# Seamless and Multi-resolution Energy Forecasting

Chenxi Wang, *Student Member, IEEE,* Pierre Pinson, *Fellow, IEEE,* Yi Wang, *Senior Member, IEEE*

***Abstract*—Forecasting is pivotal in energy systems, by providing fundamentals for operation at different horizons and resolutions. Though energy forecasting has been widely studied for capturing temporal information, very few works concentrate on the frequency information provided by forecasts. They are consequently often limited to single-resolution applications (e.g., hourly). Here, we propose a unified energy forecasting framework based on Laplace transform in the multi-resolution context. The forecasts can be seamlessly produced at different desired resolutions without re-training or post-processing. Case studies on both energy demand and supply data show that the forecasts from our proposed method can provide accurate information in both time and frequency domains. Across the resolutions, the forecasts also demonstrate high consistency. More importantly, we explore the operational effects of our produced forecasts in the day-ahead and intra-day energy scheduling. The relationship between (i) errors in both time and frequency domains and (ii) operational value of the forecasts is analysed. Significant operational benefits are obtained.**

***Index Terms*—Energy Forecasting, Machine Learning, Laplace Transform, Multi-resolution System Operation**

## I. Introduction

Energy forecasting is crucial in all segments of the energy industry. The primary goal of energy forecasting is to provide accurate information about both demand [1] and supply [2] in the future. These forecasts support decision-makers to act optimally [3]. In general, energy forecasting is a time-series prediction problem in which predictions of diverse temporal horizons and resolutions are required for different applications. Here, forecast horizon refers to how far into the future the forecasts are for, while resolution is defined as the time interval between two successive forecast points.

For instance, day-ahead energy demand and renewable energy forecasts with hourly or 15-minute resolutions are used by system operators to schedule energy generation in advance [4]. Hours-ahead forecasts at minute resolution are needed for (near) real-time energy balancing [5]. And, forecasts at second resolution levels (nowcasts) can be used for immediate control actions and grid stability analysis [6]. Therefore, energy forecasting at different horizons and resolutions is fundamental to all aspects of the operation and a management of the entire energy system. In this paper, we focus on short-term energy forecasting for electricity load and wind power generation, where the horizon is less than one day and the resolution is higher than one hour.

At one single resolution level, the energy forecasting problem has been widely studied in the context of both energy demand [7] and supply [8]. Various approaches for either single-step or multi-step forecasting situations can also be easily found in existing literature reviews [9] and books [10].

Despite the prosperity of single-resolution energy forecasting, most of the approaches merely focus on developing a mapping function between relevant temporal features and forecasting targets. However, such methods may be unsuitable for multi-resolution energy forecasting. This is since the frequency information, indicating how intense fluctuations are at various frequencies, may vary depending upon the desired forecasting resolution. Generally, the higher the resolution is, the more detailed frequency information within the energy data should be exhibited [11]. The learned forecasting model for one single resolution can overfit or underfit at other resolutions. Therefore, single model may fail to capture the precise frequency information at multiple resolutions.

Even though multiple forecasting models are trained individually for different resolutions [12], [13], there still exists challenges in terms of consistency within the temporal hierarchy of the forecasts [14]. For instance, hourly wind power data can temporally averaged by minute-resolution observations every 60-minute period. From both theoretical and practical points of view, if we forecast at both these resolutions, the forecasts generated should also respect this temporal hierarchy. Multiple models, however, by being naturally unaware of this temporal hierarchy, produce forecasts that are likely to be inconsistent. Therefore, for multi-resolution forecasting, it's no longer suitable to naively develop multiple models. By changing forecast resolution depending upon user needs, the forecasts need to seamlessly adapt, while also respecting underlying temporal hierarchies.

Some previous works aimed at tackling the above issue. For instance they focus on the post-hoc coordination of multi-resolution forecasts [15], [16], the improved training process of multiple models [17], [18], and continuous-time neural networks for time series [19]–[21]. These approaches either try to coordinate the multiple models after/during the training, or purely focus on the dynamics of the time series. They still neglect to reveal the relationship between forecasting resolutions and frequency information. Therefore, from the perspective of methodology, a unified energy forecasting model is needed to seamlessly output the hierarchical forecasts at different resolutions with the corresponding frequency information.

More importantly, energy forecasts with precise frequency information are supposed to have prominent effects on the following decision-making process [22], especially in situa-

The work was supported in part by the Research Grants Council of the Hong Kong SAR (HKU 27203723) Guangdong Basic, in part by the Applied Basic Research Foundation (2024A1515011266) and in part by the Young Elite Scientists Sponsorship Program by CAST. *(Corresponding author: Yi Wang.)*

Chenxi Wang and Yi Wang are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China (e-mail: cxwang@eee.hku.hk, yiwang@eee.hku.hk).

Pierre Pinson is with Dyson School of Design Engineering, Imperial College London, UK, Department of Technology, Management and Economics, Technical University of Denmark, Denmark and Halfspace, Denmark (e-mail: p.pinson@imperial.ac.uk).

tions of high resolution. The connection between forecasts and resulting operation results in the power system is gradually receiving attention [23]–[25], but seldom do the existing works explore the effects of errors in the frequency domain. Consequently, from the perspective of empirical analysis, the final operational effects of the forecasts in terms of both time and frequency deserve to be unveiled.

To this end, this paper aims to fill the research gaps in the context of multi-resolution energy forecasting. The main contributions are threefold.

- We proposed an innovative and unified energy forecasting framework, Hierarchical Neural Laplace (HNL) for multi-resolution energy forecasting. Given the desired resolutions, the corresponding forecasts can be seamlessly generated without re-training or post-processing.
- In the framework, inverse Laplace transform is deeply combined with Shannon sampling theorem. The former one explicitly characterizes the hierarchical frequency information of energy data, while the latter one guides how to output forecasts at different resolutions.
- Empirically, we evaluated the forecasts on both statistical accuracy and operational results of energy scheduling. The superior performance of our proposed framework is proved on both aspects. The relationship between temporal errors, frequency magnitude errors, and operational costs is revealed as well.

The rest of the paper is organized as follows. The basic problem statement of multi-resolution forecasting is introduced in Section II. Our framework HNL is presented in Section III. Case studies and numerical analysis are shown in Section IV. Finally, conclusions and future works are drawn in Section V.

## II. Problem statement

Before introducing our proposed framework, we mathematically state the setting of multi-resolution forecasting in this section. To keep our notation clear, vector variables are bold in the rest of this paper.

Given the observed energy data $\mathbf{x} = [x_{t_1}, \cdots, x_{t_n}] \in \mathbb{R}^n$ at the sampling resolution $r = n/(t_n - t_1)$, along with the external features $\mathbf{e} \in \mathbb{R}^l$ (e.g. numerical weather predictions, NWPs), we aim to produce forecasts in a forecasting horizon at $m$ ascending resolutions $r_1, \cdots, r_m$.

Here, we restrict that the largest resolution is less than the original sampling resolution, i.e. $r_m \leq r$. Otherwise, it's a super-resolution forecasting task. We also focus on the situations that the resolution $r_i, \forall i = 1, ..., m$ is a factor of $r$, i.e. $(r \mod r_i) = 0$. A ratio $K_i = r/r_i \in \mathbb{N}^+$ can be further defined to denote how many data points are needed at the original resolution $r$ to be temporally averaged to obtain the result at the resolution $r_i$.

The energy data at resolution $r$ can also be viewed as the discretely sampled points from a temporal function $f(t)$, i.e. $x_t = f(t) \in \mathbb{R}$. For each resolution $r_i$, we can denote the corresponding temporal averaged function as $f_i(t)$ as well. Therefore, given a forecasting horizon at the length of $L$, the forecasting targets at the resolution $r_i$ can be represented as $\mathbf{y}_i = [f_i(t_n + 1L/K_i), f_i(t_n + 2L/K_i), \cdots, f_i(t_n + L)] \in \mathbb{R}^{LK_i}$.

For example, given the wind power data measured every 15min in one hour, $\mathbf{x} = [x_{15}, x_{30}, x_{45}, x_{60}] = [f(15), f(30), f(45), f(60)], r = 4/\text{hr}$, we now consider how to forecast for the next hour $L = 1\text{hr}$ at resolutions $r_1 = 1/\text{hr}$, and $r_2 = 2/\text{hr}$. Then, the forecasting targets are $\mathbf{y}_1 = [f_1(60 + 60)]$ and $\mathbf{y}_2 = [f_2(60 + 30), f_2(60 + 60)]$.

Finally, given the observed energy data $\mathbf{x}$, along with the external features $\mathbf{e}$, the goal of our proposed framework is to produce multi-resolution forecasts $\hat{\mathbf{y}}_1, \cdots, \hat{\mathbf{y}}_m$ with one unified model $\mathbf{F}_{\boldsymbol{\omega}}$:

$$\hat{\mathbf{y}}_i = \mathbf{F}_{\boldsymbol{\omega}}(\mathbf{x}, \mathbf{e}; r_i), \quad \forall i = 1, \cdots, m \tag{1}$$

where $\boldsymbol{\omega}$ is the set of trainable parameters in the whole model.

## III. Hierarchical Neural Laplace

### A. Overall framework

With the clear aim of multi-resolution forecasting, we briefly introduce the proposed flexible forecasting framework, HNL, in this subsection. Unlike the traditional methods focusing on the time domain, HNL tries to learning the frequency information in the Laplace domain. Then, it utilizes the Inverse Laplace Transform (ILT) to reconstruct the temporal components with different frequency information. According to the desired forecasting resolutions, final energy forecasts can be formed seamlessly from the combination of the temporal components. As shown in Fig. 1, this framework has three core segments: an encoder, multiple Laplace decoders, and an assembler.

Firstly, the encoder is responsible for transforming observed energy data at some resolution and external features (e.g. numerical weather predictions, NWPs) into the high-dimensional representation, i.e., the hidden states. The encoder is model-agnostic, and arbitrary neural networks can be used as encoder, for example, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), etc.

Then, the encoded hidden states will go through multiple Laplace decoders and produce different temporal components. In this process, there are two sub stages: 1) generating the discrete Laplace function values on pre-determined consecutive frequency bands, and 2) converting these Laplace function values into the temporal components by Fourier-based discrete ILT. The outputs of these decoders are temporal components containing different frequency information, as for the blue curves in Fig. 1.

Finally, given the desired forecasting resolutions, the assembler will fetch groups of temporal components to form the forecasts with precise frequency information. The Shannon sampling theorem, a bridge between data resolution and frequency information, will guide the assembler to decide which groups of components need to be fetched.

In this way, HNL only needs to be trained once on the given data, and flexibly generates the energy forecasts at different resolutions that we are interested in. Details of the framework will be introduced in the following three subsections.

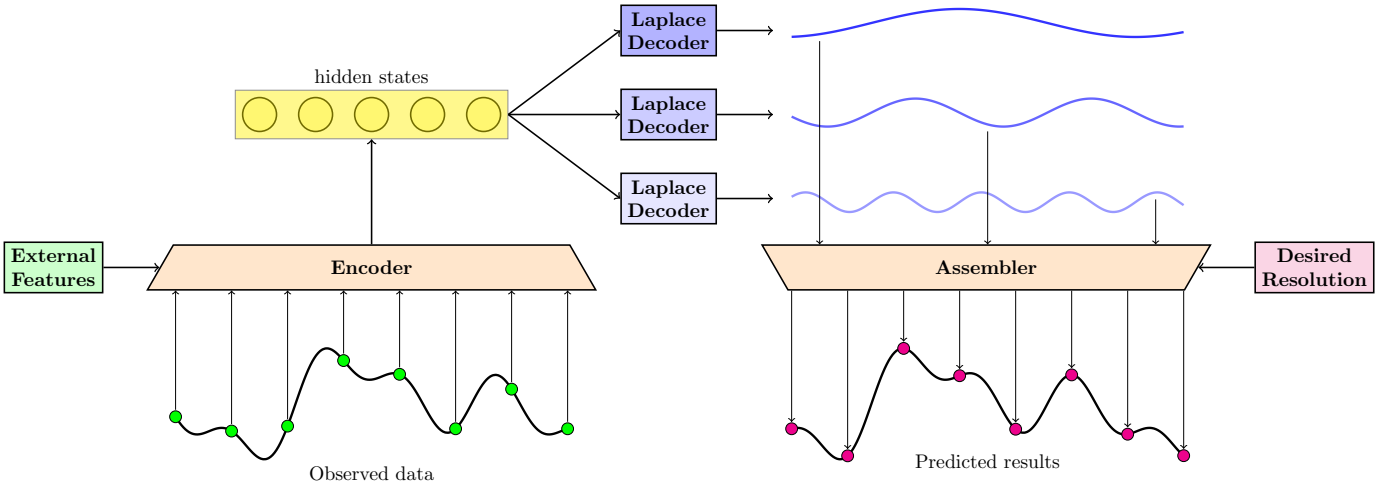



Fig. 1. The overview of the proposed Hierarchical Neural Laplace framework

### B. Encoder

As described above, the encoder $\phi_{\boldsymbol{\omega}_e}$ extracts the observed energy sequence data $\mathbf{x}$ and the external features $\mathbf{e}$ into a hidden state $\mathbf{h}$:

$$\mathbf{h} = \phi_{\boldsymbol{\omega}_e}(\mathbf{x}, \mathbf{e}) \in \mathbb{R}^d, \tag{2}$$

where $\boldsymbol{\omega}_e$ represents the trainable parameters in the encoder; $d$ is a hyperparameter, representing the number of dimension of hidden states.

In theory, the encoder can be arbitrary neural network, so the design of encoder is not the core part of this framework. In this paper, we use a GRU network and a multilayer perceptron respectively to encode the observed energy sequence and the external features. The hidden states from GRU and multilayer perceptron are concatenated to form the final hidden states shown in the Fig. 1.

### C. Multiple Laplace Decoders

*1) Laplace transform and ILT:* Laplace decoders are the core of our framework, relying on the Laplace transform and ILT. Given a temporal function $f(t)$ of energy data, the corresponding Laplace function $\bar{f}(s)$ through Laplace transform is defined as:

$$\mathscr{L}\{f(t)\} = \bar{f}(s) = \int_0^\infty e^{-st} f(t) dt. \tag{3}$$

Similarly, given a Laplace function $\bar{f}(s)$, the temporal function can be inversely transformed as its definition shows:

$$\mathscr{L}^{-1}\{\bar{f}(s)\} = f(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} \bar{f}(s) ds. \tag{4}$$

According to [26], Eq. (4) can be manipulated and discretized through the trapezoid rule:

$$\begin{aligned} f(t) &= \frac{e^{\gamma t}}{\pi} \int_0^\infty \Re\left(\bar{f}(s) e^{i\omega t}\right) d\omega \\ &\approx \frac{e^{\gamma t}}{T} \left[ \frac{\bar{f}(\gamma)}{2} + \sum_{k=1}^{N} \Re\left\{ \bar{f}\left(\gamma + \frac{ik\pi}{T}\right) \exp\left(\frac{ik\pi t}{T}\right) \right\} \right], \end{aligned} \tag{5}$$

where $T$ and $\gamma$ are usually set as constants [27], and $N$ is the number of discrete integral terms, dominating the discretization of ILT.

Further, we rewrite the Eq. (5) with the Eular formula, and obtain the following form:

$$\begin{aligned} f(t) = \frac{e^{\gamma t}}{T} \Bigg[ &\frac{\bar{f}(s_0)}{2} \\ &+ \sum_{k=1}^{N} \Re\left\{ \bar{f}(s_k) \left( \cos(\frac{k\pi t}{T}) + i \sin(\frac{k\pi t}{T}) \right) \right\} \Bigg] \end{aligned} \tag{6}$$

where for simplicity, we denote $s_k = \gamma + \frac{ik\pi}{T} \in \mathbb{C}$ and $\bar{\mathbf{f}}(\mathbf{s}) = [\bar{f}(s_0), \bar{f}(s_1), \cdots, \bar{f}(s_N)] \in \mathbb{C}^{N+1}$.

We observe that the inversely transformed $f(t)$ explicitly consists of cosine waves with exponential effects at different frequencies:

$$f_c^k = \frac{k\pi}{T \cdot 2\pi} = \frac{k}{2T}. \tag{7}$$

Consequently, if we are given the discrete Laplace function values $\bar{\mathbf{f}}(\mathbf{s})$, the original $f(t)$ can be reconstructed and an energy forecast $f(t_{n+1})$, for example, can be generated accordingly with explicit frequency information.

*2) Two-stage decoding:* Therefore, the general idea of Laplace decoders is first outputting the Laplace function values $\tilde{\mathbf{f}}(\mathbf{s})$ of the energy sequence on the forecasting horizon, and then reconstructing a temporal function $\hat{f}(t)$ for forecasting. Here, we use $\tilde{\mathbf{f}}$ and $\hat{f}$ to distinguish from the real $\bar{\mathbf{f}}$ and $f$ on the forecasting horizon.

In the first stage, we construct $p$ neural networks $g_{\boldsymbol{\omega}_1}, \cdots, g_{\boldsymbol{\omega}_p}$ to separately learn the Laplace functions on $p$ consecutive frequency bands:

$$\begin{cases} \tilde{\mathbf{f}}(\mathbf{s}_1) = g_{\boldsymbol{\omega}_1}(\mathbf{s}_1, \mathbf{h}), & \mathbf{s}_1 = [s_0, s_1, \cdots, s_{N_1}] \\ \tilde{\mathbf{f}}(\mathbf{s}_2) = g_{\boldsymbol{\omega}_2}(\mathbf{s}_2, \mathbf{h}), & \mathbf{s}_2 = [s_{N_1+1}, \cdots, s_{N_2}], \\ & \vdots \\ \tilde{\mathbf{f}}(\mathbf{s}_p) = g_{\boldsymbol{\omega}_p}(\mathbf{s}_p, \mathbf{h}), & \mathbf{s}_p = [s_{N_{p-1}+1}, \cdots, s_N], \end{cases} \tag{8}$$

where $\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_p$ are trainable parameters in these neural networks; $N_1 \leq \cdots \leq N_{p-1} (\leq N)$ are $p-1$ split points set manually.

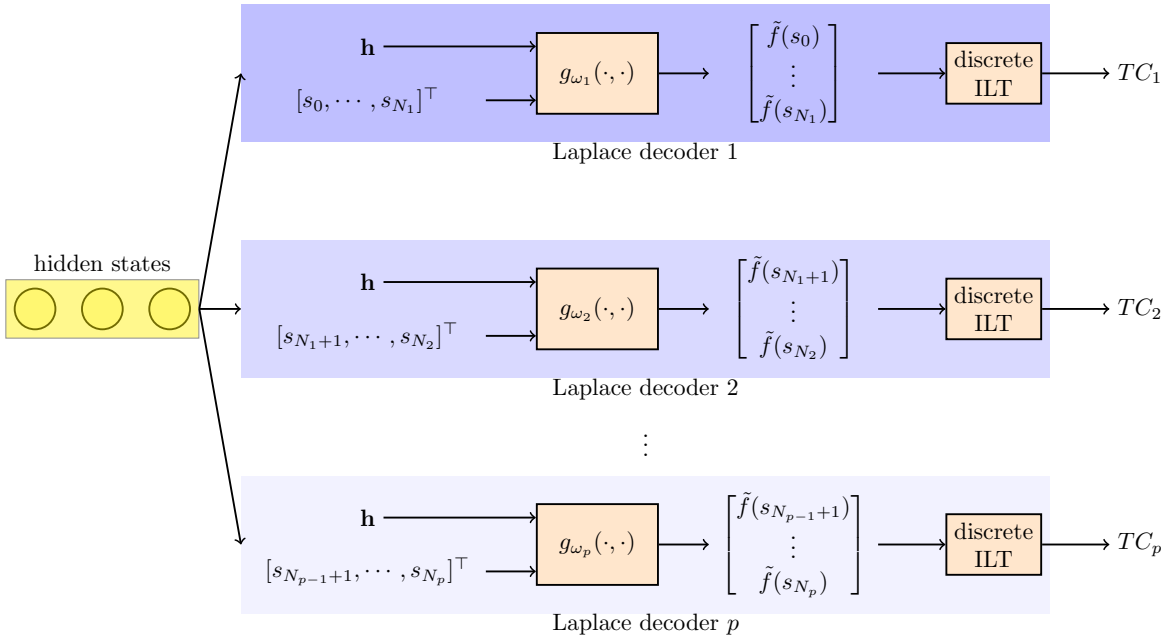



Fig. 2. The details of multiple Laplace decoders

All the neural networks share the extracted information of hidden states $\mathbf{h}$ and differ in the corresponding frequency bands. Since $\mathbf{s}_i, \tilde{\mathbf{f}}(\mathbf{s}_i) \in \mathbb{C}^{N_i - N_{i-1}+1}$ are complex vectors, we first split the values of the real and the imaginary parts, and then concatenate them together as inputs (and outputs) for decoders.

In other words, the input and output space of $g_{\boldsymbol{\omega}_i}$ are respectively $\mathbb{R}^{2(N_i-N_{i-1}+1)+l}$ and $\mathbb{R}^{2(N_i-N_{i-1}+1)}$. The architecture of these decoders are conventional multi-layer perceptrons.

In the second stage, the learned Laplace function values $\tilde{\mathbf{f}}(\mathbf{s}_1), \cdots, \tilde{\mathbf{f}}(\mathbf{s}_p)$ are fed into discrete ILT to reconstruct the temporal function. Further, the reconstructed $\hat{f}(t)$ can be presented by the combination of several temporal components:

$$\hat{f}(t) = \frac{e^{\gamma t}}{T}\Bigg[\underbrace{\frac{\tilde{f}(s_0)}{2} + \sum_{k=1}^{N_1} S_k(t)}_{TC_1} + \underbrace{\sum_{k=N_1+1}^{N_2} S_k(t)}_{TC_2} + \cdots + \underbrace{\sum_{k=N_{p-1}+1}^{N_p} S_k(t)}_{TC_p}\Bigg] \tag{9}$$

where we denote $S_k(t) = \Re\left\{\tilde{f}\left(s_k\right)\left(\exp\left(\frac{ik\pi t}{T}\right)\right)\right\}$ for simplicity and $TC_i$ represents the $i$th temporal component. The whole tow-stage decoding of these multiple Laplace decoders is demonstrated by Fig. 2.

### D. Assembler

After rebuilding a temporal function $\hat{f}(t)$ with explicit frequency information for the forecasting horizon, the assembler aims to collect the components and generate the forecasts at the desired resolutions $r_1, \cdots, r_m$.

According to Nyquist–Shannon sampling theorem [28], for a given sampling resolution $r$, the maximum informative frequency that can be expressed in the original energy data is no more than $1/2r$, i.e., $f_c^{max} \leq 1/2 \cdot r$. Note that in Eq. (7), the maximum informative frequency in the energy sequence data is $f_c^{max} = N/2T$. Therefore, if we reconstruct the temporal function through (5) without information loss, then the discretization parameter $N$ can be set as:

$$f_c^{max} = \frac{N}{2T} \leq \frac{1}{2} \cdot r \Rightarrow N \leq T \cdot r. \tag{10}$$

It tells us that by setting $N = Tr$, we can include all the necessary information at the resolution $r$. Also, if we need to forecast at a lower resolution, for instance $r_i, \forall i = 1, \cdots, m$, with precise frequency information, then we can just set $N$ to $Tr_i$.

Therefore, before collecting the components, we can first set the number $p$ and positions $N_i$ of the split points on the frequency band according to the desired resolutions:

$$\begin{cases} p = m, \\ N_i = Tr_i, \forall i = 1, \cdots, m. \end{cases} \tag{11}$$

Then, for the resolution $r_i$, $TC_1, \cdots, TC_i$ are the temporal components with necessary frequency information. In assembler, a corresponding temporal function $\hat{f}_i(t)$ is formed accordingly:

$$\hat{f}_i(t) = \frac{e^{\gamma t}}{T}\Big[TC_1 + TC_2 + \cdots + TC_i\Big]. \tag{12}$$

The energy forecasts at the resolution $r_i$ can be inferred by time steps as we introduction in Section II:

$$\hat{\mathbf{y}}_i = [\hat{f}_i(t_n + 1L/K_i), \cdots, \hat{f}_i(t_n + L)]. \tag{13}$$

**Algorithm 1** Training and inference of HNL

**Require:** training dataset $D_{train}$, epochs $E$, resolutions $r_1, \cdots, r_m$, ILT parameters $\gamma, T, N$

// *training process*
set splitting points $N_1, \cdots, N_{p-1}$ through Eq. (11)
initialize encoder $\phi_{\boldsymbol{\omega}_e}$ and decoders $g_{\boldsymbol{\omega}_1}, \cdots, g_{\boldsymbol{\omega}_p}$
prepare $\mathbf{s} = [\gamma + \frac{ik\pi}{T} \quad \text{for} \quad k = 0 \quad \text{to} \quad N]$
split $\mathbf{s} = \mathbf{s_1}, \cdots, \mathbf{s_p}$ according to $N_1, \cdots, N_{p-1}$.
**for** $i = 1$ to $E$ **do**
  **for** batch data $(\mathbf{x}, \mathbf{e}, \mathbf{y_1}, \cdots, \mathbf{y_m})$ in $D_{train}$ **do**
    $\mathbf{h} \leftarrow \phi_{\boldsymbol{\omega}_e}(\mathbf{x}, \mathbf{e})$
    $loss \leftarrow 0$
    **for** $j = 1$ to $m$ **do**
      $\tilde{\mathbf{f}}(\mathbf{s}_j) = g_{\boldsymbol{\omega}_j}(\mathbf{s_j}, \mathbf{h})$
      reconstruct to $\hat{f}_j(t)$ through Eq. (12)
      make forecasts $\hat{\mathbf{y}}_j$ through Eq. (13)
      $loss \leftarrow loss + \|\hat{\mathbf{y}}_j - \mathbf{y}_j\|_2^2$
    **end for**
    $loss \leftarrow loss/m$
    calculate gradients and update encoder and decoders.
  **end for**
**end for**

// *inference process*
sample $\mathbf{x}$, $\mathbf{e}$ from $D_{test}$
$\mathbf{h} \leftarrow \phi_{\omega_e}(\mathbf{x}, \mathbf{e})$
**for** $j = 1$ to $m$ **do**
  $\tilde{\mathbf{f}}(\mathbf{s}_j) = g_{\omega_j}(\mathbf{s_j}, \mathbf{h})$
  reconstruct to $\hat{f}_j(t)$ through Eq. (12)
  make forecasts $\hat{\mathbf{y}}_\mathbf{i}$ through Eq. (13)
**end for**

### *E. Loss function*

Since we have produced the energy forecasts for resolutions $r_1, \cdots, r_m$, we define the loss function used for training as follows:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}; \boldsymbol{\omega}) = \frac{1}{|D_{train}|} \sum_{i=1}^{|D_{train}|} \frac{1}{m} \sum_{j=1}^{m} \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2, \qquad (14)$$

where $D_{train}$ is the training dataset, and $\boldsymbol{\omega}$ is the collection of trainable parameters in the whole framework, i.e. $\boldsymbol{\omega} = \{\boldsymbol{\omega}_e, \boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_m\}$. Thus, the loss function we used for training is the averaged loss across $m$ resolutions, and any gradient-based optimizer can be applied. The complete training and inference process can be referred to Algorithm. 1.

After training, one can generate multi-resolution forecasts from $r_1$ to $r_m$ accordingly. Besides, if the model user would like to forecast at some resolution, $r' \leq r$, which is not included in the set $\{r_i, \forall i = 1, \cdots, m\}$, we can also calculate the necessary integral terms $N' = T \cdot r'$. Then, the Laplace function values will be truncated accordingly, and the predicted temporal function at resolution $r'$ will be:

$$\hat{f}'(t) = \frac{e^{\gamma t}}{T} \left[ \frac{\tilde{f}(s_0)}{2} + \sum_{k=1}^{N'} S_k(t) \right]. \qquad (15)$$

Therefore, the forecast at resolution $r'$ which is not included can also be produced without retraining.

In short, the important takeaways from our HNL framework are:

1) We try to learn the patterns in the Laplace domain, a 2-D complex plane, instead of the traditional time domain, a 1-D real plane.
2) The main advantage of learning in the Laplace domain is that the frequency information is explicitly characterized.
3) Combining Shannon sampling theorem, given desired resolutions, we are able to choose the necessary frequency in the Laplace domain, and rebuild it back to the time domain to produce the final forecasts.

## IV. Case Studies

### *A. Data and benchmarks*

We focus on multi-resolution forecasting case studies on both Multifamily Residential Electricity Dataset (MFRED) [29] and wind power data of one single site from NREL [30], along with NWP data from ECMWF [31]. MFRED records a whole year data (2019-01 to 2019-12) in the 5-minute resolution. NREL includes 6 months data (2012-7 to 2012-12) in the same resolution. Temperature and wind speed data are downloaded for consistent time periods mentioned above.

For both load and wind power datasets, we split according to the ratio of 8:1:1 to build training, validation and test datasets. We produce forecasts at three commonly used resolutions, namely 5-minute, 15-minute, and 60-minute, for the next 24 hours.

Prevailing ML-based energy forecasting models are set as benchmarks, including Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM), Gradient Boosting Regression Tree (GBTR), Temporal Fusion Transformer (TFT) [32], DLinear [33] and NBEATSx [34].

We also include the original Neural Laplace (NL) [35] as the benchmarks. It only has one Laplace decoder with a small $N$ in discrete ILT. It should be mentioned that the curse of dimensionality may happen if we only use one decoder with a large enough $N$ to cover the whole frequency domain. The ineffectiveness in such case can be found in the appendix. One naive model, Persistence, is included as the baseline model. The latest observed data is used as forecast, for all lead times.

Besides, two widely used coordination strategies for multi-resolution forecasts [14], i.e., bottom-up-based strategy (BU) and optimized-based strategy (OPT), are also applied to each of the benchmark models. Therefore, for each resolution, we have three types of benchmark forecasts:

- Raw forecasts from: MLP, LSTM, NL, and Persistence,
- BU-based coordinated forecasts from: MLP-BU, LSTM-BU, and NL-BU,
- OPT-based coordinated forecasts from: MLP-OPT, LSTM-OPT, and NL-OPT

The BU strategy means that we only forecast on the highest resolution, and the forecasts at the lower resolutions are all generated by downsampling. The OPT strategy makes full use

TABLE I
IMPORTANT PARAMETERS

| Method | Structure Parameters | Training Parameters |
|---|---|---|
| MLP | n_fc_layer=1, hidden_units=42 | batch_size=128, optimizer=Adam, learning_rate=1e-3, epochs=300, early_stop_rounds=30 |
| LSTM | n_lstm_layer=2, n_fc_layer=1, hidden_units=42 | |
| NL | encoder: RNN, n_layer=1, hidden_units=42, $N$=33 | |
| | decoder: MLP, n_fc_layer=2, hidden_units=42 | |
| HNL | encoder: RNN, n_layer=1, hidden_units=42 | |
| | decoder: MLP * 3, n_fc_layer=2, hidden_units=42 | |

of forecasts at all desired resolutions, solving a generalized least squares problem with the forecasts at resolutions and the temporal structure as inputs. We adopted the method of structure scaling to determine the covariance matrix in the generalized least squares problem. The theoretical details of these two coordination approaches can be referred to [14]. The important parameters, including structure and training parameters are list in Table. I.

The load[1] and wind power[2] datasets are both available online. The NWP data is available from the ECMWF website[3] after registration as real-name users (for researchers based in Europe, at least). The code and data for the experiments are publicly available on https://github.com/hkuedl/Multi-resolution-Energy-Forecasting.

### B. Performance metrics

The performance metrics used for evaluating the energy forecasts include RMSE in the time domain, RMSE in the frequency domain, Total Consistency Error (TCE), and the total operation costs in the energy scheduling.

In the test set $D_{\text{test}}$, we denote the $j$th forecast at the resolution $r_i$ on the following steps $K_i$ are denoted as $\hat{\mathbf{y}}_{i,j}$. The corresponding real energy data at this resolution is denoted as $\mathbf{y}_{i,j}$. Therefore, the average RMSE in the time domain for this resolution $r_i$ over the whole test set can be calculated as:

$$RMSE^{i}_{\text{time}} = \frac{1}{|D_{\text{test}}|}\sum_{j=1}^{|D_{\text{test}}|}\|\hat{\mathbf{y}}_{i,j} - \mathbf{y}_{i,j}\|_2, \tag{16}$$

where $|D_{\text{test}}|$ denotes the size of the test set.

For the RMSE in the frequency domain, we perform fast Fourier transform [36] (FFT) on for each real energy curve and corresponding forecast at the highest resolution (5-minute in this paper). Then, we take the magnitude of the results of FFT. The resulting frequency magnitudes of the $j$th real curve and the forecast are denoted as $\mathbf{z}_{i,j}$ and $\hat{\mathbf{z}}_{i,j}$. In this way, we can compute the average RMSE in the frequency domain over the whole test set:

$$RMSE_{\text{freq}} = \frac{1}{|D_{\text{test}}|}\sum_{j=1}^{|D_{\text{test}}|}\|\hat{\mathbf{z}}_{i,j} - \mathbf{z}_{i,j}\|_2, \tag{17}$$

The unit of both RMSEs in both time and frequency domain is the same as the original energy data, i.e. kW in our case studies.

In terms of evaluating forecasting consistency, the TCE over the whole test set is calculated as follows:

$$TCE = \frac{1}{|D_{\text{test}}|}\sum_{j=1}^{|D_{\text{test}}|}\sum_{a=1}^{m}\sum_{b=a+1}^{m}\|ds(\hat{\mathbf{y}}_{b,j}, r_a) - \hat{\mathbf{y}}_{a,j}\|_2^2, \tag{18}$$

where the function $ds(\cdot,\cdot)$ takes in a high-resolution forecast at $r_b$ and downsamples them into a low-resolution one at $r_a(\leq r_b)$. Then, the Euclidean distance of the downsampled forecasts and the original forecasts at the resolution $r_a$ will be calculated. For each resolution pair $(r_a, r_b)$, we calculate and sum up to obtain the final TCE of the $j$th forecasts. The average TCE over the whole test set is then calculated as above.

As for energy scheduling, we consider an economic dispatch problem in both day-ahead and intra-day fashion, including the ramping constraints of generators, battery storage devices and so on. The details of the optimization problems can be referred to the appendix in [24].

### C. Accurate forecasts at different resolutions

To comprehensively evaluate forecast performance, we run each forecast model from scratch 20 times with different random seeds. Then, RMSEs are calculated for each seed. The comparison of average RMSEs with standard deviations over all seeds is depicted in Table. II and Table. III. For each column, if our proposed HNL beats the raw forecasts from ALL benchmarks, the results will be **bold**. If our proposed method even beats the both BU and OPT coordinated forecasts from ALL benchmarks, the results will be added with $*$. The results of GBRT, TFT, DLinear and NBEATSx are included in the appendix.

For electricity load forecasting, the proposed HNL framework outperforms the baseline model by 7.28%, 7.08%, and 3.4%, respectively, at 5-minute, 15-minute, and 60-minute. The advantage of HNL is more significant with the forecasting resolution increases. At the highest resolution, HNL beats all the uncoordinated benchmarks, improving by at least 4.7%. Even though BU and OPT strategies generally lower the forecasting errors of benchmarks at all resolutions, HNL still shows superiority over most of them and achieves competitive performance with the best coordinated model (MLP-OPT).

For wind power forecasting, the performance differences are even more remarkable. Outperforming the baseline Persistence by at least 29%, the proposed HNL obtains the lowest forecasting error at all resolutions. It obtains the lowest forecasting error over all benchmarks trained individually, improving by at least 5.3% on average. Compared to coordinated benchmarks, HNL still has at least a 3% accuracy improvement over the

[1]https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X9MIDJ

[2]https://www.nrel.gov/grid/wind-toolkit.html

[3]https://www.ecmwf.int/en/forecasts/dataset/operational-archive

TABLE II
RMSE COMPARISON ON LOAD DATASET ($\times 10^{-2}$ KW)

| Method | 5min | 15min | 60min |
|---|---|---|---|
| LSTM | 3.751 ± 0.142 | 3.809 ± 0.192 | 3.746 ± 0.159 |
| LSTM-BU | 3.751 ± 0.142 | 3.708 ± 0.144 | 3.622 ± 0.146 |
| LSTM-OPT | 3.597 ± 0.099 | 3.552 ± 0.100 | 3.463 ± 0.102 |
| MLP | 3.591 ± 0.067 | 3.331 ± 0.114 | 3.130 ± 0.040 |
| MLP-BU | 3.591 ± 0.067 | 3.547 ± 0.069 | 3.462 ± 0.071 |
| MLP-OPT | 3.372 ± 0.060 | 3.325 ± 0.061 | 3.236 ± 0.062 |
| NL | 3.638 ± 0.129 | 3.579 ± 0.164 | 3.503 ± 0.150 |
| NL-BU | 3.638 ± 0.129 | 3.595 ± 0.131 | 3.501 ± 0.132 |
| NL-OPT | 3.531 ± 0.099 | 3.486 ± 0.100 | 3.390 ± 0.101 |
| Persistence | 3.697 ± 0.000 | 3.615 ± 0.000 | 3.479 ± 0.000 |
| HNL | **3.428 ± 0.082** | 3.359 ± 0.079 | 3.360 ± 0.068 |

TABLE III
RMSE COMPARISON ON WIND POWER DATASET (KW)

| Method | 5min | 15min | 60min |
|---|---|---|---|
| LSTM | 5.189 ± 0.097 | 5.308 ± 0.196 | 5.403 ± 0.160 |
| LSTM-BU | 5.189 ± 0.097 | 5.177 ± 0.097 | 5.118 ± 0.098 |
| LSTM-OPT | 5.174 ± 0.086 | 5.162 ± 0.086 | 5.103 ± 0.087 |
| MLP | 5.149 ± 0.091 | 4.928 ± 0.119 | 4.680 ± 0.087 |
| MLP-BU | 5.149 ± 0.091 | 5.126 ± 0.090 | 5.063 ± 0.090 |
| MLP-OPT | 5.028 ± 0.075 | 5.004 ± 0.073 | 4.941 ± 0.074 |
| NL | 4.937 ± 0.380 | 4.972 ± 0.267 | 4.999 ± 0.349 |
| NL-BU | 4.937 ± 0.380 | 4.930 ± 0.381 | 4.869 ± 0.384 |
| NL-OPT | 4.813 ± 0.312 | 4.805 ± 0.312 | 4.743 ± 0.316 |
| Persistence | 6.596 ± 0.000 | 6.604 ± 0.000 | 6.640 ± 0.000 |
| HNL | **4.642 ± 0.252**$^*$ | **4.617 ± 0.248**$^*$ | **4.538 ± 0.251**$^*$ |

best-performing one, showing the superiority on wind power forecasting.

Thus, on both load and wind power forecasting, the competitive performance of HNL is obvious in terms of forecast quality in the time domain at all resolutions.

In addition to the performance in the time domain, we also show that our proposed method can capture more precise information in the frequency domain. Table. IV collects the RMSEs on frequency magnitudes at the highest resolution. BU-based forecasts are omitted because they are the same as the raw forecasts at the highest resolution.

HNL attains the best performances in the frequency domain for both datasets, owing to the hierarchical frequency learning. With the limited capability of frequency learning, there is at least 5% performance lag for NL-based methods. Despite the close performance of MLP-OPT with HNL in the load forecasting, MLP-based models fail to accurately portray the actual frequency information of wind power. Similarly, mainly focusing on the temporal dependence, LSTM-based models cause around 15% more error in the frequency domain than our proposed HNL. Therefore, besides the competitive forecasting performance in the time domain, our proposed method can capture the information in the frequency domain more accurately than the prevailing forecasting models.

TABLE IV
RMSE COMPARISON ON FREQUENCY DOMAIN (MAGNITUDE)

| Method | Load ($\times 10^{-2}$) | Wind power |
|---|---|---|
| LSTM | 58.771 ± 3.180 | 88.487 ± 2.710 |
| LSTM-OPT | 56.694 ± 2.297 | 88.548 ± 2.412 |
| MLP | 55.202 ± 1.805 | 92.625 ± 2.526 |
| MLP-OPT | 51.102 ± 1.508 | 90.352 ± 2.025 |
| NL | 55.665 ± 2.403 | 82.677 ± 7.025 |
| NL-OPT | 54.430 ± 1.602 | 80.987 ± 5.703 |
| HNL | **50.003 ± 1.359**$^*$ | **76.477 ± 4.576**$^*$ |

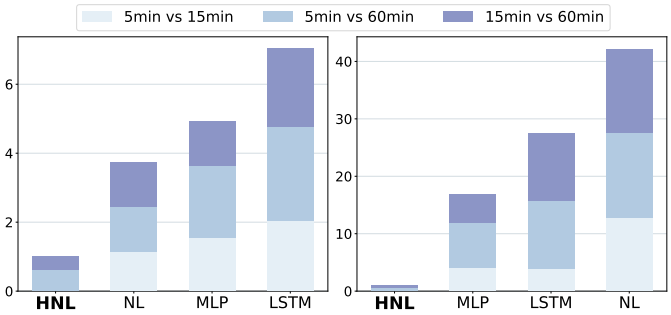



Fig. 3. TCE comparison. The TCEs are scaled based on the performance of the HNL. Left: electricity load forecasts. Right: wind power forecasts

### D. Consistent forecasts across resolutions

Beyond the accuracy at each resolution, we also evaluate the consistency of the forecasts across these resolutions. Fig. 3 presents the comparison of our proposed HNL and uncoordinated models in terms of consistency. The most obvious inconsistency happens between the highest and the lowest resolution, i.e., 5-minute and 60-minute. The proposed HNL brings significant advantages in both electricity load and wind power forecasting. In contrast, the raw forecasts from the benchmarks without any coordination apparently cause extremely high consistency errors.

Besides, the leading position of HNL in wind power forecasting is more striking than electricity load forecasting. Our proposed HNL is at least 10 times better than the benchmarks in the wind power forecasting case. Fig. 4 depicts an example of multi-resolution forecasts comparison for wind power generation.

From the perspective of accuracy, traditional deep learning-based models, LSTM and MLP, fail to produce accurate forecasts in terms of both time and frequency domains. Especially in high-resolution situations, forecasts from MLP-related models introduce redundant high-frequency information. Though the forecasts from NL-related models track the trend of wind power, they fail to represent the information across relevant frequency ranges, owing to the limited frequency learning capability. These phenomena reflect the conclusion that the HNL method yield more accurate forecasts when assessed in both time and frequency domains.

From the perspective of consistency, there appears abrupt gaps during the transition of forecasting resolutions among raw forecasts from benchmarks. For example, forecasts from

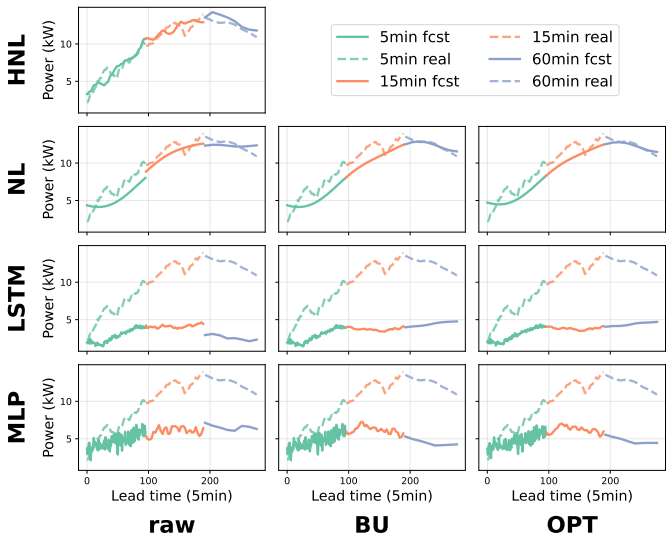



Fig. 4. Multi-resolution forecasts comparison on wind power. One forecasting day is separated into three even horizons for demonstration. Each row represents the model type, and each column represents the coordination strategy. The proposed HNL belongs to raw forecasting methods since it doesn't experience any coordination.

LSTM somehow become lower if the forecasting resolution is switched from 15-minute to 60-minute. Both post-coordination strategies (BU and OPT) eliminate these jumps in the raw forecasts, enhancing the consistency of the benchmark approaches. In contrast, even without any coordination strategies, the forecasts from HNL can still behave consistently across resolutions. The transition from one resolution to another is natural because of the framework design that generates energy forecasts at different resolutions hierarchically. Therefore, beyond the high accuracy in both time and frequency domains, forecasts from HNL are highly consistent across resolutions.

### *E. Supportive effects for decision making*

The final goal of energy forecasting is to support decision-making, for a range of operational problems. Besides the accuracy and consistency of the forecasts, we also analyze the operational benefits brought by the forecasts in the downstream decision-making process.

Specifically, we start with the day-ahead energy scheduling problem for a local region. It is a cost minimization problem that takes the 24-hour load forecast as the demand to be met, and outputs the optimal schedule of energy supply for each time step for the next day. Though the scheduling horizon is fixed to 24 hours, the scheduling resolution can vary, depending on how detailed the schedules are expected. Consequently, for different scheduling resolutions, the forecasts with the corresponding resolutions are required. Since it is impossible to forecast perfectly, additional real-time operations like external energy purchases are needed to balance the real energy demand and scheduled energy supply. This will cause extra operation costs. Therefore, regarding the extra costs, we evaluate the operational effects of the day-ahead forecasts at each resolution.

Fig. 5(a)(b)(c) illustrate the comparison of operational costs from all forecasting models at different resolutions. We observe that with the resolution becoming higher, the advantage of our proposed HNL is more remarkable. Compared to the baseline model, the percentage savings of the operational costs are 14.8%, 19.9%, and 20.6% at the resolutions of 60-minute, 15-minute, and 5-minute, respectively. For the 60-minute resolution, the prevailing machine-learning model MLP even achieves lower operation costs. The possible reason is that hourly day-ahead forecasts only require 24 prediction points. The number of output dimensions is relatively small and contains limited frequency information. Traditional machine learning methods can perform well in such low-dimension forecasting settings.

However, when it comes to the higher resolution cases, the energy curves will reveal more detailed frequency information. It may also affect the operation results probably because when the scheduling resolution is higher, the ramping constraints of generators are tighter and more strict. Then, if the forecasts introduce redundant frequencies, more frequent battery charging and discharging operations will happen to compensate.

We collected each day-ahead forecast from all the models and computed the corresponding temporal RMSE, frequency magnitude RMSE, and total operational cost. Fig. 5 (d) depicts that forecasting errors in both time and frequency domains generally follow a linear relationship. Lower errors jointly in the time and frequency domains lead to lower additional operation costs. The blue shaded part shows a large proportion of special situations. The temporal errors of the forecasts are at a similar level, but the frequency magnitude errors varies a lot. In this vertical slice, we observe that with similar temporal accuracy, generally, the lower the errors in the frequency domain are, the lower the additional operational costs are in the day-ahead scheduling. Therefore, while our proposed HNL approach achieved similar temporal performance with the state-of-the-art benchmarks, the performance advantage of HNL in the frequency domain makes it substantially better in supporting decision-making.

In addition to day-ahead scheduling, energy system operators launch intra-day scheduling on a regular basis (for example, every 4 hours) to modify the day-ahead schedules. Distinct from day-ahead scheduling, intra-day scheduling concentrates on the operation decisions for a shorter horizon (usually 4 hours) and higher resolution (at least 15-minute). This is also a cost minimization problem which requires the updated energy forecasts at the according horizon and resolution.

Here, to conduct a more realistic case study, we jointly consider the day-ahead and intra-day scheduling, and denote as integrated scheduling. Apparently, in this case, multi-resolution energy forecasts cooperatively contribute to the final operation decisions. To quantify the value of multi-resolution forecasts in this context, we calculate the total operation costs, including the day-ahead costs for energy arrangements, the intra-day costs for adjustments, and the real-time costs for energy balancing.

To fully illustrate the operational effects of both energy demand and supply forecasts, we enumerate all the possible pairs of forecasting approaches for electricity load and wind

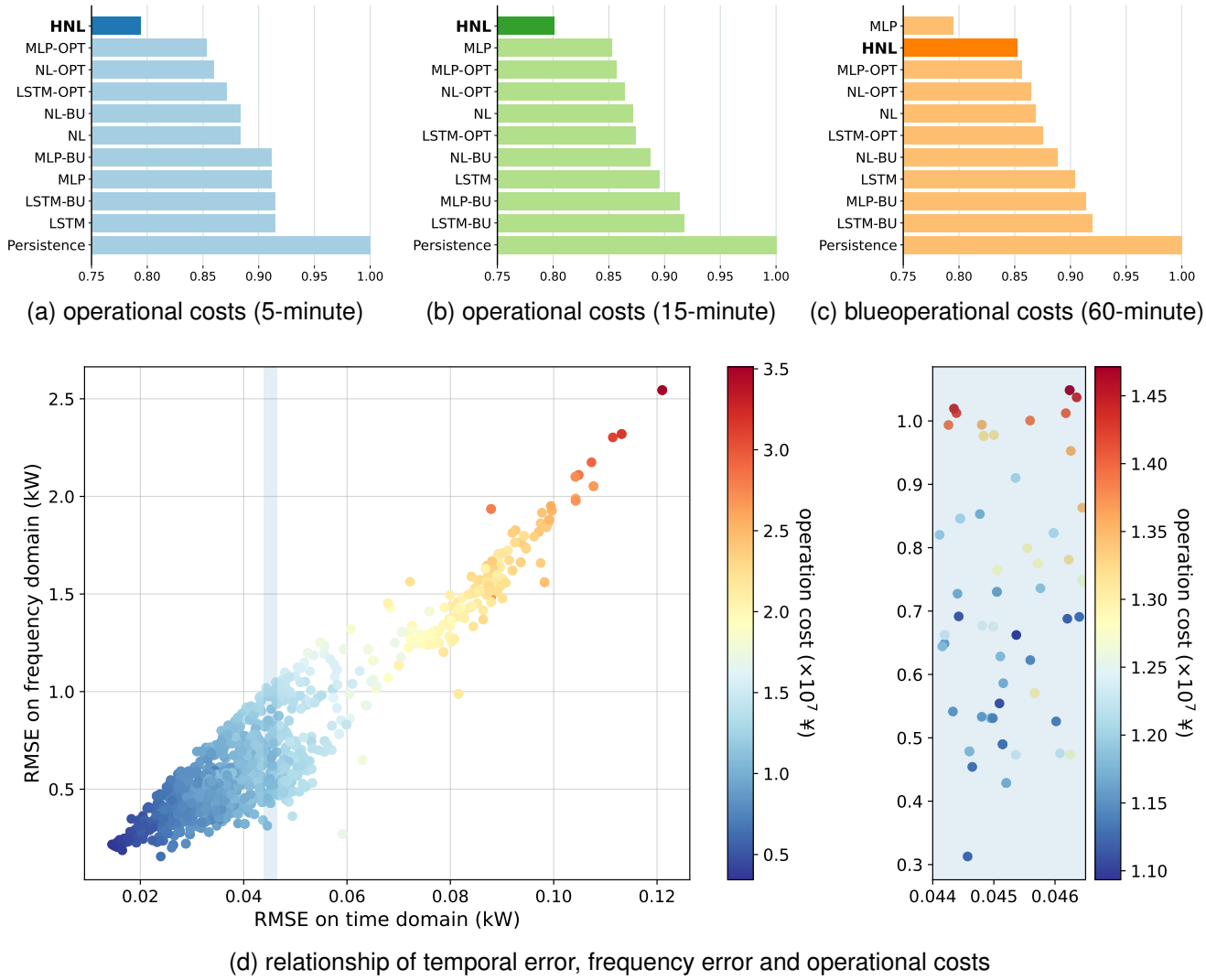



Fig. 5. Comparison of total operational costs in day-ahead scheduling. (a),(b) and (c) show the comparison based on 5-minute, 15-minute, and 60-minute resolution forecasts, respectively. The operational costs in (a), (b), and (c) are scaled based on the performance of the Persistence model. (d) shows the relationship among temporal error, frequency error, and operational costs at the highest resoltuion.

power generation, and calculate the resulting total operational costs. Fig. 6 presents the results in integrated scheduling of different model pairs. It is proved that our proposed HNL brings significant supportive effects in this large-scale realistic integrated scheduling case. Compared with the Persistence-based methods, the saved costs are more than 25%. We also observe that, in this case, the differences between each column were remarkable, which means the wind forecasting methods dominate in the integrated scheduling. This is probably because the penetration rate of the wind power is set as a relatively high ratio (50%). The results with other wind power penetration rates can be found in the appendix.

Therefore, with high accuracy in both the time and frequency domain, our proposed energy forecasting framework, HNL, can provide more supportive forecasts for day-ahead scheduling. The relationship between temporal error, frequency error, and operation costs is explored as well. The striking advantage of HNL continues to be effective in the integrated scheduling process.

## V. Discussion

In this section, we will discuss some key points based on our methodology and the empirical case studies.

### A. Learning in the Laplace domain

Frequency information of energy data is explicitly condensed in the Laplace domain. This is the reason why we design our framework in this domain. Fourier transform (FT), though also able to conduct time-frequency transform, is actually a subset of Laplace transform. The result of FT falls into the imaginary axis in the 2D complex frequency plane. Beyond FT, the Laplace transform provides the damping effects as well, which is helpful for modeling energy data with complicated properties, like wind power data. Therefore, learning in the Laplace domain provides more capability than traditional FT.

Meanwhile, theoretically, the Laplace function values $\bar{\mathbf{f}}(\mathbf{s})$ may encounter singularities, which are related to the poles

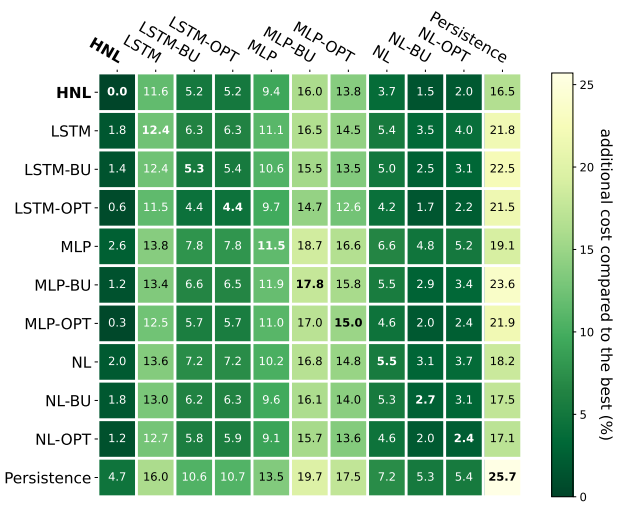



Fig. 6. Comparison of operational costs for integrated scheduling. Columns: wind forecasting methods. Rows: electricity load forecasting methods. Entries: additional costs compared to the best in percentage value. Zero means the most economic situation.

in the Laplace domain. Such singularities are usually difficult for neural networks to approximate. In this paper, we didn't actually run into the instability of training. This is possibly because unlike theoretical signals like sine waves, realistic energy data is not polarized so severely in the Laplace domain. The magnitudes of Laplace function values in both low and high-frequency components are informational. It makes the singularities less significant than the situation of ideal signals.

If severe singularity problems appear in practice, we can first utilize stereographic projection to transform complex numbers in the Laplace domain to a Riemann sphere (Eq. (19)) to stabilize the training process [35], and then inversely transform back to Laplace domain for ILT reconstruction (Eq. (20)):

$$(\theta, \phi) = \left( \arctan\left( \frac{\Im(s)}{\Re(s)} \right), \arcsin\left( \frac{|s|^2 - 1}{|s|^2 + 1} \right) \right) \tag{19}$$

$$s = \tan\left( \frac{\phi}{2} + \frac{\pi}{4} \right) e^{i\theta}, \tag{20}$$

where $s \in \mathbb{C}, (\theta, \phi) \in (-\pi, \pi) \times (-\pi/2, \pi/2)$. In this way, the singularities will be located at the north pole of this sphere, and the theoretical problem of singularities can be alleviated.

### B. Connection to differential equations

Laplace transform is often used to solve ordinary differential equations. Consequently, learning the Laplace function values also implies fitting differential equations with neural networks in the Laplace domain [35].

Such idea of fitting differential equations presents advantages on modelling energy data which may involve physical processes. Some previous works on wind power/speed forecasting leveraged such related models for improving the forecasting accuracy [37], [38]. This may explain why our HNL framework outperformed more significantly on the NREL dataset in the case studies.

### C. Comparison to super resolution

Though we focus on multi-resolution forecasting in this paper, it relates to the topic of super resolution (SR) of energy profiles, but some differences exist.

The aim of SR is to upsample the low-resolution data into high-resolution data for the same time period with consistency. In contrast, the aim of our paper is that given the energy data at the fixed sampling resolution, we need to produce forecasts at multiple resolutions that are less than the original sampling resolution.

For example, given the electricity load data at a 5-minute resolution for one day, SR tries to generate the data for the same day at a 1-minute level, while our HNL framework tries to generate 5-minute, 15-minute, and 60-minute level forecasts for the next day.

Although the aims are a bit different, we think it's possible to empower SR by our HNL framework or vice versa. In recent years, the research focus in SR shifted from *training different models separately for different target resolutions* [39], [40] to *arbitrary SR with one model* [41]. To support arbitrary SR by our HNL framework, we can increase the discretization parameter $N$ in (10) to include higher-frequency components. Under the guidance of different high-resolution data labels, more detailed energy profiles can be reconstructed through Laplace decoders.

On the other hand, if we are interested in forecasting beyond the original sampling resolutions, we can also first utilize some SR models to generate the high-resolution data, then apply HNL to obtain the forecasts whose resolutions are even higher than the original resolution.

## VI. Conclusion and Future Works

In this paper, an innovative energy forecasting framework, HNL, is proposed for multi-resolution forecasting, providing unified modeling at different resolutions. Without the need for multiple models, the proposed framework is designed to characterize the hierarchical frequency information of the energy data based on the Laplace transform. Forecasts at desired resolutions can be flexibly produced by aggregating corresponding temporal components based on the Shannon sampling theorem.

Case studies have demonstrated that the HNL framework attains satisfactory accuracy improvement on the competitive benchmarks and baseline model. Thanks to the hierarchical framework design, more seamless forecasts across the resolutions can be generated than benchmarks. The focus on frequency learning gives the HNL an additional advantage in the accuracy in the frequency domain where the HNL can capture more precise variation patterns of the energy data.

More importantly, in the follow-up day-ahead energy scheduling, this advantage of HNL can lead to better schedules with lower operation costs. The analysis of the relationship between temporal error, frequency error, and operation costs further implies that with similar temporal accuracy, forecasts with more precise frequency information may lead to better decisions. Consequently, it is suggested that accuracy in both time and frequency domains matters in energy forecasting,

especially for the downstream decision-making. Besides, in the complicated integrated scheduling, multi-resolution forecasts of electricity load and wind power from HNL resulted in a salient reduction in terms of the total operation costs as well.

As for future works, we have two potential directions: 1) extension to probabilistic forecasting considering the uncertainty in the frequency domain, and 2) explore the extension for arbitrary super resolution of load profile.

## APPENDICES

### A. Ineffectiveness of single Laplace decoder

We launched the experiment with one large Laplace decoder, i.e. Neural Laplace with a large enough frequency parameter $N$, trying to learn the whole spectrum of the energy data. However, the experiment turns out to be ineffective when only using one large Laplace decoder. The forecasting result can be observed in Fig. 7.

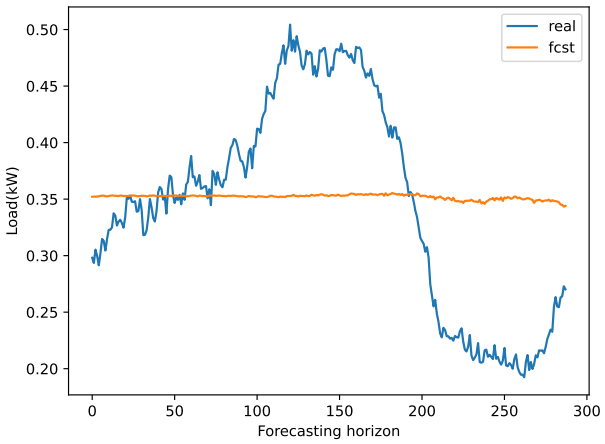



Fig. 7. Ineffectiveness when only using one large Laplace decoder

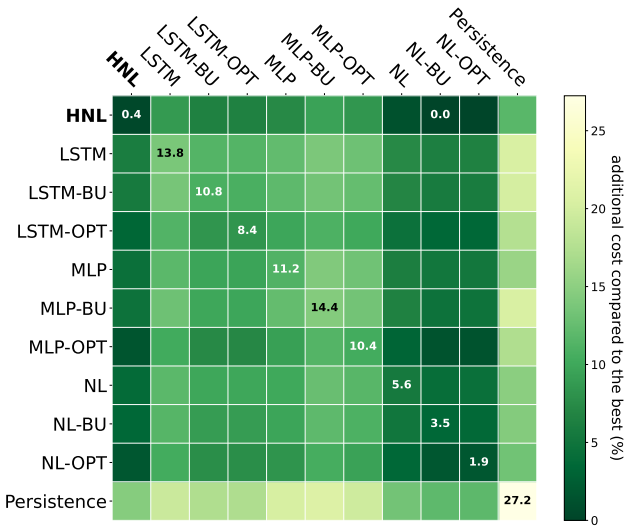
Therefore, when launching the benchmarks, we set the frequency parameter in the Neural Laplace as $N = 33$ which is the same setting in the original code. In comparison, our proposed HNL use multiple Laplace decoders to learn the frequency information in the energy data.

### B. Operation costs comparison under different penetrations

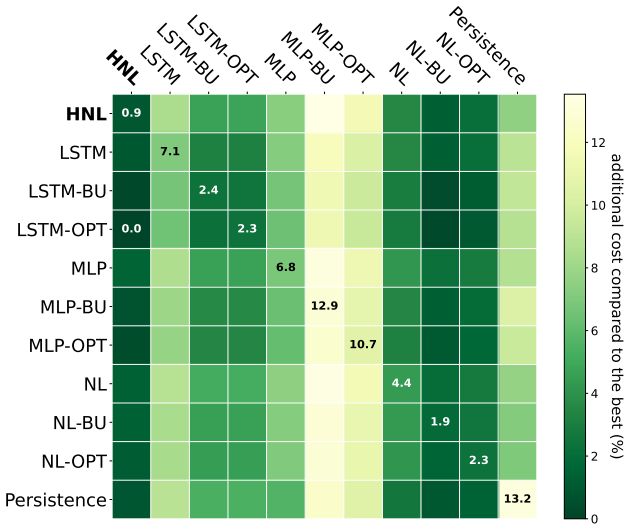
Fig. 8 shows the results under different penetration rate of wind power in the large-scale integrated scheduling problem.

When the penetration rate is low (0.2), where load forecasts dominate the decision making, the lowest operation costs are achieved in the row (load forecasting methods) of HNL.

Similarly, when the penetration rate is high (0.8), where wind power forecasts dominate, the lowest operation costs happens in the column (wind power forecasting methods) of HNL. It suggested that HNL shows dominant operational effects in all situations considered.




(a) penetration rate: 20%




(b) penetration rate: 80%

Fig. 8. Comparison of operational costs for integrated scheduling.

### C. Comparison with SOTA methods

Table. V and Table. VI respectively records the forecasting accuracy of GBRT, TFT, NBEATSx, and DLinear.

For load forecasting, HNL outperforms all benchmarks trained individually except for 60-minute level. This is probably because hourly day-ahead forecasts only require 24 prediction points. The number of output dimensions is relatively small and contains limited frequency information. SOTA models can perform well in such low-dimension forecasting settings.

However, at the highest resolution, only the NBEATSx with OPT coordination strategy obtained the similar performance with HNL.

For wind power forecasting, HNL leads all benchmarks with all post coordination strategies, showing the significant improvements.

TABLE V
RMSE COMPARISON ON LOAD DATASET ($\times 10^{-2}$ KW)

| Method | 5min | 15min | 60min |
|---|---|---|---|
| GBRT | 3.443 ± 0.001 | 3.432 ± 0.003 | 3.656 ± 0.009 |
| GBRT-BU | 3.443 ± 0.001 | 3.374 ± 0.001 | 3.278 ± 0.001 |
| GBRT-OPT | 3.453 ± 0.001 | 3.384 ± 0.001 | 3.292 ± 0.001 |
| TFT | 3.924 ± 0.150 | 3.998 ± 0.190 | 3.807 ± 0.096 |
| TFT-BU | 3.924 ± 0.150 | 3.883 ± 0.152 | 3.789 ± 0.152 |
| TFT-OPT | 3.705 ± 0.096 | 3.662 ± 0.097 | 3.569 ± 0.097 |
| NBEATSx | 3.497 ± 0.058 | 3.467 ± 0.086 | 3.335 ± 0.093 |
| NBEATSx-BU | 3.497 ± 0.058 | 3.428 ± 0.062 | 3.327 ± 0.065 |
| NBEATSx-OPT | 3.426 ± 0.050 | 3.355 ± 0.054 | 3.255 ± 0.056 |
| DLinear | 3.554 ± 0.042 | 3.421 ± 0.032 | 3.307 ± 0.025 |
| DLinear-BU | 3.554 ± 0.042 | 3.441 ± 0.043 | 3.328 ± 0.046 |
| DLinear-OPT | 3.516 ± 0.035 | 3.402 ± 0.034 | 3.297 ± 0.035 |
| Persistence | 3.697 ± 0.000 | 3.615 ± 0.000 | 3.479 ± 0.000 |
| HNL | **3.428 ± 0.082** | **3.359 ± 0.079** | 3.360 ± 0.068 |

TABLE VI
RMSE COMPARISON ON WIND POWER DATASET (KW)

| Method | 5min | 15min | 60min |
|---|---|---|---|
| GBRT | 4.795 ± 0.010 | 4.820 ± 0.031 | 4.774 ± 0.060 |
| GBRT-BU | 4.795 ± 0.010 | 4.767 ± 0.010 | 4.700 ± 0.010 |
| GBRT-OPT | 4.799 ± 0.011 | 4.771 ± 0.011 | 4.706 ± 0.012 |
| TFT | 5.730 ± 0.261 | 5.596 ± 0.215 | 4.921 ± 0.339 |
| TFT-BU | 5.730 ± 0.261 | 5.723 ± 0.262 | 5.665 ± 0.265 |
| TFT-OPT | 5.506 ± 0.187 | 5.498 ± 0.187 | 5.441 ± 0.189 |
| NBEATSx | 5.244 ± 0.165 | 5.189 ± 0.221 | 5.039 ± 0.215 |
| NBEATSx-BU | 5.244 ± 0.165 | 5.179 ± 0.167 | 5.101 ± 0.171 |
| NBEATSx-OPT | 5.115 ± 0.159 | 5.048 ± 0.159 | 4.974 ± 0.161 |
| DLinear | 5.179 ± 0.073 | 5.120 ± 0.069 | 5.040 ± 0.049 |
| DLinear-BU | 5.179 ± 0.073 | 5.151 ± 0.074 | 5.082 ± 0.075 |
| DLinear-OPT | 5.150 ± 0.069 | 5.122 ± 0.069 | 5.056 ± 0.070 |
| Persistence | 6.596 ± 0.000 | 6.604 ± 0.000 | 6.640 ± 0.000 |
| HNL | **4.642 ± 0.252**$^*$ | **4.617 ± 0.248**$^*$ | **4.538 ± 0.251**$^*$ |

## REFERENCES


[1] S. Haben, S. Arora, G. Giasemidis, M. Voss, and D. V. Greetham, "Review of low voltage load forecasting: Methods, applications, and recommendations," *Applied Energy*, vol. 304, p. 117798, 2021.

[2] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Applied Energy*, vol. 304, p. 117766, 2021.

[3] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.

[4] B. F. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam, "Analysis of the value for unit commitment of improved load forecasts," *IEEE Transactions on Power Systems*, vol. 14, no. 4, pp. 1342–1348, 1999.

[5] R. Tawn and J. Browell, "A review of very short-term wind and solar power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 153, p. 111758, 2022.

[6] D. V. Pombo, T. Göçmen, K. Das, and P. Sørensen, "Multi-horizon data-driven wind power forecast: From nowcast to 2 days-ahead," in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE, 2021, pp. 1–6.

[7] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting—a review," *Renewable and sustainable energy reviews*, vol. 16, no. 2, pp. 1223–1240, 2012.

[8] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019.

[9] F. Petropoulos, D. Apiletti, V. Assimakopoulos, M. Z. Babai, D. K. Barrow, S. B. Taieb, C. Bergmeir, R. J. Bessa, J. Bijak, J. E. Boylan *et al.*, "Forecasting: theory and practice," *International Journal of Forecasting*, vol. 38, no. 3, pp. 705–871, 2022.

[10] G. Kariniotakis, *Renewable energy forecasting: from models to applications*. Woodhead Publishing, 2017.

[11] H. Li, Z. Wang, T. Hong, A. Parker, and M. Neukomm, "Characterizing patterns and variability of building electric load profiles in time and frequency domains," *Applied Energy*, vol. 291, p. 116721, 2021.

[12] Y. Amara-Ouali, M. Fasiolo, Y. Goude, and H. Yan, "Daily peak electrical load forecasting with a multi-resolution approach," *International Journal of Forecasting*, vol. 39, no. 3, pp. 1272–1286, 2023.

[13] H. Liu, Z. Duan, and C. Chen, "Wind speed big data forecasting using time-variant multi-resolution ensemble model with clustering auto-encoder," *Applied Energy*, vol. 280, p. 115975, 2020.

[14] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and F. Petropoulos, "Forecasting with temporal hierarchies," *European Journal of Operational Research*, vol. 262, no. 1, pp. 60–74, 2017.

[15] D. Yang, H. Quan, V. R. Disfani, and C. D. Rodríguez-Gallegos, "Reconciling solar forecasts: Temporal hierarchy," *Solar Energy*, vol. 158, pp. 332–346, 2017.

[16] P. Nystrup, E. Lindström, P. Pinson, and H. Madsen, "Temporal hierarchies with autocorrelation for load forecasting," *European Journal of Operational Research*, vol. 280, no. 3, pp. 876–888, 2020.

[17] M. Nejati, N. Amjady, and H. Zareipour, "A new multi-resolution closed-loop wind power forecasting method," *IEEE Transactions on Sustainable Energy*, 2023.

[18] H. Liu, R. Yang, and Z. Duan, "Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction," *Energy Conversion and Management*, vol. 217, p. 112995, 2020.

[19] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," *Advances in neural information processing systems*, vol. 31, 2018.

[20] R. Hasani, M. Lechner, A. Amini, L. Liebenwein, A. Ray, M. Tschaikowski, G. Teschl, and D. Rus, "Closed-form continuous-time neural networks," *Nature Machine Intelligence*, vol. 4, no. 11, pp. 992–1003, 2022.

[21] P. Kidger, J. Morrill, J. Foster, and T. Lyons, "Neural controlled differential equations for irregular time series," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6696–6707, 2020.

[22] M. Anvari, E. Proedrou, B. Schäfer, C. Beck, H. Kantz, and M. Timme, "Data-driven load profiles and the dynamics of residential electricity consumption," *Nature communications*, vol. 13, no. 1, p. 4593, 2022.

[23] C. Zhao, C. Wan, and Y. Song, "Cost-oriented prediction intervals: On bridging the gap between forecasting and decision," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 3048–3062, 2021.

[24] J. Zhang, Y. Wang, and G. Hug, "Cost-oriented load forecasting," *Electric Power Systems Research*, vol. 205, p. 107723, 2022.

[25] G. Li and H.-D. Chiang, "Toward cost-oriented forecasting of wind power generation," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2508–2517, 2016.

[26] K. S. Crump, "Numerical inversion of laplace transforms using a fourier series approximation," *Journal of the ACM (JACM)*, vol. 23, no. 1, pp. 89–96, 1976.

[27] K. L. Kuhlman, "Review of inverse laplace transform algorithms for laplace-space numerical approaches," *Numerical Algorithms*, vol. 63, pp. 339–355, 2013.

[28] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[29] C. J. Meinrenken, N. Rauschkolb, S. Abrol, T. Chakrabarty, V. C. Decalf, C. Hidey, K. McKeown, A. Mehmani, V. Modi, and P. J. Culligan, "Mfred, 10 second interval real and reactive power for groups of 390 us apartments of varying size and vintage," *Scientific Data*, vol. 7, no. 1, p. 375, 2020.

[30] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015.

[31] "MARS - the ECMWF meteorological archive," 2018, pages: 1 hour Publication Title: Software and computing services. [Online]. Available: https://www.ecmwf.int/node/18124

[32] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[33] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.
[34] K. G. Olivares, C. Challu, G. Marcjasz, R. Weron, and A. Dubrawski, "Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx," *International Journal of Forecasting*, vol. 39, no. 2, pp. 884–900, 2023.
[35] S. I. Holt, Z. Qian, and M. van der Schaar, "Neural laplace: Learning diverse classes of differential equations in the laplace domain," in *International Conference on Machine Learning*. PMLR, 2022, pp. 8811–8832.
[36] E. O. Brigham, *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.
[37] X. Liu, L. Yang, and Z. Zhang, "The attention-assisted ordinary differential equation networks for short-term probabilistic wind power predictions," *Applied Energy*, vol. 324, p. 119794, 2022.
[38] R. Ye, X. Li, Y. Ye, and B. Zhang, "Dynamicnet: A time-variant ode network for multi-step wind speed prediction," *Neural Networks*, vol. 152, pp. 118–139, 2022.
[39] L. Song, Y. Li, and N. Lu, "Profilesr-gan: A gan based super-resolution method for generating high-resolution load profiles," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3278–3289, 2022.
[40] F. Li, D. Lin, and T. Yu, "Improved generative adversarial network-based super resolution reconstruction for low-frequency measurement of smart grid," *IEEE Access*, vol. 8, pp. 85 257–85 270, 2020.
[41] Y. He, F. Luo, and G. Ranzi, "Load reconstruction with arbitrary super resolutions," *IEEE Transactions on Power Systems*, 2023.

**Chenxi Wang** received the B.S. degree in Electrical Engineering and its Automation from South China University of Technology, Guangzhou, China, in 2022. He is now pursuing a Ph.D. in Electrical and Electronic Engineering at the University of Hong Kong. His current research interests include energy forecasting and machine learning.

**Pierre Pinson** received the M.Sc. degree in applied mathematics from the National In- stitute of Applied Sciences (INSA), Toulouse, France, in 2002 and the Ph.D. degree in energetics from Ecole des Mines de Paris, France, Paris, France, in 2006. He is currently the Chair of data-centric design engineering with Dyson School of Design Engineer- ing, Imperial College London, London, U.K. He is also an affiliated Professor of operations research and analytics with the Technical University of Denmark and a chief Scientist with Halfspace (Denmark). His research interests include analytics, forecasting, optimization and game theory, with application to energy systems mostly, but also logistics, weather-driven industries and business analytics. He is an Editor-in-Chief of the International Journal of Forecasting. He is an INFORMS member and a Director of the International Institute of Forecasters.

**Yi Wang** received the B.S. degree from Huazhong University of Science and Technology in June 2014, and the Ph.D. degree from Tsinghua University in January 2019. He was a visiting student with the University of Washington from March 2017 to April 2018. He served as a Postdoctoral Researcher in the Power Systems Laboratory, ETH Zurich from February 2019 to August 2021.

He is currently an Assistant Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include data analytics in smart grids, energy forecasting, multi-energy systems, Internet-of-things, cyber-physical-social energy systems.