegressionTree}(R_2, r, depth - 1, min\_obs)$
  return tree with split (j^*, s^*), left, right
end procedure

REFERENCES

1. Jim Crompton, *Data management from the DCS to the historian and HMI*, Editor(s): Patrick Bangert, Machine Learning and Data Science in the Power Generation Industry, Elsevier, 2021, Pages 93-122, https://doi.org/10.1016/B978-0-12-819742-4.00005-6.
2. Eskom, *Eskom Integrated Report for the year ended 31 March 2024*, 2024. Accessed: 7 April 2025. https://www.eskom.co.za/wp-content/uploads/2024/12/Eskom-integrated-report-2024.pdf.
3. J. Leprince, H. Madsen, J. K. Møller, and W. Zeiler, *Hierarchical learning, forecasting coherent spatio-temporal individual and aggregated building loads*, Applied Energy, vol. 348, p. 121510, 2023. https://doi.org/10.1016/j.apenergy.2023.121510
4. V. Almeida, R. Ribeiro, and J. Gama, *Hierarchical time series forecast in electrical grids*, in Information Science and Applications (ICISA) 2016, pp. 995–1005, 2016. https://doi.org/10.1007/978-981-10-0557-2_95
5. M. E. Hansen, N. Peter, J. K. Møller, and M. Henrik, *Reconciliation of wind power forecasts in spatial hierarchies*, Wind Energy, vol. 26, no. 6, pp. 615–632, 2023. https://doi.org/10.1002/we.2819
6. R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, *Optimal combination forecasts for hierarchical time series*, Computational statistics & data analysis, vol. 55, no. 9, pp. 2579–2589, 2011. https://doi.org/10.1016/j.csda.2011.03.006
7. T. Silveira Gontijo and M. Azevedo Costa, *Forecasting Hierarchical Time Series in Power Generation*, Energies, vol. 14, no. 13, p. 1996-1073, 2020. https://doi.org/10.3390/en13143722
8. G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and A. Panagiotelis, *Forecast reconciliation: A review*, International Journal of Forecasting, vol. 40, pp. 430–456, 2024. https://doi.org/10.1016/j.ijforecast.2023.10.010
9. H. Wen and P. Pinson, *Value-oriented forecast reconciliation for renewables in electricity markets*, ArXiv Preprint Version 1, pp. 1–30, 2025. https://doi.org/10.48550/arXiv.2501.16086
10. S. Taghiyeh, D. C. Lengacher, A. H. Sadeghi, A. Sahebi-Fakhrabad, and R. B. Handfield, *A novel multi-phase hierarchical forecasting approach with machine learning in supply chain management*, Supply Chain Analytics, vol. 3, p. 100032, 2023. https://doi.org/10.1016/j.sca.2023.100032
11. J. Rombouts, M. Ternes, and I. Wilms, *Cross-temporal forecast reconciliation at digital platforms with machine learning*, International Journal of Forecasting, vol. 41, pp. 321–344, 2025. https://doi.org/10.1016/j.ijforecast.2024.05.008
12. N. Kafa, M. Z. Babai, and W. Klibi, *Forecasting mail flow: A hierarchical approach for enhanced societal wellbeing*, International Journal of Forecasting, vol. 41, pp. 51–65, 2025. https://doi.org/10.1016/j.ijforecast.2024.07.001
13. Florek, P., & Zagdański, A. (2023). *Benchmarking state-of-the-art gradient boosting algorithms for classification*, ArXiv, 2023. https://arxiv.org/abs/2305.17094
14. Michael P. Rogers, Paul C. Kuo *Machine learning to predict mortality risk in coronary artery bypass surgery*, Cardiovascular and Coronary Artery Imaging , 2022 pp 195-210
15. Aurélien Géron *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, (O'Reilly Media) - A comprehensive guide to machine learning, offering practical guidance on gradient boosting algorithms, including a detailed comparison and application of XGBoost, 2022
16. W. Abouarghoub, N. K. Nomikos, and F. Petropoulos, *On reconciling macro and micro energy transport forecasts for strategic decision making in the tanker industry*, Transportation Research Part E: Logistics and Transportation Review, vol. 113, pp. 225–238, 2018. https://doi.org/10.1016/j.tre.2017.10.012
17. X. Wang, R. J. Hyndman, and S. L. Wickramasuriya, *Optimal forecast reconciliation with time series selection*, European Journal of Operational Research, vol. 323, no. 2, pp. 455–470, 2025. https://doi.org/10.1016/j.ejor.2024.12.004
18. S. L. Wickramasuriya, *Properties of point forecast reconciliation approaches*, arXiv preprint arXiv:2103.11129, 2021. https://doi.org/10.48550/arXiv.2103.11129
19. T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016. https://doi.org/10.1145/2939672.2939785
20. J. H. Friedman, *Stochastic gradient boosting*, Comput. Stat. Data Anal., vol. 38, pp. 367–378, 2002. https://doi.org/10.1016/S0167-9473(01)00065-2
21. P. Mpfumali, C. Sigauke, A. Bere, and S. Mulaudzi, *Day Ahead Hourly Global Horizontal Irradiance Forecasting: Application to South African Data*, Energies, vol. 12, p. 3569, 2019. https://doi.org/10.3390/en12183569
22. J. H. Friedman, *Greedy function approximation: A gradient boosting machine*, Ann. Stat., vol. 29, pp. 1189–1232, 2001. https://doi.org/10.1214/aos/1013203451
23. T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, *The elements of statistical learning: data mining, inference and prediction*, Math. Intell., vol. 27, pp. 83–85, 2005. https://doi.org/10.1007/BF02985802
24. A. Mahdi, R.J. Hyndman, E. Spiliotis, and C. Bergmeir, *Model selection in reconciling hierarchical time series*, Machine Learning, vol. 111, pp. 739–789, 2022. https://doi.org/10.1007/s10994-021-06126-z
25. E. Nikulchev and A. Chervyakov, *Prediction Intervals: A Geometric View*, Symmetry, vol. 15, p. 781, 2023. https://doi.org/10.3390/sym15040781
26. R. J. Hyndman, A. J. Lee, and E. Wang, *Fast computation of reconciled forecasts for hierarchical and grouped time series*, Computational Statistics and Data Analysis, vol. 97, pp. 16–32, 2016. https://doi.org/10.1016/j.csda.2015.11.007
27. S. Ben Taieb, J. W. Taylor, and R. J. Hyndman, *Coherent probabilistic forecasts for hierarchical time series*, in Proceedings of the 34th international conference on machine learning (D. Precup and Y. W. Teh, eds.), vol. 70 of Proceedings of machine learning research, pp. 3348–3357, PMLR, 2017. https://proceedings.mlr.press/v70/taieb17a.html