"Highly-Accurate Electricity Load Forecasting via Hybrid Machine Learning" Supplementary File

Di Wu, *Member, IEEE,* Yuting Ding, Song Deng, *Member, IEEE,* Yi He, *Member, IEEE,* and Xin Luo, *Senior Member, IEEE*

S.I. INTRODUCTION

THIS is the supplementary file for the paper entitled "Highly-Accurate Electricity Load Forecasting via Hybrid Machine Learning" in Transactions on Industrial Electronics. We have put some contents cited by the paper here, including Section S.II. Preliminaries, Section S.III. Cited Tables, and Section S.IV XASX's Sensitivity to Its Hyper-parameters.

S.II. PRELIMINARIES

A. Symbols and Notations

Table S.I summarizes the adopted main symbols.

TABLE S.I
THE ADOPTED MAIN SYMBOLS AND THEIR DESCRIPTIONS.

Symbol	Description
L, $\bar{L}$	The real and forecasted monthly electricity load time-series.
N	The length of L.
R_i, $\hat{R}_i$	The i-th real and forecasted value of L, $i\in\{1, 2, ..., N\}$.
Lt, $\bar{Lt}$	The decomposed and forecasted trend component of L.
Ls, $\bar{Ls}$	The decomposed and forecasted seasonal component of L.
Li, $\bar{Li}$	The decomposed and forecasted irregular component of L.
$\boldsymbol{F}$	The feature set of external factors.
p	The number of autoregression terms.
q	The number of moving average terms.
d	The differential order.
y_i, $\hat{y}_i$	A general value and its forecasted one.
N	The length of the load series.

B. Census X12 Decomposition Technique

The quarterly or monthly load time-series are non-stationary and regional. They are heavily affected by economic development and industrial restructuring, seasonal changes, non-linearity, and other uncertain factors. Hence, they can be decomposed into trend, seasonal, and irregular components. The trend component reflects the long-term change pattern of load influenced by the region's economic development and industrial structure adjustment. The seasonal component reflects the cyclic change of load influenced by the region's climate change in the same season of different years. The irregular component refers to the irregular influences caused by abnormal events, natural disasters, noise, *etc*. X12 [1], which is frequently adopted to analyze economic time-series, can extract these three components from load time-series. The basic decomposition methods of X12 include the addition decomposition and the multiplication decomposition as follows:

$$L = Lt + Ls + Li, \tag{1}$$

$$L = Lt \times Ls \times Li. \tag{2}$$

where L is the load time-series, Lt is the trend component, Ls is the seasonal component, and Li is the random component.

C. ARIMA Model

ARIMA (autoregressive integrated moving average) [2] is a traditional time-series forecasting model. It can be divided into autoregressive, moving average, and autoregressive moving average models. Its modeling process is to first judge the stationary degree, then use the difference method to stabilize the non-stationary time-series, and finally select AR(p) and MA(q) to determine the order of the model. The general expression of the ARIMA(p, d, q) is as follows:

$$\begin{aligned} w_t &= \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots \phi_p w_{t-p} \\ &+ \phi + u_t - \theta_1 u_{t-1} - \theta_2 u_{2-1} - \cdots \theta_q u_{t-q}. \end{aligned} \tag{3}$$

where (ϕ_1, ϕ_2, …, ϕ_p) is an autoregressive coefficient, p is the order of autoregression, (θ_1, θ_2, ⋯, θ_q) is the moving average coefficient, u_t is a white noise sequence (zero mean, normal distribution, and independent from each other). The values of p and q can be determined by autocorrelation and bias correlation plots.

D. SVR Model

SVR (support vector regression) [3] allows maximum ε deviation between the model output $f(x)$ and the real output y. In other words, the *loss* is calculated only when the absolute value of the difference between $f(x)$ and y is greater than ε. When the sample falls into the interval band with a width of 2ε, the prediction is considered correct, which enables SVR to avoid overfitting to a certain extent and improve generalization ability.

E. XGBoost Model

XGBoost (extreme gradient boosting) [4] is an optimization of the Boosting algorithm. It integrates several weak classifiers into a strong classifier. XGBoost generates a new tree to fit the residuals of the previous tree by iterating continuously, and the accuracy improves as the number of iterations increases. Therefore, XGBoost can fit the load time-series better to reduce the forecasting errors.

The regression model adopted by XGBoost is the CART tree model. Its objective function is as follows:

$$\hat{y}_i = \sum_{t=1}^{m} f_t(x_i).,\ f_t \in \Phi\ , \tag{4}$$

where m is the number of trees, F is the set of all possible CARTs, f_t is a forecasting function on the space Φ, $\hat{y}_i$ is the forecasted value, and x_i is the i-th input sample.

S.III. Cited Tables

TABLE S.II
ALGORITHM 1. XASX

Steps	Input: $\boldsymbol{F}$, L; Output: $\bar{L}$	Cost
1	**Initialize** *p , q, d, max_depth, learning_rate*	$\Theta(1)$
2	**Initialize** *Lt* randomly	$\Theta(N)$
3	**Initialize** *Ls* randomly	$\Theta(N)$
4	**Initialize** *Li* randomly	$\Theta(N)$
5	**Decompose** monthly electricty load series	$\Theta(N^2)$
6	**Update** *Li*, *Ls*, *Lt*	$\Theta(1)$
7	**While *Lt* not static**	$\times N$
8	Update $Lt_i = Lt_i - Lt_{i-1}$	$\Theta(1)$
9	**end while**	--
10	**Update** all the parameters of (13) using OLS	$\Theta(N^2)$
11	**For** *i=1 to n*	$\times N$
12	Update Lt_i according to (17)	$\Theta(1)$
13	**end for**	$\Theta(1)$
14	**Update** $\boldsymbol{F}$ according to (12)	$\Theta(N)$
15	**do**	--
16	Update $\bar{L}t$=SVR({temperature, month, holiday, *Lt*})	$\Theta(N^3)$
17	**end**	--
18	**do**	--
19	Update $\bar{L}i$=XGBoost({temperature, month holiday, *Li*})	$\Theta(NlogN)$
20	Update $\bar{L}s$=XGBoost({temperature, month holiday, *Ls*})	$\Theta(NlogN)$
21	**end**	--
22	**Update** $\bar{L}= \bar{L}t\times\bar{L}i\times \bar{L}s$	$\Theta(1)$

TABLE S.Ⅲ
STATISTICS OF THE STUDIED DATASETS

Name	Time Range	Feature
D1	01/01/2013-31/12/2021	12
D2	01/01/2013-31/12/2021	12
D3	01/01/2013-31/12/2021	12
D4	01/01/2016-31/12/2021	12

TABLE S.IV
THE DESCRIPTIONS OF COMPARISON MODELS.

Model	Description
ETS	ETS is a common local statistical algorithm for time-series prediction.
X12-ARIMA	X12 multiplication model is used to decompose monthly load time-series, and ARIMA is used to forecast.
GRNN	The general regression neural network (GRNN) model is a four-layer neural network with Gaussian nodes focused on the training pattern.
MLP	Multilayer perceptron (MLP) with a single hidden layer and sigmoidal neurons.
XGBoost	XGBoost is an optimized distributed gradient enhancement library designed to be efficient, flexible, and portable.
SVR	SVR is a "tolerant regression model" with strict linear regression .
LSTM	LSTM is a recursive neural network that can learn and forecast time-series.
NBEATS	A deep neural architecture modeled by fully connecting layers connected with forward and backward residual links.
ETS+RD-LSTM	It combines ETS, advanced LSTM, and ensemble learning to build a hybrid model.
APLF	It is a probabilistic LF method based on adaptive on-line learning of hidden Markov model.

S.IV. XASX's Sensitivity to Its Hyper-parameters

To evaluate XASX's sensitivity to its hyper-parameters, we carry out experiments with different values of p (the number of autoregression terms) and q (the number of moving average terms) of ARIMA as well as different values of md (the max depth) and lr (the learning rate) of XGBoost. The results of MAPE are presented in Fig. S.1. Fig. S.1(a) shows that the performance of XASX relies on the setting of p and q. On the different datasets, XASX achieves the lowest MAPE with different p and q. On the other hand, Fig. S.1(b) shows that different md and lr lead to different MAPEs on different datasets, where we see that md is data-dependent and lr is relatively stable between 0.1 and 0.11. Therefore, we conclude that our XASX is relatively robust to the learning rate of XGBoost and data-dependent to the numbers of autoregression and moving average terms of ARIMA and the max depth of XGBoost. These data-dependent hyper-parameters can be tuned on a partial training set.

References

[1] G. Xie, N. Zhang, and S. Wang, "Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology," *Transportation Research Part E: Logistics and Transportation Review*, vol 108, pp.160-178, 2017.

[2] M. H. Amini, A. Kargarian, and O. Karabasoglu, "ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation," *Electric Power Systems Research*, vol. 140, no. nov., pp. 378–390, 2016.

[3] Y. Chen et al., "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, Jun. 2017.

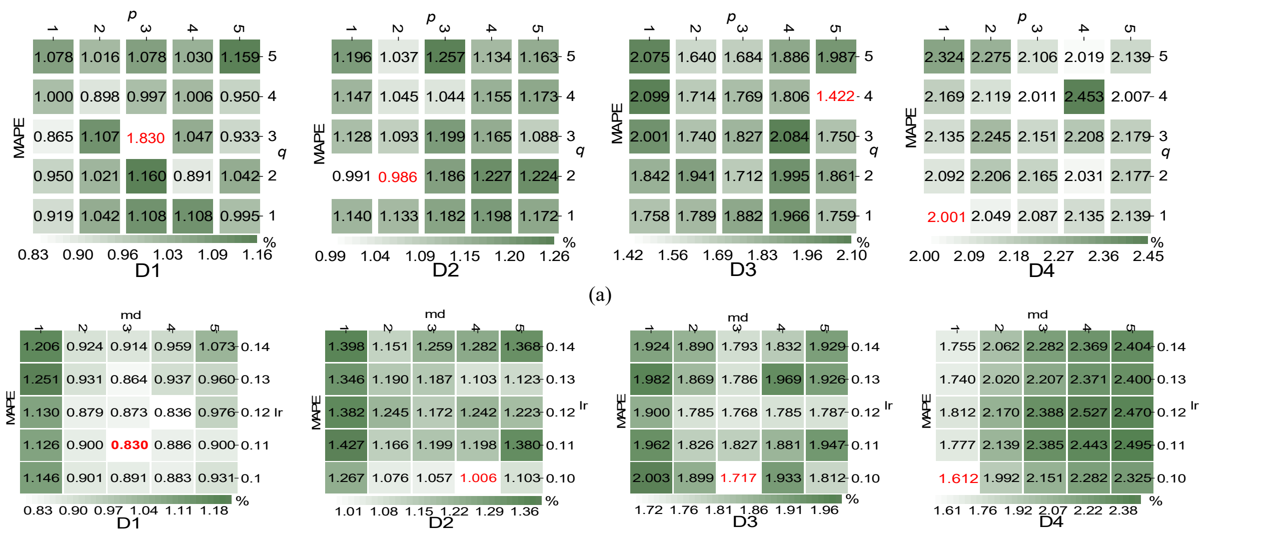
[4] B. Patnaik, M. Mishra, R. C. Bansal, and R. K. Jena, "MODWT-XGBoost based smart energy solution for fault detection and classification in a smart microgrid," *Applied Energy*, vol. 285, no. 113503, p. 116457, 2021.

TABLE S.V
THE COMPARISON RESULTS OF MAPE ON D3.

Month	X12-ARIMA	ETS	LSTM	NBEATS	MLP	GRNN	SVR	XGBOOST	RD-ETS+LSTM	APLF	XASX
1	9.55%	10.15%	4.39%●	3.68%●	18.93%	14.14%	14.04%	14.22%	7.51%	24.30%	6.45%
2	13.78%	12.82%	7.04%	1.65%	48.31%	0.88%	44.25%	0.36%	13.37%	4.09%	0.07%
3	1.78%●	7.94%	1.38%●	6.69%	2.88%●	11.59%	5.13%	31.55%	1.66%●	22.29%	3.15%
4	0.12%●	2.27%	2.51%	4.71%	4.65%	7.31%	2.26%	0.00%●	1.08%●	17.55%	1.27%
5	1.95%	0.81%	0.24%●	5.99%	5.57%	8.27%	4.42%	17.90%	3.15%	15.02%	0.40%
6	3.72%	5.46%	6.07%	4.69%	0.67%●	13.57%	3.02%●	3.45%●	2.42%●	16.68%	3.64%
7	2.22%●	0.00%●	4.89%	1.71%●	16.58%	3.63%●	18.40%	8.86%	2.00%●	22.69%	4.09%
8	2.03%	2.45%	5.15%	7.32%	18.33%	13.32%	20.94%	10.95%	0.45%●	5.68%	0.95%
9	1.74%	8.20%	9.08%	5.50%	7.49%	3.25%	4.32%	1.81%	4.47%	1.06%	0.15%
10	8.74%	14.21%	16.69%	0.23%	16.05%	3.13%	13.84%	2.13%	10.16%	4.34%	0.14%
11	2.33%	1.01%	3.32%	5.06%	0.98%	11.48%	1.69%	0.54%	3.71%	6.32%	0.38%
12	5.67%	1.83%	4.29%	2.08%	13.99%	10.49%	16.48%	2.70%	2.36%	5.98%	0.66%
Mean-MAPE	4.47% ±4.14%	5.60% ±4.95%	5.42% ±4.29%	4.11% ±2.24%	12.87% ±13.1%	8.42% ±4.7%	12.40% ±12.18%	7.87% ±9.54%	4.36% ±3.97%	12.17% ±8.43%	**1.78% ±2.06%**
Win/Loss	9/3	11/1	9/3	10/2	10/2	11/1	11/1	10/2	7/5	12/0	**100/20**
F-rank	4.8333	5.6667	6.0833	5.5833	7.6667	7.6667	7.3333	5.6667	4.5000	8.3333	**2.6667**
***p*-value**	**0.0261**	**0.0061**	**0.0105**	**0.0134**	**0.0017**	**0.0004**	**0.0004**	**0.0034**	0.0647	**0.0002**	—

TABLE S.VI
THE COMPARISON RESULTS OF MAPE ON D4.

Month	X12-ARIMA	ETS	LSTM	NBEATS	MLP	GRNN	SVR	XGBOOST	RD-ETS+LSTM	APLF	XASX
1	5.77%	7.15%	1.33%●	5.45%	5.19%	3.20%	8.42%	10.32%	3.11%	30.58%	2.41%
2	18.13%	15.03%	8.00%	13.23%	55.47%	14.75%	65.68%	0.31%	16.92%	5.90%	0.21%
3	1.49%●	3.83%	5.12%	5.35%	4.81%	17.15%	0.25%●	33.50%	1.97%●	21.83%	2.08%
4	1.60%	2.77%	1.61%	6.37%	5.53%	7.92%	6.71%	0.00%●	2.43%	18.06%	0.34%
5	1.40%	1.22%	4.19%	4.32%	2.01%	4.04%	3.75%	17.17%	2.50%	16.19%	0.93%
6	3.19%	3.91%	1.26%	0.34%●	1.75%	1.64%	0.12%●	4.91%	1.21%	13.68%	0.98%
7	1.04%●	0.55%●	0.35%●	0.96%●	11.66%	5.80%	12.68%	9.95%	0.78%●	19.19%	3.03%
8	0.77%●	2.16%	10.08%	7.28%	16.05%	4.82%	14.22%	4.75%	1.33%●	5.05%	1.54%
9	5.40%	8.48%	20.10%	5.72%	6.07%	8.07%	6.14%	0.01%●	6.89%	1.19%	0.98%
10	1.80%	3.77%	11.37%	6.32%	6.08%	7.04%	7.00%	1.50%	3.88%	1.35%	0.12%
11	0.58%	2.37%	5.30%	3.32%	0.20%●	5.21%	0.65%	0.00%●	2.77%	2.52%	0.33%
12	16.99%	12.59%	9.99%●	8.19%●	24.56%	14.48%	22.65%	14.02%	14.64%	14.14%	12.50%
Mean-MAPE	4.84% ±6.18%	5.32% ±4.59%	6.56% ±5.72%	5.57% ±3.37%	11.62% ±15.42%	7.84% ±4.99%	12.36% ±18.03%	8.04% ±9.97%	4.78% ±5.36%	12.47% ±9.33%	**2.12% ±3.4%**
Win/Loss	9/3	11/1	9/3	9/3	11/1	12/0	10/2	9/3	9/3	12/0	**101/19**
F-rank	4.9167	5.6667	6.1667	6.0833	7.0833	7.5833	7.3333	5.6667	5.2500	7.6667	**2.5833**
***p*-value**	**0.0261**	**0.0046**	**0.0134**	**0.0105**	**0.0004**	**0.0002**	**0.0024**	**0.0081**	**0.0134**	**0.0002**	—



(a)

(b)

Fig. S.1. The result of XASX: (a) MAPE with different hyper-parameters *p* and *q*, (b) MAPE with different hyper-parameters *md* and *lr*.