

# Functional Data Analysis for Peak Shape Forecasting

Shreyashi Shukla<sup>a</sup>, Tao Hong<sup>a</sup>, Rob J Hyndman<sup>b</sup>

<sup>a</sup>*University of North Carolina at Charlotte, Charlotte, 28223, North Carolina, USA*

<sup>b</sup>*Monash University, Clayton, 3800, Victoria, Australia*

---

## Abstract

Peak load forecasting is critical for power system operations and planning. Electric utilities are increasingly adopting peak-shaving strategies as they look for sustainable solutions to improve operational excellence. During the past few decades, hourly load forecasts have been used in practice as the input to inform peak-shaving decisions. The BigDEAL Challenge 2022 first proposed the concept of peak (load) shape forecasting, which showed that a better hourly load forecast may not result in a better peak shape forecast. This paper tackles peak shape forecasting, an under-served aspect of peak load forecasting, which is critical for deploying those peak shaving programs. We adopt functional data analysis in exploring new approaches to peak shape forecasting. Using the data published by ISONE, the experiments show that the peak shape forecasts are more robust when compared with the state-of-the-art hourly forecasting models, reducing the errors by 14–18%. The experiment also demonstrates that no single model excels across all attributes, emphasizing the need to build task-specific models.

**Keywords:** Peak shape forecasting, peak timing, functional data analysis, functional principal component, benchmark, peak shape error

---

## 1. Introduction

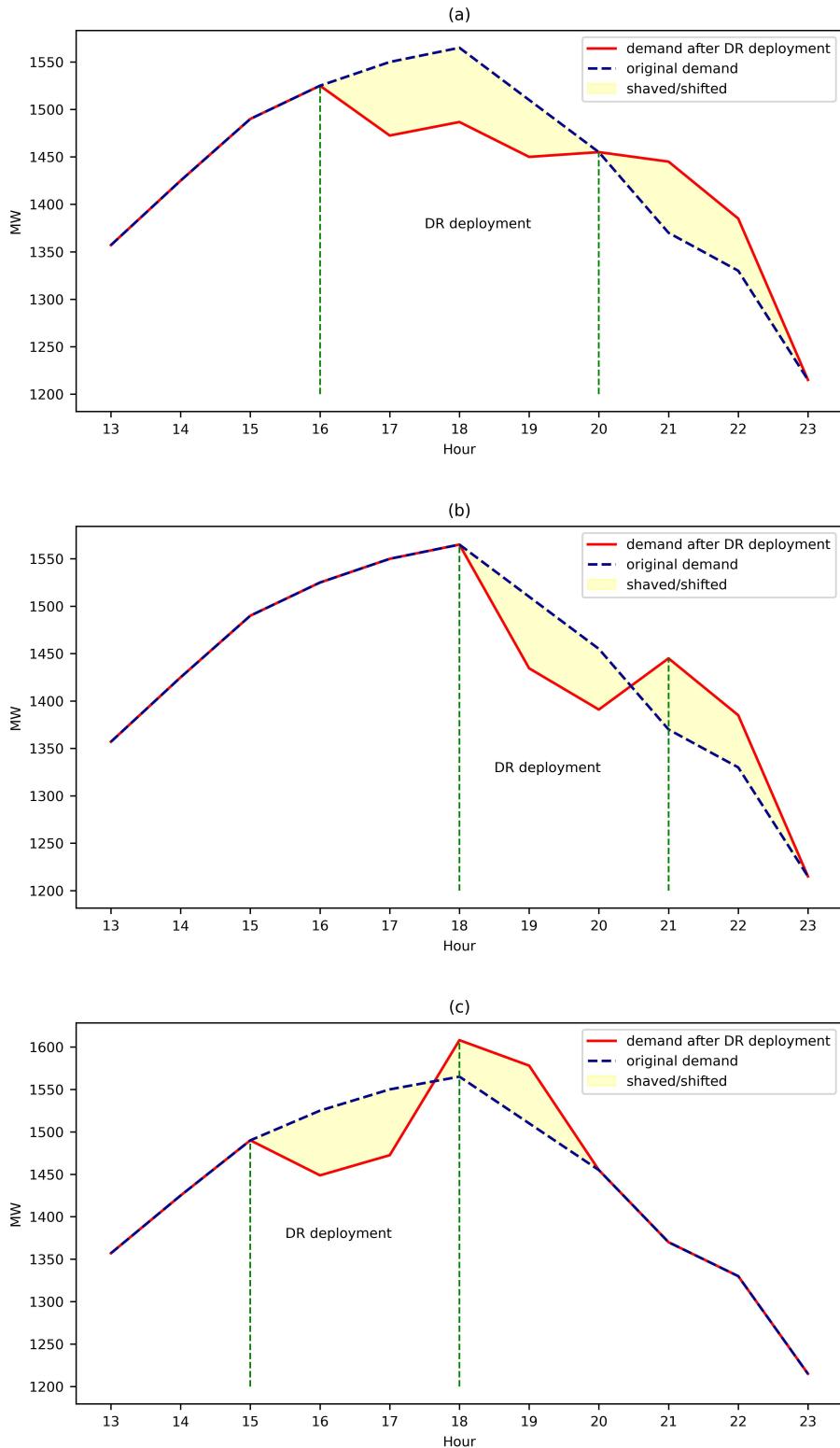
Peak electricity demand forecasting is a critical task in power system planning and operations. Power companies face the challenge of meeting electricity demand while maintaining economical operations and preventing outages. Different types of generators are used to serve the demand at various levels. For example, nuclear plants are typically used to serve base load due to their low unit cost, but they have limited load-following capability. On the other hand, serving peak load requires the generators to have fast ramping capabilities so that they can adapt to significant demand fluctuations quickly. However, many of these generators have a high marginal cost (O'Connell et al., 2014).

For over a century, power companies have been trying to flatten the load profiles to increase the utilization rates of the power systems. Initially, they encouraged and diversified the electricity end uses, which led to high growth of the load. In the recent few decades, the industry started to shave the peaks in order to mitigate the impact of peak demand on energy prices and infrastructure stress. For peak

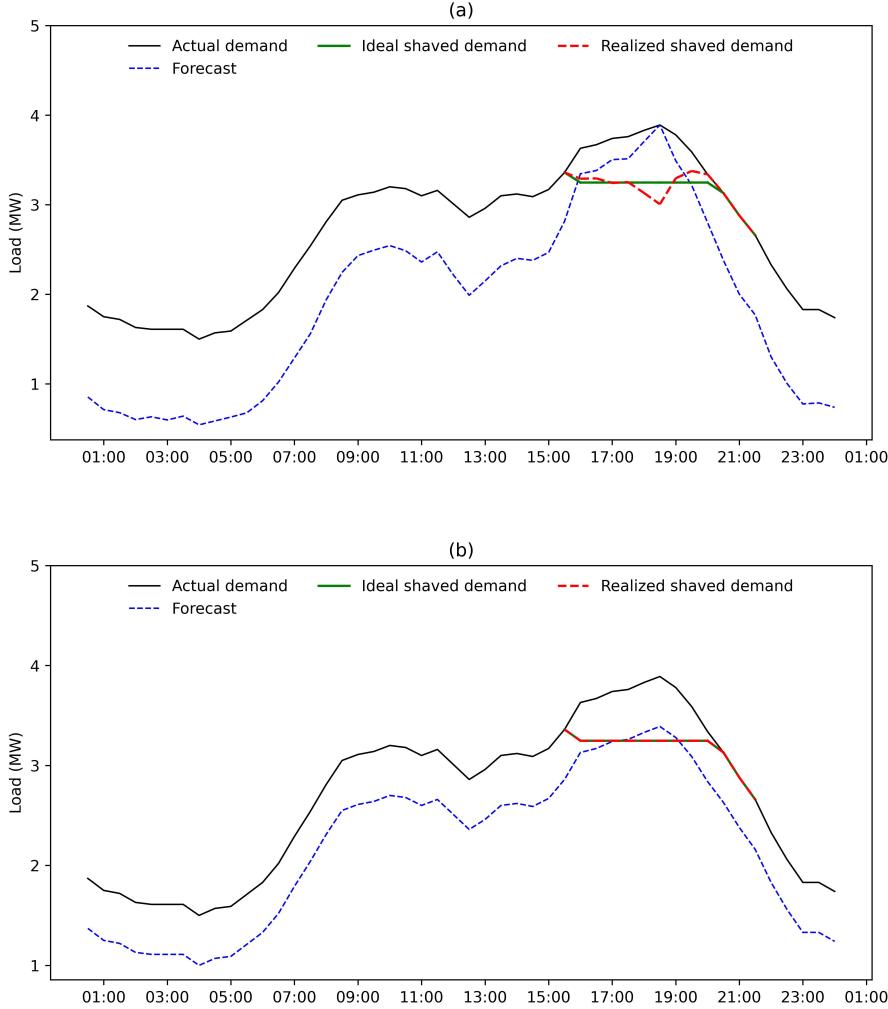
shaving, several solutions have been adopted in today's industry with varying popularity, such as demand-side management (DSM) (Gellings, 1985), energy storage system (ESS) integration (Twitchell, 2019), and vehicle-to-grid (V2G) integration (Kempton and Tomic, 2005). DSM reduces peak demand by deploying various demand response (DR) programs, while ESS and V2G integration provide auxiliary sources to balance the grid. The effectiveness of these peak load management strategies, such as demand response and energy storage systems, relies on accurate forecasts of the peak load day (Saxena et al., 2019) and identification of the optimal time of the day to schedule the event (Amara-Ouali et al., 2023). Moreover, forecasting the shape of demand during the peak period is critical, which has been mostly overlooked. Recently, the BigDEAL Challenge 2022 (Shukla and Hong, 2024), was organized with the theme of peak load forecasting. The competition introduced the concept of peak shape forecasting in one of its three tracks.

DR programs are typically executed by load curtailment and load shifting. The DR events are called on the days when the demand is expected to be unusually high. To maximize the benefit of these programs, utilities need to accurately forecast the day of peak demand, and also the period during the day on which a program is being executed. Ideally, we can achieve optimal peak shaving if an event is scheduled during the right period, as shown in Figure 1(a). Missing the peak period and calling events during the wrong window would jeopardize the peak shaving opportunity as shown in Figure 1(b), where the DR is deployed after the actual peak hour. If the DR event is called much earlier than the actual peak period, it would shift the demand to the true peak hours, raising the original peak demand even more (Figure 1(c)). Knowing just the magnitude and the peak hour is not enough. It is also important to know the shape of ramps leading to and trailing from the peak hour so that an appropriate window to execute these programs can be decided. This is where forecasting the peak shape becomes useful. It takes into account not only the timing of peak demand but also the relative load values around the peak hour.

Peak shape forecasts also play a critical role in scheduling the discharge profile of an ESS during peak periods. The ESS discharge profile specifies the amount of energy to be released during each time interval of the discharge period. An optimal peak shaving is achieved when the discharge profile precisely tracks the actual load profile. In Figure 2, we demonstrate how an accurate peak shape forecast helps. The evening peak demand is to be reduced between 16:00 and 21:00 hours by discharging an ESS, which has a limited capacity. The black line presents the actual load, while the blue dash line is the forecast. The green line is the ideal shaved load profile that can be achieved. In scenario (a), the realized shaved load, shown by the red line, is obtained by using the forecast to optimize the distribution of the discharge of ESS. Since the peak shape forecast is not aligned with the actual shape of the load, the full potential of peak shaving



**Figure 1:** (a) DR deployment in the right window; (b) DR deployment with peak period forecast between 18 h -21 h (later than actual); (c) DR deployment with peak period forecast between 15 h -18 h (earlier than actual).



**Figure 2:** Significance of forecasting load shape for discharge scheduling.

is not achieved. In scenario (b), the peak shape forecast perfectly matches the actual peak shape, although the forecast is biased. Consequently, the peak shaving is effective.

Traditionally, forecasting the shape of the peak is achieved by generating the hourly forecasts for the entire day's profile. However, no formal study has been conducted to understand whether a model that improves hourly forecast accuracy also enhances the peak period shape forecast. In other words, does forecasting the peak shape require a dedicated model? In this paper, we investigate this question. We first evaluate peak shape forecasts from a well-established benchmark model, popularly known as the Vanilla model (Hong et al., 2014). We then analyze the peak shape forecasts from a state-of-art model which has shown a superiority over the Vanilla model by 12% to 15% on hourly forecasts (Wang et al., 2016). We also adopt an innovative approach to this problem by formulating it using the perspective of functional data analysis (FDA), which is a rapidly evolving domain in statistics (Ramsay and Silverman, 2005).

In the context of FDA, the data are analyzed as continuous functions rather than discrete points, with a primary assumption that the measurements originate from an underlying function. Instead of focusing on individual data points, the emphasis is on the function or curve that underlies the data. This shift in perspective allows for a richer analysis of the data structure and dynamics. The functional approach has previously been applied to load forecasting. Shang (2013) applied functional principal component analysis to reduce the dimensions of load curves and obtained orthonormal functions. Then, the principal component scores were predicted using univariate time series models to forecast short-term load. Goia et al. (2010) classified daily load curves using a functional clustering process for short-term peak electricity demand forecasting. Feng and Ryan (2016) used epi-splines to approximate the effect of temperature and dew point on load to forecast the day-ahead hourly load. Mestre et al. (2020) explored functional seasonal autoregressive models with exogenous covariates to forecast the load curves in the short term. Chaouch (2014) presented a day-ahead household-level curve using nonparametric functional time series forecasting.

Peak shape forecasting is consistent with the idea of generating functions from discrete data points, making FDA a promising approach to the problem. The smoothing technique involved in FDA helps in reducing noise and capturing the essential features of the data. Hence, in this paper, we investigate using this technique for peak shape forecasting. We evaluate the performance of the proposed model against benchmark and state-of-the-art hourly load forecasting models. By using the load data published by ISO New England, we compare day-ahead peak shape forecasts in ex-post and ex-ante forecasting set-ups.

This paper is organized as follows: Section 2 gives some background on the two benchmark models used for evaluation, and describes the data used for the experiment; Section 3 introduces the error measures used for the evaluation of different attributes of load forecasting; Section 4 introduces our approach of forecasting daily load curves using functional data analysis; Section 5 present results from the FDA approach and compared them with the two benchmark models; Section 6 concludes the paper with some key takeaways.

## 2. Background

### 2.1. Benchmark models

#### 2.1.1. Vanilla model

Typically, peak shape forecasts are derived from hourly load forecasting models. In this study, we will use Tao's Vanilla Benchmark model (Hong et al., 2015; Wang et al., 2016), the Vanilla model hereafter, as our benchmark model. It uses multiple linear regression (MLR) to model the relationship between load and predictor variables, and has been used as a benchmark model in several recent load forecasting

competitions (Hong et al., 2014, 2016, 2021). Predictions from the model can be obtained using

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 H_t + \beta_3 W_t + \beta_4 M_t + \beta_5 H_t W_t + f(T_t, H_t, M_t), \quad (1)$$

where  $t$  denotes the time index ( $t = 1, 2, 3, \dots$ );  $\hat{y}_t$  is the predicted load at time  $t$ ;

$$f(T_t, H_t, M_t) = \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t H_t + \beta_{10} T_t^2 H_t + \beta_{11} T_t^3 H_t + \beta_{12} T_t M_t + \beta_{13} T_t^2 M_t + \beta_{14} T_t^3 M_t$$

describes the effect of temperature  $T_t$  on load; and  $H_t$ ,  $W_t$ ,  $M_t$  are categorical variables indicating the coincident hour of the day, day of the week, and month of the year, respectively. The multiplication of the predictor variables allows for interaction effects.

### 2.1.2. Recency effect model

The framework of the Vanilla model can be extended to include the impact of the temperatures of the preceding hours on the load, providing a “recency” effect (Wang et al., 2016):

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 H_t + \beta_3 W_t + \beta_4 M_t + \beta_5 H_t W_t + \sum_{h=0}^H f_h(T_{t-h}, H_{t-h}, M_{t-h}) + \sum_{d=1}^D g_d(\tilde{T}_{t,d}, M_t), \quad (2)$$

where  $T_{t-h}$  is the lagged hourly temperature at the preceding  $h^{\text{th}}$  hour ( $h = 0, 1, 2, \dots$ ), and  $\tilde{T}_{t,d}$  is the 24-hour moving average temperature of the preceding  $d^{\text{th}}$  day:

$$\tilde{T}_{t,d} = \frac{1}{24} \sum_{h=24d-23}^{24d} T_{t-h}. \quad (3)$$

Each  $f_h$  function is of the same form as  $f$ , each with different coefficients, while each  $g_d$  function is of the same form as  $f$  without the terms involving hour-of-day  $H_t$ , and with different coefficients. The framework detailed in Wang et al. (2016) was used to thoroughly explore the number of recency effect terms ( $H$  and  $D$ ) to minimize the forecast error.

The recency effect model has been shown to improve forecasts by 10–12% on different datasets (Wang et al., 2016), and has been used as the second benchmark to set a higher bar in forecasting competitions (Hong et al., 2021; Shukla and Hong, 2024). In short-term forecasting practice, adding autoregressive terms (i.e., lagged load values) is valuable to the model, as they capture the short-term trends. In practice, as the forecast for the next day is generated in the morning of the current day, the lagged 24 hours of load data are not available for all hours of the next day. Hence, the most recent load information that can be used are the load values 48 hours before the forecast hours. By adding the lagged dependent variable ( $y_{t-48}$ ) to the Recency effect model, the model can now be represented by (4). We will refer to this model as Recency-L48 from now on. This benchmark is a state-of-the-art model used by utilities. It would be interesting to see its performance in forecasting peak shapes.

$$\hat{y}_t = \beta_0 + \beta_1 t + \beta_2 H_t + \beta_3 W_t + \beta_4 M_t + \beta_5 H_t W_t + \beta_6 y_{t-48} + \sum_{h=0}^H f_h(T_{t-h}, H_{t-h}, M_{t-h}) + \sum_{d=1}^D g_d(\tilde{T}_{t,d}, M_t), \quad (4)$$

## 2.2. Data description

The experiment utilizes publicly available data published by ISO New England (ISONE), which covers the six New England states: Connecticut (CT), Massachusetts (MA), Maine (ME), New Hampshire (NH), Rhode Island (RI), and Vermont (VT). Massachusetts is divided into three load zones — Northeastern MA and Boston (NEMASSB), Southeastern MA (SEMASS), and Western Central MA (WCMASS) — while the remaining five states each constitute their own load zone. The combined data from all zones is referred to as ISONE. The summary statistics of the nine zones are shown in Table 1. The dataset includes hourly load data from the nine load zones and corresponding hourly weather history from weather stations over six years (2014–2019). However, historical temperature forecasts are not included. For this paper, we use data from three years (2017–2019) for out-of-sample testing, with the two preceding years serving as the training period.

Historical weather forecast history is obtained from data archives published by the European Centre for Medium-Range Weather Forecasts (Yang et al., 2022). This dataset contains four years of historical weather forecast history (2017–2020), and 14 weather forecast variables at a spatial resolution of  $0.5^\circ \times 0.5^\circ$ , covering most of North America. We leverage this data to get the day-ahead temperature forecast for all the zones by mapping the coordinates of the weather station corresponding to each load zone, to the closest grid point. We use the hourly temperature forecasts for the three test years (2017–2019).

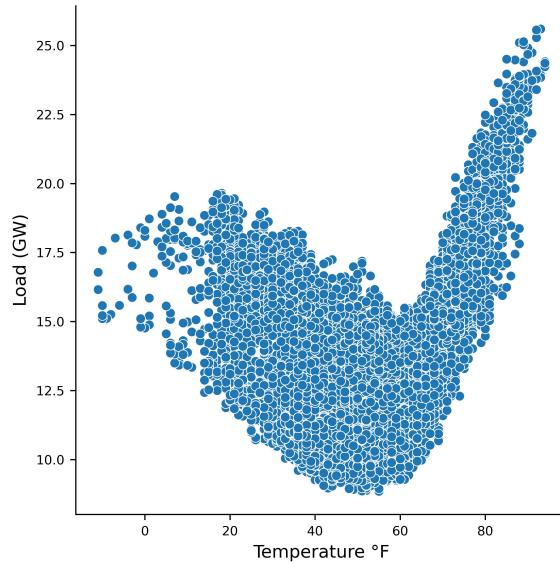
### 2.2.1. Exploratory data analysis

Figure 3 shows the scatter plots of load and historical temperature in 2016. The graph suggests a strong nonlinear correlation between load and temperature. Figure 4 shows the average load curve across

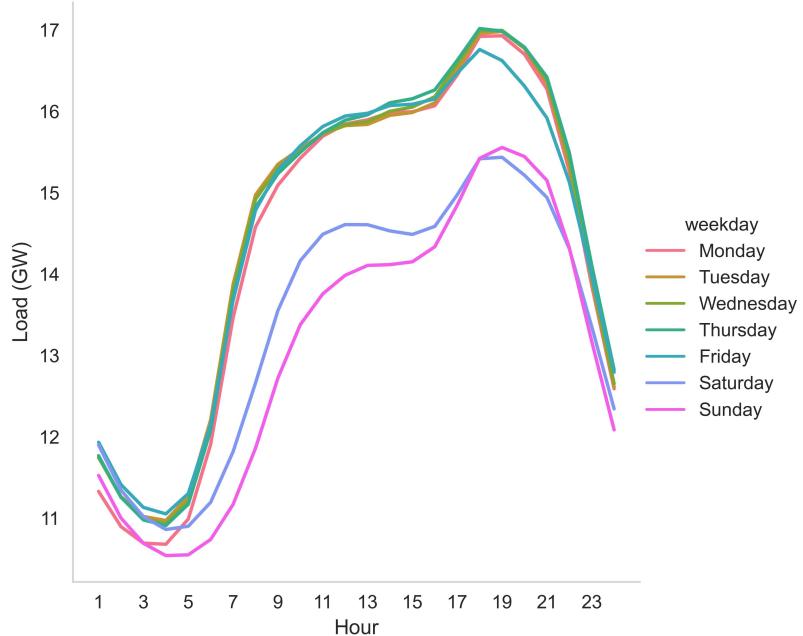
**Table 1:** Summary statistics of load (MW) under ISONE.

Load Zone	Mean	Std	Min	Max
CT	3410	763	1372	7219
ISONE	14212	2829	7892	27707
ME	1303	206	712	2135
NEMASSB	2830	564	1666	5658
NH	1317	262	521	2433
RI	916	204	365	1967
SEMASS	1659	385	875	3645
VT	628	111	178	985
WCMASS	1922	373	739	3650

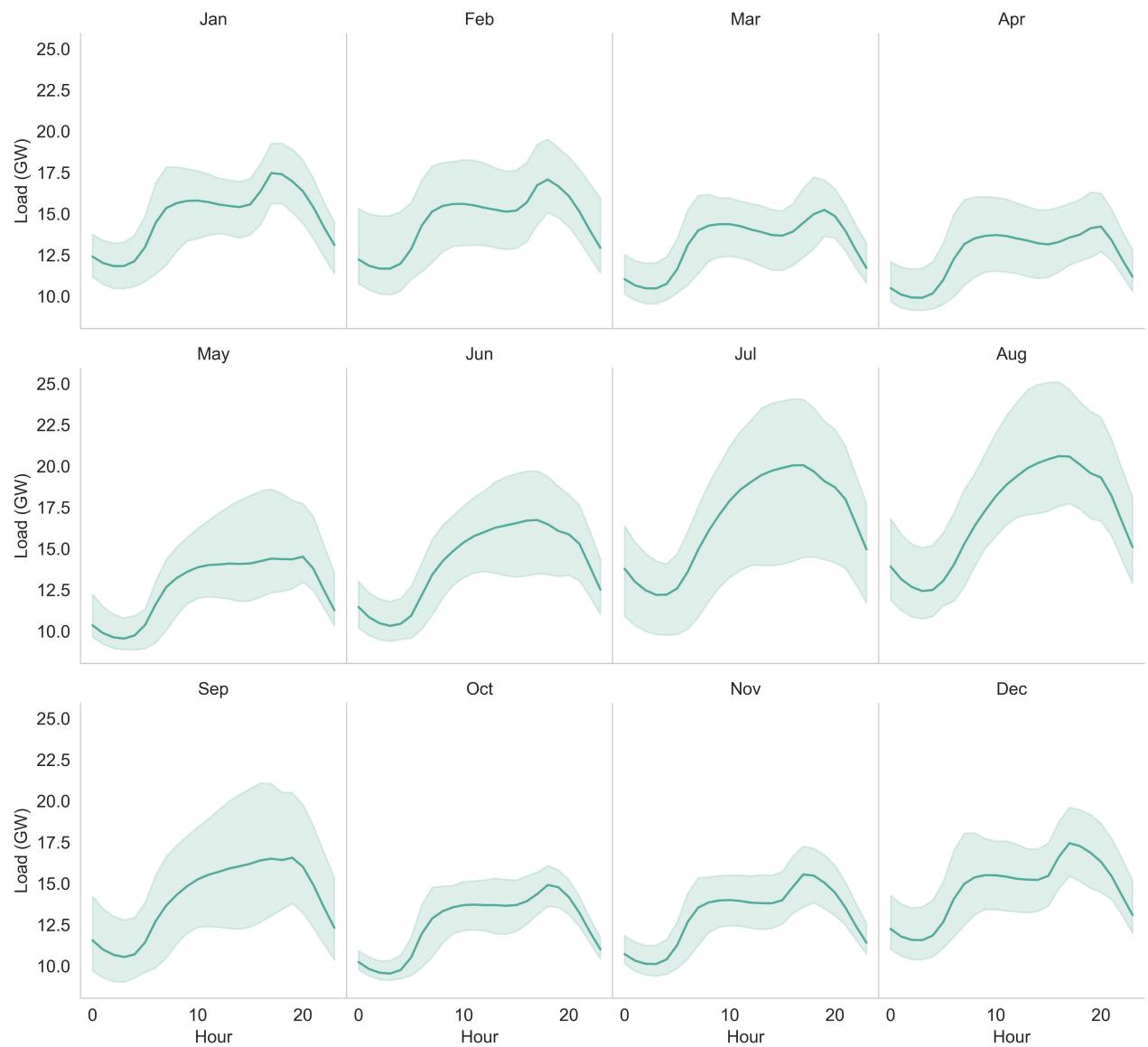
different days of week. The weekend daily load curves are distinctly different from those on weekdays. Figure 5 shows the line plot of the smoothed daily load curve averaged over a month (bold lines) and the range of the observed load values between the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles is depicted by the shaded area. There is a notable difference in the daily load curve between months. The summer months are peaking late evenings, while for the winter months the load curve is bimodal.



**Figure 3:** Scatter plot depicting the non-linear relationship between temperature and load for ISONE.



**Figure 4:** Average load profiles on different days of week.



**Figure 5:** Average load profile in different months of the year 2016 for ISONE, shown by the bold line, while the presents the percentile interval from the sampled observations.

### 3. Error measure

To define peak shape error measures, we first introduce a different notation for the load observations:  $y_{h,i}$  denotes the load at hour  $h$  on the  $i^{\text{th}}$  day.

There are not many studies focused on peak shape forecasting. Singleton and Grindrod (2021) introduced an error metric to evaluate load forecasts that were purposed to be used for optimizing the charging and discharging schedule. The proposed method was implemented in the winning entry of the Western Power Distribution Presumed Open Data 6 MWh battery storage capacity forecasting competition. The metric, named time series shape error, was used as the loss function in the forecasting model. It was only evaluated for evening hours, with 6 pm used as the reference hour. The measure compares the difference between the forecast in hour  $h$  ( $\hat{y}_{h,i}$ ) and at 6 pm ( $\hat{y}_{6PM,i}$ ) on the same day, and the difference between the actual load in hour  $h$  ( $y_{h,i}$ ) and at 6 pm ( $y_{6PM,i}$ ) on the same day, yielding

$$\varepsilon_{h,i} = [(\hat{y}_{h,i} - \hat{y}_{6PM,i}) - (y_{h,i} - y_{6PM,i})]^2. \quad (5)$$

The time series shape error (TSE) for a day is defined as the square root of the mean of  $\varepsilon_{h,i}$  over the discharge hours (i.e., 5pm to 8pm) for that day.

While the metric is implementable, it is biased to the choice of the reference hour. To demonstrate, let's assume three scenarios as shown in Table 2. In case 1, the forecast error is at 5 pm and TSE is 1.5. In case 2, the same error is now at 7 pm, and TSE is the same as in case 1. However, in the case 3, when the same error is in the reference hour load, TSE takes a higher value (2.6).

	Case 1			Case 2			Case 3		
Hour $h$	$y_{h,i}$	$\hat{y}_{h,i}$	$\varepsilon_{h,i}$	$y_{h,i}$	$\hat{y}_{h,i}$	$\varepsilon_{h,i}$	$y_{h,i}$	$\hat{y}_{h,i}$	$\varepsilon_{h,i}$
5 pm	12	15	9	12	12	0	12	12	9
6 pm	14	14	0	14	14	0	14	17	0
7 pm	13	13	0	13	16	9	13	13	9
8 pm	12	12	0	12	12	0	12	12	9
TSE <sub>i</sub>	$\sqrt{\frac{1}{4} \sum_h \varepsilon_{h,i}}$		1.5	1.5			2.6		

**Table 2:** Assessment of time series shape error.

The BigDEAL Challenge 2022 (Shukla and Hong, 2024) introduced peak shape forecasting, where a novel error measure, called mean peak shape error (MPSE), was used to evaluate the forecasts. Let  $P_i$  denote the daily peak load on the  $i^{\text{th}}$  day (the maximum of  $y_{1,i}, \dots, y_{24,i}$ ). Then load shapes are standardized by dividing the load values  $y_{h,i}$  by the daily peak load  $P_n$ , to give  $S_{h,i} = y_{h,i}/P_i$ . Shape forecasts are similarly standardized to obtain  $\hat{S}_{h,i} = \hat{y}_{h,i}/\hat{P}_i$ . The resulting mean peak shape error (MPSE)

can be computed as the average of the absolute error of the shape forecast for the on-peak period ( $Q$ ), over the available days  $i = 1, \dots, n$ :

$$\text{MPSE} = \frac{1}{n} \sum_{i=1}^n \sum_{h \in Q} |S_{h,i} - \hat{S}_{h,i}|. \quad (6)$$

We will use the MPSE to evaluate peak shape forecast in this study.

The next question is how we should define the on-peak period  $Q$  for the problem. The on-peak periods are selected based on the seasonal load profile of the system load. A fixed broad window of hours is chosen for the on-peak period to cover when peaks are most likely to occur. Utilities typically designate the evening hours during summers and morning hours during winters. It is crucial for the implementation of dispatchable peak-shaving techniques and tariff designing to cover the peak hour of day. However, this way of defining  $Q$  may not always contain the peak hour. Hence, instead of using a generic broad time window, the BigDEAL Challenge 2022 introduced a new definition of an on-peak period:  $\pm 2$  hours adjoining the peak hour for the same day, which can be written as

$$\dot{Q} = \{H_i - 2, H_i - 1, H_i, H_i + 1, H_i + 2\},$$

where  $H_i = \arg \max_h y_{h,i}$  denotes the peak hour of the  $i^{\text{th}}$  day.

We further extend this definition by incorporating a threshold for the load values. For example, the threshold could be 90% of the daily peak load, resulting in an alternative  $Q$  given by

$$\ddot{Q} = \dot{Q} \cap \{h : y_{h,i} \geq P_i \times 0.90\}.$$

It is notable that under both definitions, the peak demand period must be contiguous and including the peak hour. It is also noteworthy that we will evaluate peak shape errors only over the actual on-peak period. By not including the evaluation of the on-peak period forecast, we avoid the situation of redundantly penalizing the same errors twice, sometimes called a double penalty (Haben et al., 2014). In addition, we will use the mean absolute percentage error (MAPE) to evaluate hourly load forecasts, as defined in (7). Despite its well-known flaws (Hyndman and Koehler, 2006), it is one of the most widely used error measures in the load forecasting field due to its scale-independency and interpretability. We will also evaluate the daily peak forecast, for which we will use the Peak MAPE defined in (8).

$$\text{MAPE} = \frac{1}{24n} \sum_{i=1}^n \sum_{h=1}^{24} \left| \frac{y_{h,n} - \hat{y}_{h,n}}{y_{h,n}} \right| \times 100\% \quad (7)$$

$$\text{Peak MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - \hat{P}_i}{P_i} \right| \times 100\%. \quad (8)$$

Hyndman and Koehler (2006) proposed scaled errors as an alternative to using percentage errors when evaluating forecast accuracy. We will use the root mean square scaled errors (RMSSE) to evaluate

hourly and peak forecasts, as described in (9) and (10). The errors are scaled based on the seasonal naïve hourly forecasts for RMSSE, and the seasonal naïve peak forecasts for PEAK RMSSE, that is, the actual load values from the previous week.

$$\text{RMSSE} = \sqrt{\frac{\sum_{i=1}^n \sum_{h=1}^{24} (y_{h,i} - \hat{y}_{h,i})^2}{\sum_{i=1}^n \sum_{h=1}^{24} (y_{h,i} - y_{h-168,n})^2}} \quad (9)$$

$$\text{PEAK RMSSE} = \sqrt{\frac{\sum_{i=1}^n (P_n - \hat{P}_n)^2}{\sum_{i=1}^n (P_n - P_{n-7})^2}} \quad (10)$$

Note that the scaling denominators are defined using test data, rather than the training data as proposed by Hyndman and Koehler (2006). This is because there is sufficient test data in this evaluation to obtain reliable scaling constants, and scaling by measures computed on the test data allows easier interpretation of the results.

#### 4. Functional data analysis

The proposed method uses the perspective of functional data analysis to forecast the daily load curve. For each day, we assume that there exists some underlying smooth and continuous load function over the 24 hours of day. Let  $\ell_i(x)$  denote the daily load curve for day  $i$  at hour  $x$ , where  $x \in [1, 24]$ . The value of the function at any point is equal to the total load over the preceding hour. We assume that we observe  $\ell_i(x)$  with error and at discrete points of  $x$ . Hence, the observations for day  $i$  are  $\{x_j, y_i(x_j)\}$ , where  $x_j = 1, \dots, 24$ ,  $y_i(x_j) = \ell_i(x_j) + \varepsilon_{i,j}$ , and  $\varepsilon_{i,j}$  is a random error term. We are interested in forecasting the load values for the next day, that is  $y_{i+1}(x)$ , using all available data up to day  $i$ . The framework of FPCA for forecasting was first proposed by (Hyndman and Ullah, 2007), and has been adopted by (Shang, 2013) in the context of load forecasting.

##### 4.1. Smooth load curves

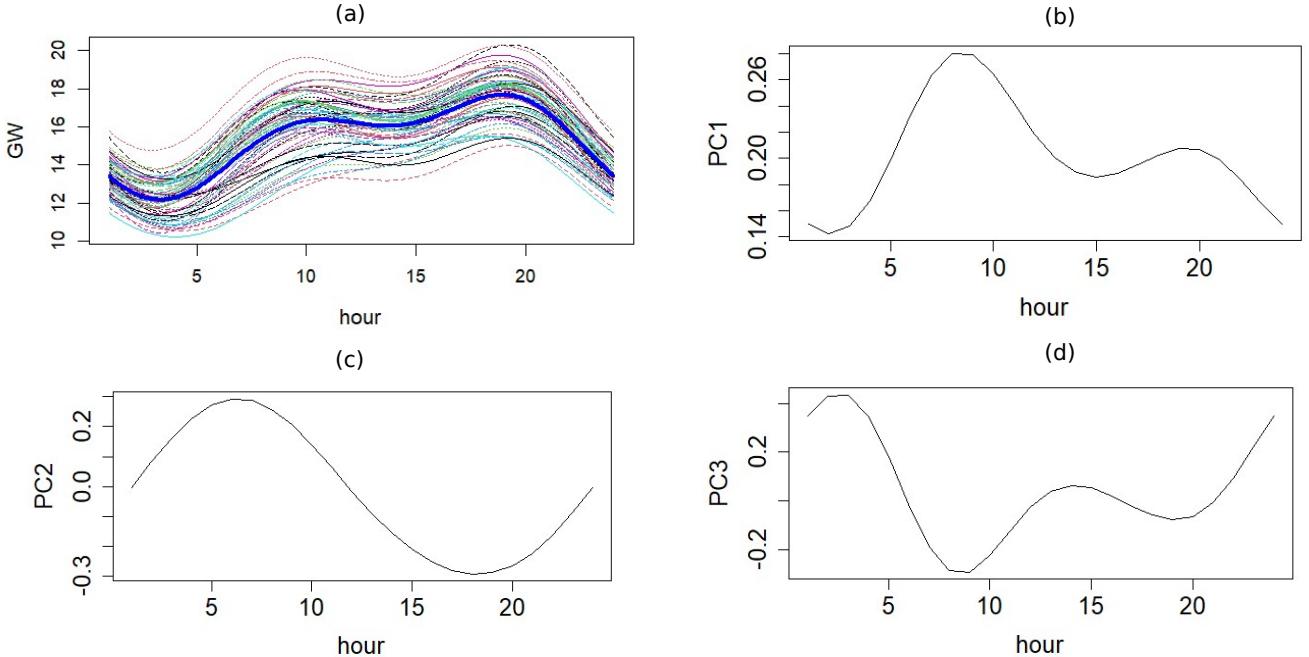
The first step is to derive smooth functional data,  $\ell_1(x), \dots, \ell_n(x)$ . Given that the daily load observations are of hourly resolution in our study, periodic Fourier basis functions defined over 24 hours are used for smoothing. We use all Fourier basis functions with an L2 roughness penalty  $\lambda$  on the coefficients to control the smoothness of the fitted curves. A small  $\lambda$  allows variations in the data to be captured, minimizing loss of information. In our analysis, we have used  $\lambda = 1$  for the smoothing process of load curves.

#### 4.2. Functional principal component analysis

Functional principal component analysis (FPCA) is commonly used in forecasting functional data. In this process, the curves are first centralized by removing the estimated mean curve,  $\mu(x)$ . The centralized load curves are then transformed to extract independent curves that explain the maximum possible variance of the original load curves. The orthogonal basis functions that explain the maximum variance in the direction of eigenfunctions are obtained as functional principal components,  $\phi_1(x), \phi_2(x), \dots$ . In this way, each of the original curves can be represented by linear combinations of the functional principal components. By projecting the original curves onto the principal components, we obtain the coefficients  $\{\beta_{k,i}\}$ , also called principal functional scores. Thus, the load curves can be represented as

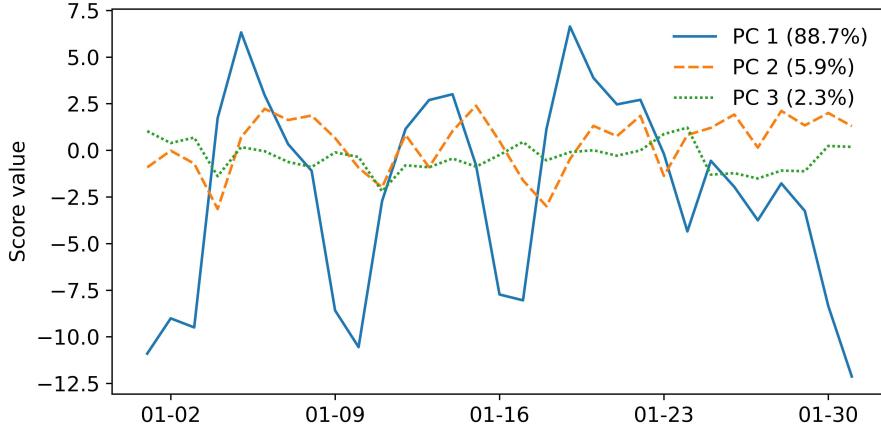
$$\ell_i(x) = \mu(x) + \sum_{k=1}^K \beta_{k,i} \phi_k(x) + e_i(x), \quad (11)$$

where  $K$  is the number of principal components we will use, and  $e_i(x)$  is a functional random error due to  $K$  being less than its maximum value.



**Figure 6:** Estimated mean curve and the first three principal component functions for load curves of January 2015-2016.

To implement FPCA, we have used the `pca.fd()` function from the `fda` package (Ramsay et al., 2005). The analysis is done month-wise, so we will have twelve models, one for each month, with between 56 and 62 curves available for each model. Figure 6(a) shows the estimated mean curve (shown by the bold blue line), and the first three principal component functions (Figure 6(b,c,d)) fitted to the data from

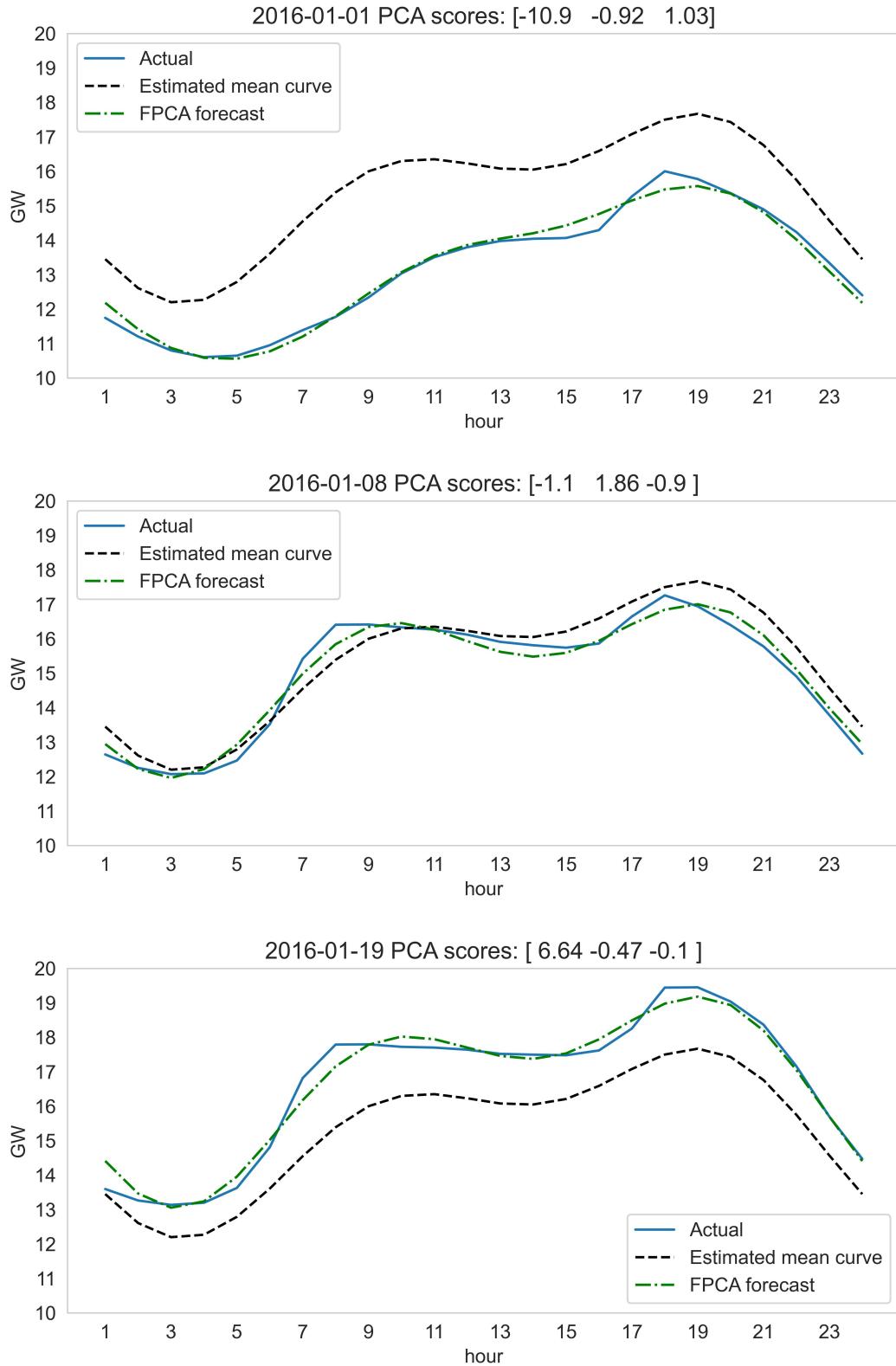


**Figure 7:** The principal component scores of the first three principal components (PC) functions for the load curves of ISONE (Jan 2016). Notably, the first component scores capture the weekly pattern. The percentages indicate the amount of variance explained from each of the first three PCs.

January 2015 and 2016. The mean curve has two peaks, with the evening peak slightly higher than the morning peak. The first principal component mostly captures the morning peak, with a small contribution from the evening peak; while the second principal component captures the difference between the early morning and later afternoon load. Higher order principal components become increasingly difficult to interpret. The first three principal components explain up to 97.5% of variance as shown in Figure 7.

The first component scores have a much greater influence compared to the other two scores. These scores are positive when the estimated daily load curve exceeds the mean curve for all  $x$ , and negative when it falls below for all  $x$ . When the second component score is a large positive number, the difference between the morning and evening peaks is reduced; conversely, a negative second component score increases this difference. Figure 8 illustrates the analysis on three different days in the validation year. The first panel shows a day when the load curve is lower than the mean curve, so the first score is a large negative number ( $-10.9$ ). The second panel shows a day when the load curve is close to the mean curve, but the morning and evening peaks are close. Hence, the first score is a relatively small number, while the second score dominates with a positive value ( $1.86$ ). The third panel shows a day when the load curve is higher than the mean curve; consequently, the first score takes a large positive value ( $6.64$ ).

Once we have determined the principal components from the training data, the next step is to forecast the coefficients  $\{\beta_{k,i}\}$ ,  $k = 1, \dots, K$ , and construct the forecasts for the test data. Since these scores are uncorrelated to each other by construction, we can fit a univariate time series model to each series of coefficients  $\{\beta_{k,1}, \beta_{k,2}, \beta_{k,3} \dots \beta_{k,n}\}$  in order to forecast the future load curves (Shang, 2013).



**Figure 8:** Actual, forecast, and mean load curve for three different days in January 2016 for ISONE. The values of the principal component scores are shown in the title of each panel.

As illustrated in Figure 4, the daily load shape varies across different days of the week, hence, the day of week  $W_i$  could be a useful predictor in each model. Additionally, temperature impacts the load, so incorporating the daily average temperature  $\bar{T}_i$ , daily maximum temperature  $T_i^+$ , and daily minimum temperature  $T_i^-$ , can also be useful for forecasting the scores. Using these four predictors, we fit a multiple linear regression model on the training data:

$$\hat{\beta}_{k,i} = \theta_{k,0} + \theta_{k,1}\bar{T}_i + \theta_{k,2}T_i^+ + \theta_{k,3}T_i^- + \theta_{k,4}W_i. \quad (12)$$

The principal component scores for the test days are predicted using the day of week and temperature features of the test day. We have used  $K = 3$  principal components for FPCA, as for most months in the validation year, three components were sufficient to explain the variability of curves. The daily load curve forecasts are constructed using the estimated mean curve, the predicted scores for the test days, and the principal component functions. The validation results under ISONE, presented by the MPSE values are shown in Table 3. The FPCA model forecast does not beat the forecasts from the Vanilla model for any of the attributes, including peak shape errors. This is because the model lacks hourly temperature information, unlike the Vanilla model. However, it is noteworthy that the peak MAPE and peak RMSSE from FPCA are close to those from the Vanilla model, indicating that the FPCA approach might be advantageous for forecasting temporally higher levels of load.

**Table 3:** Validation results under ISONE for 2016, comparing the Vanilla model with the Functional data model (FPCA).

MAPE		Peak MAPE		RMSSE		Peak RMSSE		MPSE ( $\dot{Q}$ )		MPSE ( $\ddot{Q}$ )	
Vanilla	FPCA	Vanilla	FPCA	Vanilla	FPCA	Vanilla	FPCA	Vanilla	FPCA	Vanilla	FPCA
3.46	5.30	3.50	3.70	0.51	0.76	0.51	0.52	0.09	0.14	0.08	0.12

#### 4.2.1. Function-on-function regression

As seen in the previous section, the peak shape forecast from the FPCA approach is incompetent on its own. Hence, we enhanced the functional analysis by adding temperature curves using a function-on-function regression method. For this approach, we use smoothed daily temperature curves as predictors to forecast peak shapes. The function-on-function regression model has been implemented using penalized function-on-function regression (PFFR) as described in Ivanescu et al. (2015), which has been implemented in the `pffr` function of the R package **refund** (Goldsmith et al., 2024). PFFR is an additive regression framework in the functional form where coefficient functions are estimated using penalized spline regression. This is our base model B1:

$$\ell_i(x) = \alpha_0(x) + \int \alpha_1(x) \dot{T}_i(x) dt + e_i(x). \quad (13)$$

Here,  $\dot{T}_i(x)$  denotes the temperature curve of day  $i$ , centered by subtracting the estimated mean curve over all days in the training data. Similar to FPCA analysis, the models are trained month-wise, so we will have twelve models, one for each month. For example, to forecast the load curves from the month of January of the test year, we used the model trained on load curves using January data from the previous two years.

Some other variants of model B1 were also tried. Model B2 adds scalar covariates such as day of week and daily temperature features to B1:

$$\ell_i(x) = \alpha_0(x) + \int \alpha_1(x) \dot{T}_i(x) dt + \alpha_2 \bar{T}_i + \alpha_3 T_i^+ + \alpha_4 T_i^- + \alpha_5 W_i + e_i(x). \quad (14)$$

Finally, we tried adding the three coefficient forecasts from FPCA analysis  $\{\hat{\beta}_{k,i}\}$ , which implicitly contain information on daily temperature features and weekday features. This variant of the model is called B3:

$$\ell_i(x) = \alpha_0(x) + \int \alpha_1(x) \dot{T}_i(x) dt + \sum_{k=1}^K \alpha_{1+k}(x) \hat{\beta}_{k,i} + e_i(x). \quad (15)$$

Table 4 presents the MPSE under  $\dot{Q}$  and  $\ddot{Q}$  from the three models compared with the benchmark. The base model B1 reduces the MPSE of the benchmark model by a substantial margin. The add-on model B2 does not bring much further improvement, however B3 brings the error down further. This suggests that adding daily temperature and day-of-week features directly to the base model does not help. Instead, the principal components score forecasts, derived from FPCA analysis, help to improve the peak shape forecast. Hence, model B3 is chosen from the validation process.

Once the historical training data is divided into twelve monthly segments, the subsequent steps under model B3 for each monthly segment  $m$  are outlined below:

*Step 1:* Smooth the daily load curves from historic data of the same month using Fourier basis functions to estimate  $\ell_i(x)$  from the observations  $\{x_j, y_i(x_j)\}$ ,  $x_j = 1, 2, \dots, 24$ ,  $i = 1, \dots, n$ .

*Step 2:* Decompose the fitted curves into  $K$  principal component functions using (11).

**Table 4:** Validation results from different models under function-on-function regression.

Model	MPSE $\dot{Q}$	MPSE $\ddot{Q}$
Vanilla benchmark	0.090	0.081
FPCA	0.141	0.115
B1	0.056	0.050
B2	0.059	0.050
B3	<b>0.052</b>	<b>0.047</b>

*Step 3:* Fit multiple linear regression models to each of the coefficients  $\{\beta_{k,i}\}$ , for  $k = 1, \dots, K$ , using daily temperature features such as the maximum, minimum, and average temperature of day, and day of  $W_n$  as in (12).

*Step 4:* Predict the coefficients  $\hat{\beta}_{k,i}$ , for the period covered by the test data, using the models from *Step 3*.

*Step 5:* Smooth the temperature curves from historical data of the same month using Fourier basis functions and estimate the mean curve  $\mu_s(x)$ .

*Step 6:* Use  $\mu_s(x)$  to get the centered temperature predictors  $\tilde{T}_i(x)$  for training and test data.

*Step 7:* Fit function-on-function regression model B3 to the daily load curve as described in (15).

*Step 8:* Finally, construct the daily load curve forecast for the next day.

We will call this model FoF, hereafter.

## 5. Results

In this section, we compare the performance of all models for different attributes including hourly forecast, daily peak forecast, and peak shape forecasts under the three test years. The error measures used for each of these forecasting problems are described in Section 3. Table 5A shows the ex-post forecasting results. The columns present the average hourly MAPE, Peak MAPE, RMSSE, Peak RMSSE, and MPSE across different zones from the Recency-L48 model and FoF model compared side by side with the Vanilla model. The bold font indicates superior results. The Recency-L48 model improves the hourly MAPE as well as RMSSE over the Vanilla model for all zones. RMSSE values from FoF get close to the Vanilla model, while as per MAPE, FoF does not outperform the Vanilla model. For Peak MAPE, the performance of the Recency-L48 model dominates in all zones but one, while the FoF model beats the Vanilla model in most of the zones. Coming to MPSE values, the FoF model is unbeatable across all zones for both on-peak periods. Therefore, while the FoF model does not enhance hourly forecasts, it outperforms the Recency-L48 model in forecasting peak shape forecasting.

Ex-ante forecasting results from the three models, shown in Table 5B, follow a similar trend as ex-post forecasting. The experiment setup for ex-ante mimics a real-time scenario, using day-ahead temperature forecasts for the test data, rather than the observed temperatures. Both the Recency-L48 model and FoF model outperform the Vanilla model in all three attributes. However, for peak shape errors, FoF outdoes even the Recency-L48 model.

Table 6 summarizes the percentage difference in the three overall forecast errors from the Recency-L48 model and FoF model when compared with the Vanilla model. While the FoF model does not improve the

**Table 5:** Test results for the Vanilla model (1), the Recency-L48 model (4), and the Function-on-function regression model (15).

**A. Ex-post forecasting results**

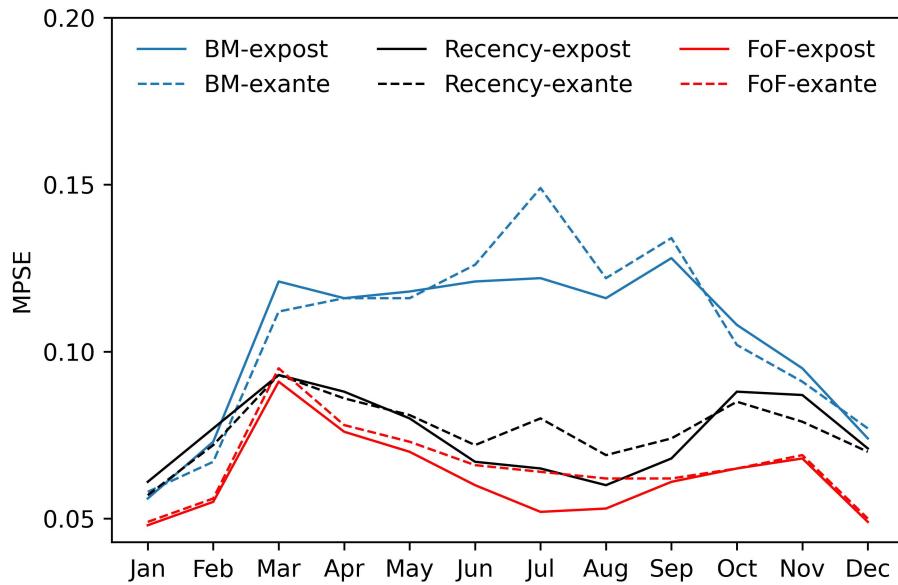
	MAPE			Peak MAPE			RMSSE			RMSSE-PEAK			MPSE( $\bar{Q}$ )		
	Vanilla	Recency	FoF	Vanilla	Recency	FoF	Vanilla	Recency	FoF	Vanilla	Recency	FoF	Vanilla	Recency	FoF
CT	4.76	<b>3.51</b>	5.30	4.78	<b>3.35</b>	4.66	0.53	<b>0.40</b>	0.56	0.52	<b>0.40</b>	0.53	0.115	0.077	<b>0.065</b>
ISONE	3.77	<b>3.00</b>	4.24	3.71	<b>2.91</b>	3.79	0.50	<b>0.41</b>	0.54	0.49	<b>0.42</b>	0.52	0.090	0.063	<b>0.053</b>
ME	3.86	<b>3.25</b>	3.49	3.98	3.32	<b>3.11</b>	0.67	<b>0.57</b>	0.62	0.74	<b>0.62</b>	0.087	0.071	<b>0.059</b>	0.081
NEMASSB	4.38	<b>3.84</b>	4.91	4.69	<b>4.07</b>	4.59	0.54	<b>0.47</b>	0.58	0.55	<b>0.49</b>	0.54	0.101	0.074	<b>0.060</b>
NH	3.89	<b>2.85</b>	3.88	3.78	<b>2.79</b>	3.58	0.55	<b>0.43</b>	0.54	0.53	<b>0.42</b>	0.52	0.097	0.070	<b>0.057</b>
RI	4.15	<b>3.17</b>	4.39	4.50	3.26	4.28	0.54	<b>0.43</b>	0.55	0.56	<b>0.43</b>	0.53	0.105	0.070	<b>0.058</b>
SEMASS	4.89	<b>3.78</b>	5.20	4.26	<b>3.31</b>	4.48	0.54	<b>0.42</b>	0.55	0.51	<b>0.42</b>	0.56	0.122	0.082	<b>0.067</b>
VT	6.74	<b>5.49</b>	6.23	3.67	<b>3.33</b>	3.37	0.73	<b>0.60</b>	0.69	0.70	<b>0.69</b>	0.68	0.119	0.097	<b>0.079</b>
WCMASS	4.92	<b>3.82</b>	4.99	4.21	<b>3.28</b>	4.08	0.59	<b>0.48</b>	0.58	0.55	<b>0.47</b>	0.56	0.100	0.076	<b>0.062</b>
Average	4.60	<b>3.63</b>	4.74	4.18	<b>3.29</b>	3.99	0.58	<b>0.47</b>	0.58	0.57	<b>0.49</b>	0.56	0.104	0.076	<b>0.062</b>
															<b>0.056</b>

**B. Ex-ante forecasting results**

	MAPE			Peak MAPE			RMSSE			RMSSE-PEAK			MPSE( $\bar{Q}$ )		
	Vanilla	Recency	FoF	Vanilla	Recency	FoF	Vanilla	Recency	FoF	Vanilla	Recency	FoF	Vanilla	Recency	FoF
CT	5.21	<b>4.08</b>	5.34	5.19	<b>4.08</b>	4.83	0.58	<b>0.46</b>	0.58	0.59	<b>0.48</b>	0.56	0.114	0.078	<b>0.071</b>
ISONE	4.20	<b>3.37</b>	4.29	4.05	<b>3.35</b>	3.85	0.55	<b>0.45</b>	0.55	0.55	<b>0.48</b>	0.54	0.092	0.063	<b>0.058</b>
ME	3.95	<b>3.29</b>	3.61	4.02	3.30	<b>3.19</b>	0.70	<b>0.59</b>	0.63	0.76	<b>0.64</b>	0.63	0.087	0.067	<b>0.059</b>
NEMASSB	4.72	<b>4.09</b>	4.82	4.67	<b>4.08</b>	4.54	0.57	<b>0.50</b>	0.59	0.57	<b>0.51</b>	0.56	0.098	0.071	<b>0.062</b>
NH	4.27	<b>3.32</b>	4.30	4.40	<b>3.43</b>	3.93	0.60	<b>0.48</b>	0.58	0.60	<b>0.49</b>	0.56	0.102	0.074	<b>0.059</b>
RI	4.73	<b>3.86</b>	4.70	5.23	<b>4.46</b>	4.79	0.61	<b>0.50</b>	0.58	0.64	<b>0.54</b>	0.57	0.110	0.077	<b>0.064</b>
SEMASS	5.47	<b>4.40</b>	5.48	5.45	<b>4.49</b>	5.02	0.61	<b>0.48</b>	0.57	0.64	<b>0.53</b>	0.60	0.111	0.080	<b>0.069</b>
VT	6.96	<b>5.98</b>	6.53	3.96	<b>3.90</b>	3.45	0.75	<b>0.65</b>	0.71	0.76	<b>0.75</b>	0.69	0.119	0.097	<b>0.083</b>
WCMASS	5.39	<b>4.14</b>	4.96	4.93	<b>3.85</b>	4.17	0.64	<b>0.53</b>	0.58	0.63	<b>0.56</b>	0.57	0.119	0.082	<b>0.067</b>
Average	4.99	<b>4.06</b>	4.89	4.66	<b>3.88</b>	4.20	0.62	<b>0.52</b>	0.60	0.64	<b>0.55</b>	0.59	0.106	0.076	<b>0.066</b>
															<b>0.060</b>

**Table 6:** Percentage difference ( $\Delta\%$ ) in errors compared to the Vanilla model from Recency-L48 and FoF.

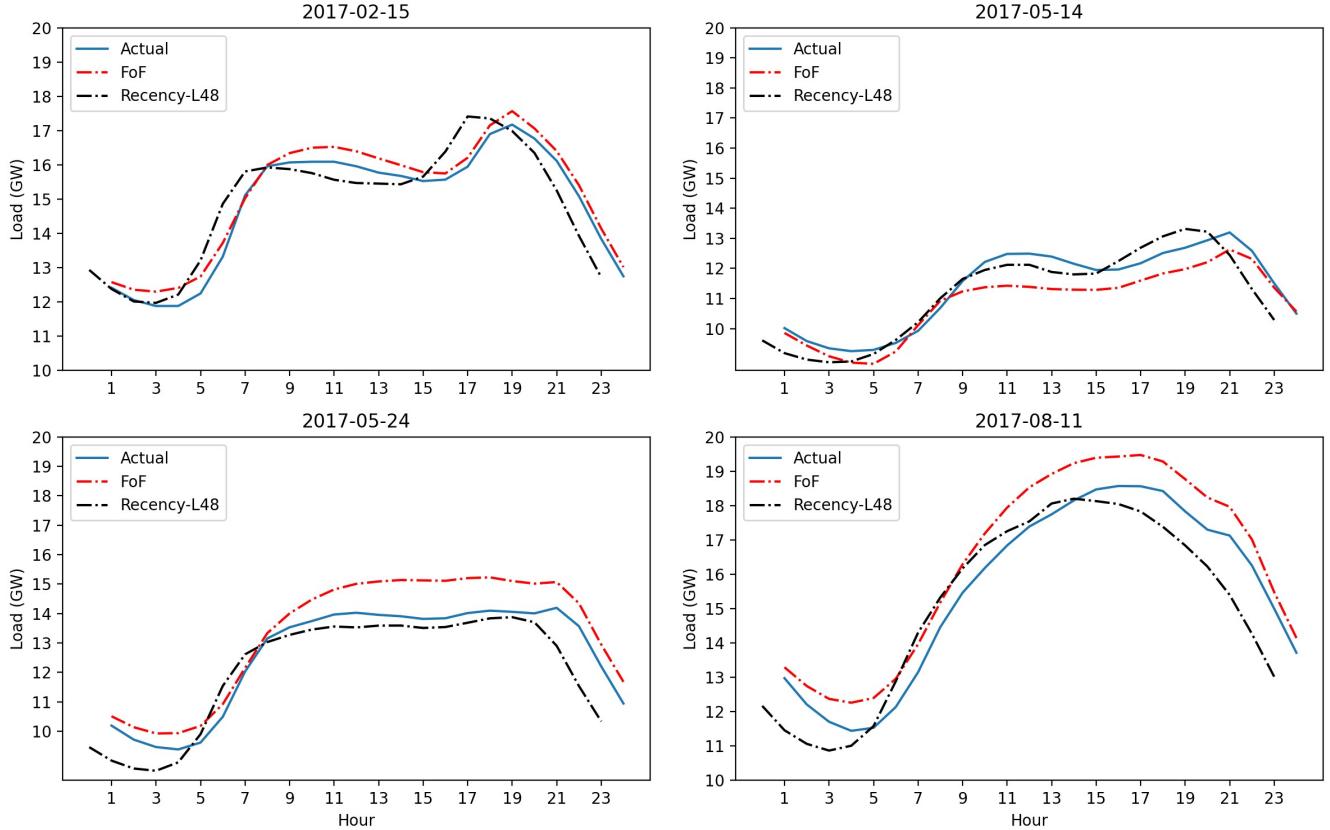
	$\Delta\text{MAPE}$	$\Delta\text{Peak MAPE}$	$\Delta\text{RMSSE}$ ( $\dot{Q}$ )	$\Delta\text{Peak RMSSE}$ ( $\ddot{Q}$ )	$\Delta\text{MPSE}$ ( $\dot{Q}$ )	$\Delta\text{MPSE}$ ( $\ddot{Q}$ )
<i>Ex-post forecast results</i>						
Recency-L48	<b>20.9%</b>	<b>21.2%</b>	<b>18.7%</b>	<b>15.2%</b>	27.3%	28.9%
FoF	-3.1%	4.4%	-0.2%	2.0%	<b>40.2%</b>	<b>40.8%</b>
<i>Ex-post forecast results</i>						
Recency-L48	<b>18.7%</b>	<b>16.6%</b>	<b>17.3%</b>	<b>13.2%</b>	27.8%	28.8%
FoF	1.9%	9.9%	4.3%	7.9%	<b>37.9%</b>	<b>38.8%</b>



**Figure 9:** Line plots showing the mean PSE over different months.

hourly forecasts from the Vanilla model, it enhances the peak shape forecasts by 37.9 to 40.2%, surpassing even peak shape forecasts from the Recency-L48 model by 14 to 18%.

Figure 9 analyses the MPSE across different months for the three models in ex-post and ex-ante setup. The green, black lines and red lines represent the Vanilla model, Recency-L48 model, and FoF model, respectively. The solid lines and dashed lines are used for ex-post and ex-ante forecasting respectively. In general, the peak shape forecasts are close for ex-post and ex-ante for the winter months for any model. However, for summer months, the FoF seems more robust to the temperature forecast errors as the ex-post and ex-ante MPSE values are closer.



**Figure 10:** Daily load curves from Recency-L48 model and FoF model compared with actuals.

Figure 10 shows the daily load curves for selected days from the test year 2017. While the FoF forecasts may exhibit some bias, they closely track the actual load shape during peak periods. This demonstrates that no single model excels across all attributes.

## 6. Conclusion

Peak shape refers to the shape of electricity demand during the peak hour and its surrounding hours. The peak shape forecasts are crucial for the effective implementation of various peak-shaving programs. Traditionally, power companies derive daily peaks from the forecasts of hourly load based on an underlying assumption that the improvement in hourly load forecasting can be translated to the improvement in peak shave forecasting. However, the BigDEAL Challenge 2022 (Shukla and Hong, 2024) revealed some counterexamples against this assumption, which highlighted the importance of developing task-specific load forecasting models.

In this paper, we have tackled an emerging problem, peak shape forecasting, which concerns the shapes of electricity demand during on-peak periods. Instead of focusing on the original hourly load, we have proposed a methodology based on functional data analysis to directly target the peak shape.

The proposed methodology involves applying functional principal component analysis and function-on-function regression. The computational experiment is based on the data published by ISONE, with the Vanilla model (Hong et al., 2015) and the recency effect model (Wang et al., 2016) serving as the benchmarks. The proposed model significantly outperforms both benchmarks with respect to shape forecasting errors in ex-post and ex-ante forecasting scenarios. The proposed model also exhibits robustness against temperature forecast errors, as the ex-post and ex-ante results are closely aligned.

In conclusion, we have proposed a model to forecast load shapes during on-peak periods, based on a functional data approach. This study opens up new avenues for leveraging FDA for more precise and reliable load forecasts. One future research direction is to investigate the applicability of the proposed method to the shapes of other peaks, such as monthly peaks and annual peaks. Another direction is to investigate whether FDA can be beneficial to the traditional hourly load forecasting problems.

## References

- Amara-Ouali, Y., Fasiolo, M., Goude, Y. and Yan, H. (2023), 'Daily peak electrical load forecasting with a multi-resolution approach', *International Journal of Forecasting* **39**(3), 1272–1286.
- Chaouch, M. (2014), 'Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves', *IEEE Transactions on Smart Grid* **5**(1), 411–419.
- Feng, Y. H. and Ryan, S. M. (2016), 'Day-ahead hourly electricity load modeling by functional regression', *Applied Energy* **170**, 455–465.
- Gellings, C. W. (1985), 'The concept of demand-side management for electric utilities', *Proceedings of the IEEE* **73**(10), 1468–1470.
- Goia, A., May, C. and Fusai, G. (2010), 'Functional clustering and linear regression for peak load forecasting', *International Journal Of Forecasting* **26**(4), 700–711.
- Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., Reiss, P. T. and Cui, E. (2024), *refund: Regression with Functional Data*. R package version 0.1-37.  
**URL:** <https://CRAN.R-project.org/package=refund>
- Haben, S., Ward, J., Greetham, D. V., Singleton, C. and Grindrod, P. (2014), 'A new error measure for forecasts of household-level, high resolution electrical energy consumption', *International Journal of Forecasting* **30**(2), 246–256.
- Hong, T., Pinson, P. and Fan, S. (2014), 'Global energy forecasting competition 2012', *International Journal of Forecasting* **30**(2), 357–363.
- Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. and Hyndman, R. J. (2016), 'Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond', *International Journal of Forecasting* **32**(3), 896–913.
- Hong, T., Wang, P. and White, L. (2015), 'Weather station selection for electric load forecasting', *International Journal of Forecasting* **31**(2), 286–295.
- Hong, T., Xie, J. and Black, J. (2021), 'Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting', *International Journal of Forecasting* **37**(3), 1303–1303.
- Hyndman, R. J. and Koehler, A. B. (2006), 'Another look at measures of forecast accuracy', *International Journal of Forecasting* **22**(4), 679–688.

- Hyndman, R. J. and Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.
- Ivanescu, A., Staicu, A., Scheipl, F. and Greven, S. (2015), 'Penalized function-on-function regression', *Computational Statistics* **30**, 539–568.
- Kempton, W. and Tomic, J. (2005), 'Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy', *Journal of Power Sources* **144**(1), 280–294.
- Mestre, G., Portela, J., San Roque, A. and Alonso, E. (2020), 'Forecasting hourly supply curves in the Italian day-ahead electricity market with a double-seasonal SARMAHX model', *International Journal of Electrical Power & Energy Systems* **121**.
- O'Connell, N., Pinson, P., Madsen, H. and O'Malley, M. (2014), 'Benefits and challenges of electrical demand response: A critical review', *Renewable & Sustainable Energy Reviews* **39**, 686–699.
- Ramsay, J., Hooker, G. and Graves, S. (2005), *fda: Functional Data Analysis*. R package version 6.2.0.  
**URL:** <https://CRAN.R-project.org/package=fda>
- Ramsay, J. and Silverma, B. (2005), *Functional Data Analysis*, 2nd edn, Springer, New York.
- Saxena, H., Aponte, O. and McConky, K. T. (2019), 'A hybrid machine learning model for forecasting a billing period's peak electric load days', *International Journal of Forecasting* **35**(4), 1288–1303.
- Shang, H. (2013), 'Functional time series approach for forecasting very short-term electricity demand', *Journal of Applied Statistics* **40**, 152–168.
- Shukla, S. and Hong, T. (2024), 'BigDEAL Challenge 2022: Forecasting peak timing of electricity demand', *IET Smart Grid*.
- Singleton, C. and Grindrod, P. (2021), 'Forecasting for battery storage: Choosing the error metric', *Energies* **14**(19).
- Twitchell, J. (2019), 'A review of state-level policies on electrical energy storage', *Current Sustainable/Renewable Energy Reports* **6**(2), 35–41.
- Wang, P., Liu, B. and Hong, T. (2016), 'Electric load forecasting with recency effect: A big data approach', *International Journal of Forecasting* **32**(3), 585–597.
- Yang, D., Wang, W. and Hong, T. (2022), 'A historical weather forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for energy forecasting', *Solar Energy* **232**, 263–274.