Machine Learning ID

Functional Data Analysis for Energy Consumption Modelling using R

Novri Suhermi
Department of Mathematics and Statistics, Lancaster University UK

Saturday, 15 August 2020

Outline

- Functional Data
- Representing Functional Data
- Exploring functional data
- Regression for Functional Data
- Real World Application

Functional Data Analysis (FDA)

- FDA is an analysis of data that are in the form of functions, images and shapes, or more general objects [1]
- Simplest dataset form in FDA

$$x_n\left(t_{j,n}\right) \in \mathbb{R}, \quad t_{j,n} \in \left[T_1, T_2\right], \quad n = 1, 2, \ldots, N, \quad j = 1, \ldots, J_n$$

  - N curves are observed on a common interval $[T_1, T_2]$
  - The values of the curves are never known at all points $t \in [T_1, T_2]$
  - They are available only at some specific points $t_{j,n}$

Functional data

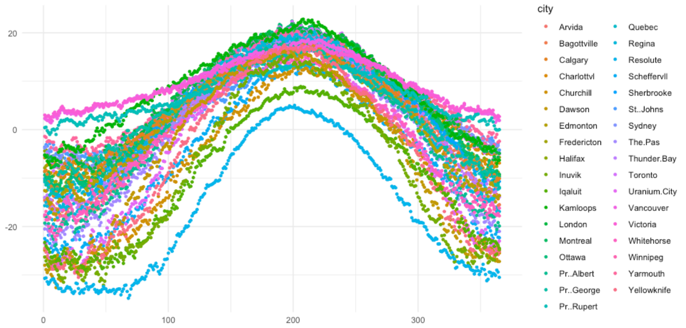
$x(t)$ can be represented using basis expansions [2]

$$x(t) = \sum_{k}^{K} c_k \phi_k(t) = \Phi(t)\mathbf{c}$$

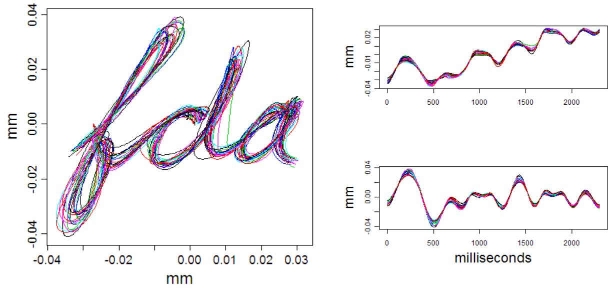
where $\Phi(t)$ is called basis system and $\mathbf{c}$ is a coefficient vector.

Functional data = the functions $x_i(t)$

Functional data analysis (FDA) = analysis of data that are functions



Daily temperatures in Canada



Handwriting data

Smoothing functional data by least squares

Recall

$$x(t) = \sum_{k}^{K} c_j \phi_k(t) = \Phi(t)\mathbf{c}$$

Minimize

$$\text{SMSSE}(\mathbf{y}|\mathbf{c}) = \sum_{j=1}^{n} \left[y_j - \sum_{k}^{K} c_k \phi_k \left(t_j \right) \right]^2 = (\mathbf{y} - \mathbf{\Phi c})'(\mathbf{y} - \mathbf{\Phi c})$$

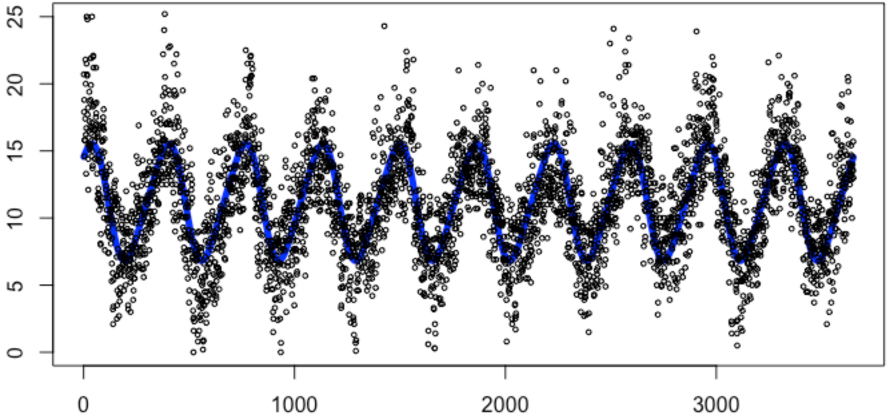
Solution

$$\hat{\mathbf{c}} = \left(\mathbf{\Phi}' \mathbf{\Phi} \right)^{-1} \mathbf{\Phi}' \mathbf{y}$$

Basis expansions

Fourier basis (for periodic data)

$$x(t) = c_1 + c_2 sin(\omega t) + c_3 cos(\omega t) + c_4 sin(2\omega t) + c_5 cos(2\omega t) + \ldots$$

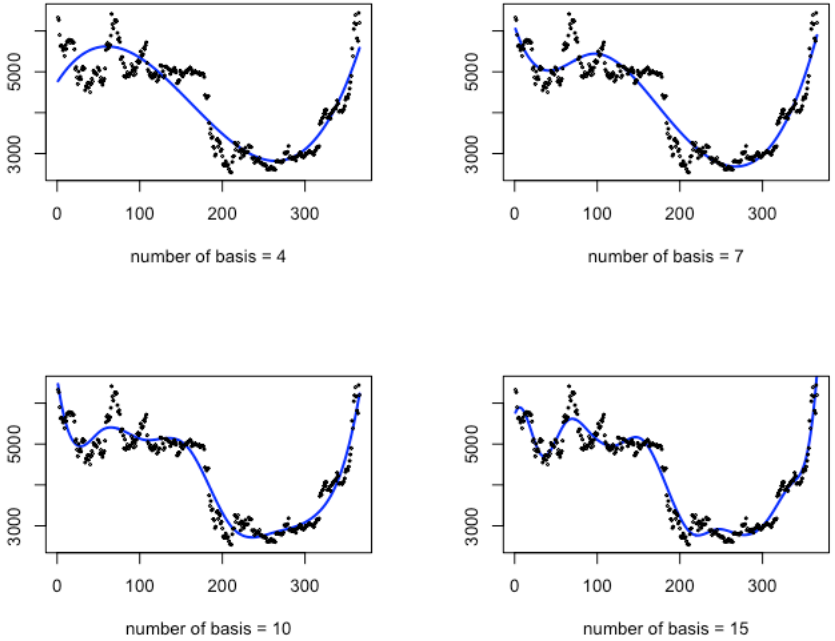


Smoothing using Fourier Basis

Basis expansions

B-Spline basis

$$x(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau)$$



Smoothing using B-Spline

Basis expansions

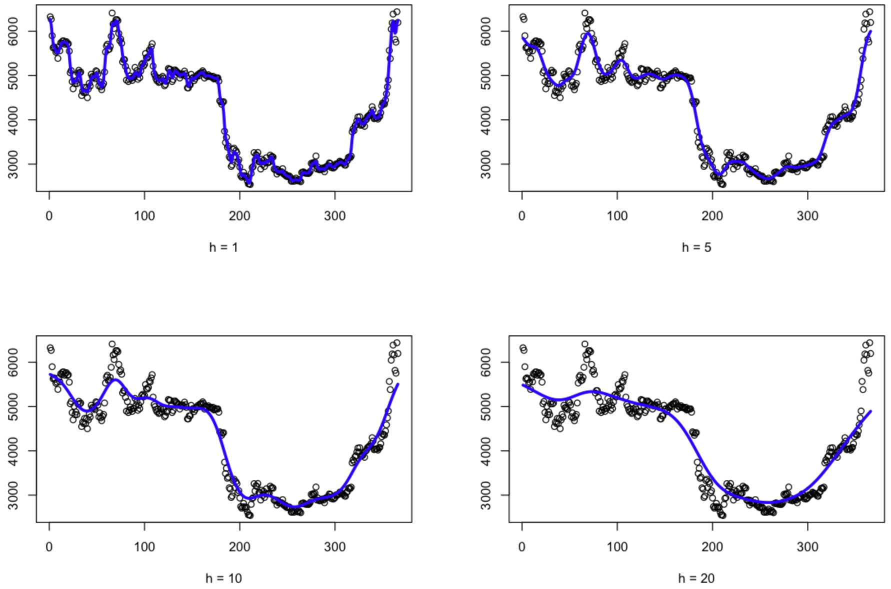
Kernel smoothing (local estimator)

$$x(t) = \sum_{j=1}^{n} S_j(t) y_j$$

where $S(t)$ is a Nadaraya-Watson kernel estimator given by

$$S_j(t) = \frac{\mathrm{Kern}[(t_j - t)/h]}{\sum_r \mathrm{Kern}[(t_r - t)/h]}$$

and $\mathrm{Kern}(.)$ is a kernel function.



Kernel smoothing for Bitcoin price

Summary statistics for functional data

Functional mean

$$\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t)$$

Functional variance

$$\text{var}_X(t) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t) - \bar{x}(t)]^2$$

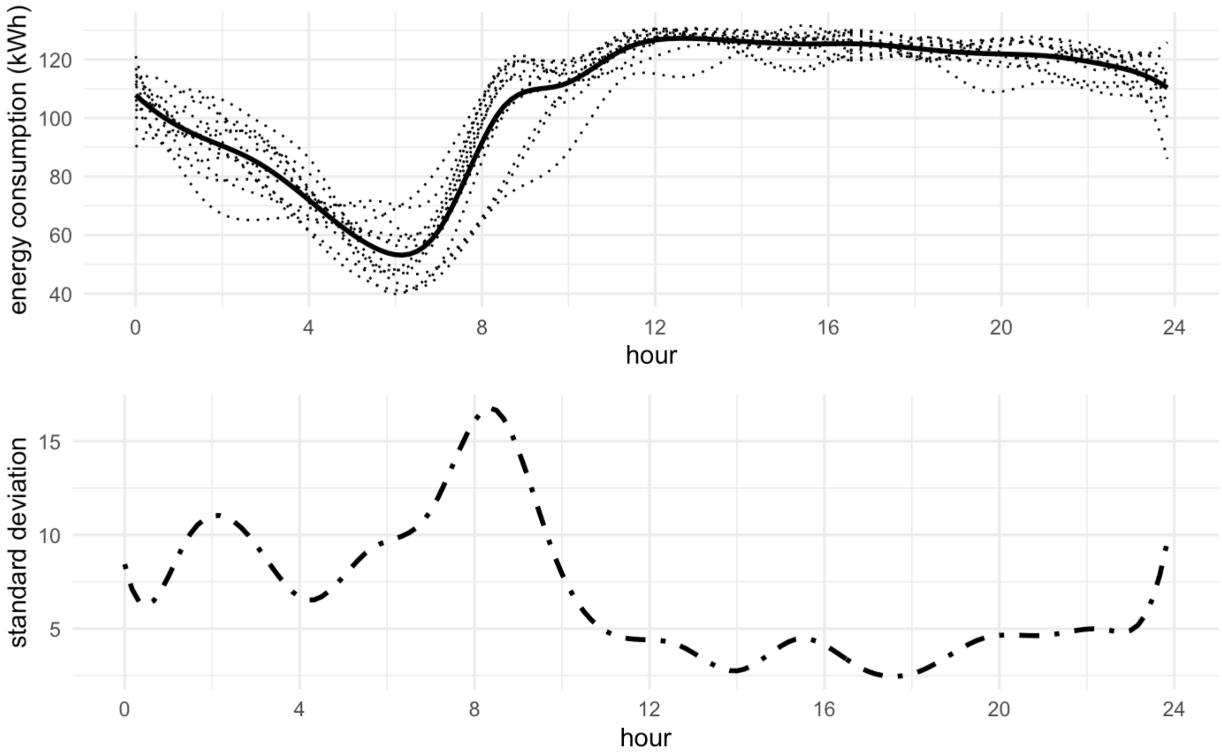
Functional covariance

$$\text{cov}_X(t_1, t_2) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$

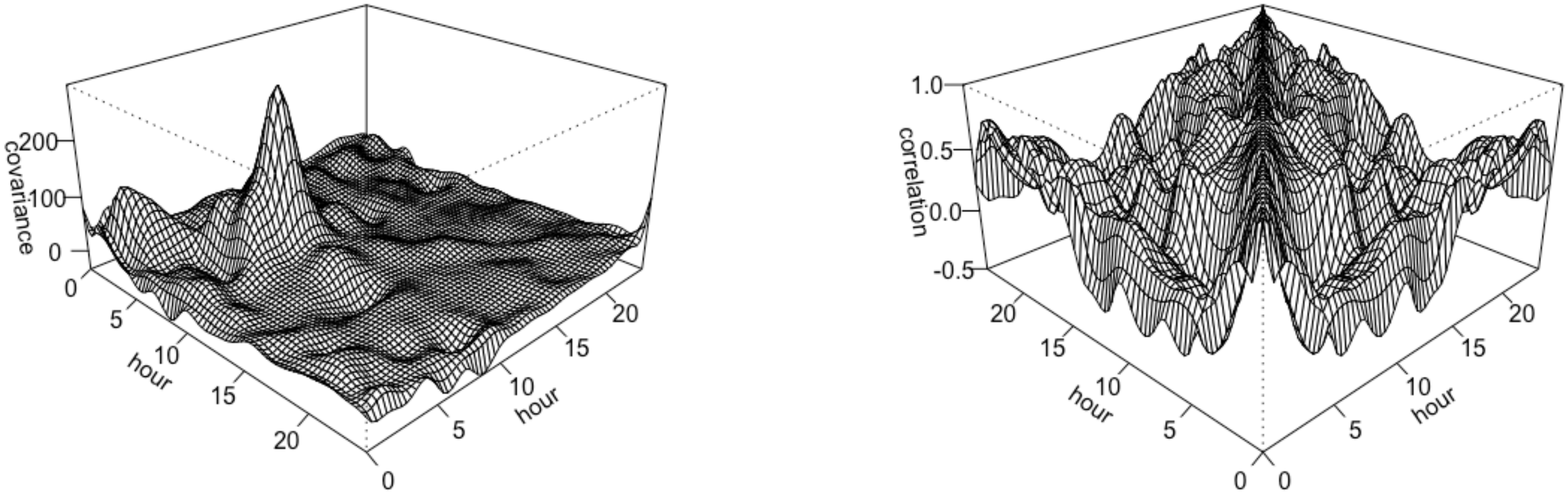
Functional correlation

$$\text{cov}_X(t_1, t_2) = \frac{\text{cov}_X(t_1, t_2)}{\sqrt{\text{var}_X(t_1)\text{var}_X(t_2)}}$$

Functional mean and variance example



Functional covariance and correlation example



Functional regression model

- Functional regression is a regression model that includes functional data as a response variable or a covariate
- General form of functional regression

$$y_i(t) = \mathbf{X}_i(t)'\boldsymbol{\beta}(t) + e_i(t), \quad i = 1, 2, \cdots, n$$

- Also called Time-Varying Coefficient (TVC) Regression [3]

Estimation using local kernel method

- Minimize [4]

$$\arg\min_{\boldsymbol{\beta}(t),\boldsymbol{\beta}(t)^{(1)}} \sum_{i=1}^{T} \left[y(t) - \mathbf{X}(t)'\boldsymbol{\beta}(t) - (t_i - t)\mathbf{X}(t)'\boldsymbol{\beta}(t)^{(1)} \right]^2 K_b(t_i - t)$$

- where $K_b(t_i - t) = b^{-1} K\left(\frac{t_i - t}{b}\right)$ is kernel function with bandwith of b
- $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ (Gaussian kernel)
- Solution

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}(t) \\ \hat{\boldsymbol{\beta}}(t)^{(1)} \end{pmatrix} = \begin{pmatrix} S_{T,0}(t) & S_{T,1}^{\top}(t) \\ S_{T,1}(t) & S_{T,2}(t) \end{pmatrix}^{-1} \begin{pmatrix} T_{T,0}(t) \\ T_{T,1}(t) \end{pmatrix}$$

$$S_{T,s}(t) = \frac{1}{T} \sum_{i=1}^{T} \mathbf{X}(t_i)'\mathbf{X}(t_i) (t_i - t)^s K_b(t_i - t)$$

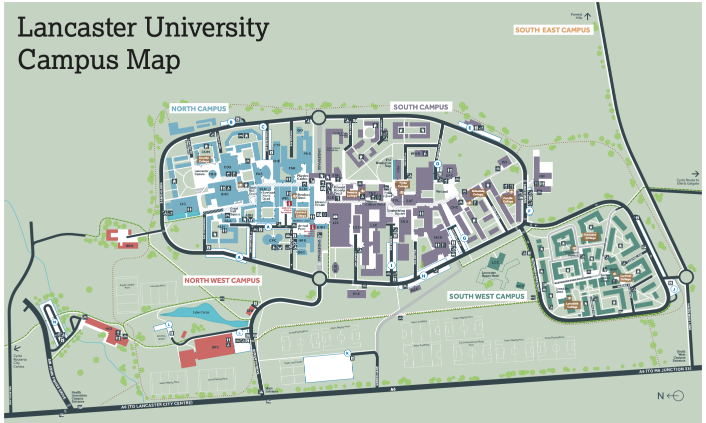
$$T_{T,s}(t) = \frac{1}{T} \sum_{i=1}^{T} \mathbf{X}(t_i)'\mathbf{X}(t_i) (t_i - t)^s K_b(t_i - t) \, y(t_i)$$

Application in Energy Consumption Modelling

- Building energy consumption is influenced by many factors
- Data is large, high frequency, and dense (real-time recorded)
- Necessary to build a model that could quantify the relationship between energy consumption and its factors
- Understanding the model will help us improve methods for reducing energy consumption in buildings
- Utilize the model to predict future consumption and cluster the buildings based on the factor effect's characteristics

Available Datasets

- Over 800 energy meters, measuring energy consumption in over 100 buildings
- Over 30,000 sensors in the Building Management System, including internal temperatures, ventilation and heating.
- Weather data from an on-site meteorology station
- Building occupancy data using WiFi data and room booking data
- Room and building data, using the asset management system



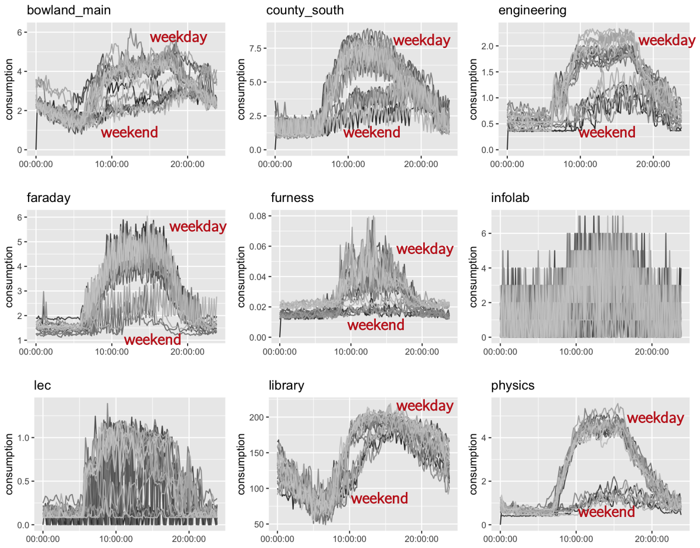
Lancaster University Campus Map

Data (current study)

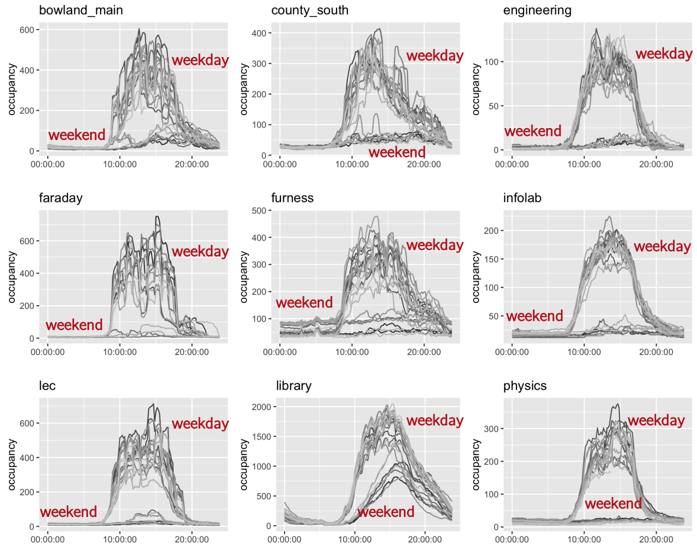
- Datasets: lighting consumption (kWh) and building occupancy (number of person) of 9 buldings at Lancaster University campus
- 25 day period data
- Recorded every 10 minutes
- Response variable (Y): lighting consumption
- Predictor variable (X): building occupancy

Results

- Daily time series plots of energy consumption and building occupancy in 9 building at Lancaster University Campus



Time series plot of energy consumption



Time series plot of building occupancy

Results

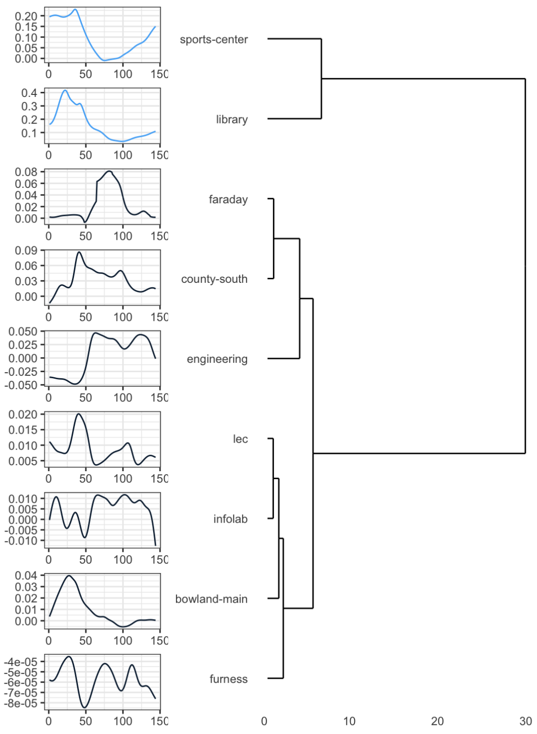
- Since there are different patterns between weekend and weekday, we add this information to the model
- Regression model with weekend-weekday indicator variable *D(t)*

$$y(t) = \beta_0(t) + \beta_1(t)X(t) + \beta_2 D(t) + \beta_3 D(t)X(t) + e(t)$$

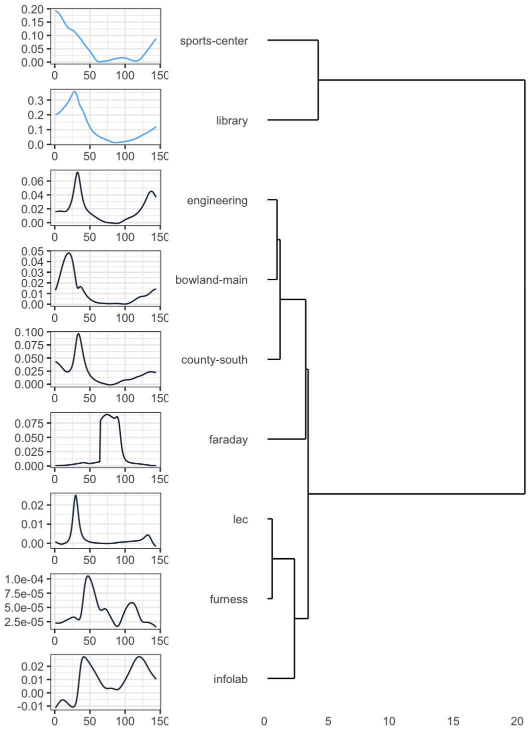
- We fit the model for each building and we compare the effect of building occupancy (model coefficients) to the energy consumption

Results

- In order to find the similarity between buildings, we conduct hierarchical time series clustering with dynamic time warping to the model coefficient curves



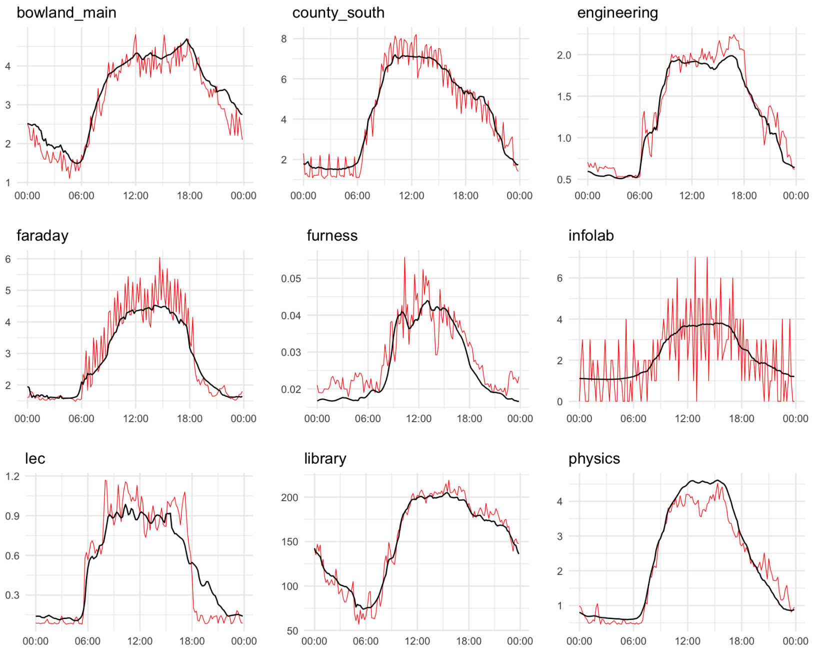
Occupancy effect curves in weekend



Occupancy effect curves in weekday

Results

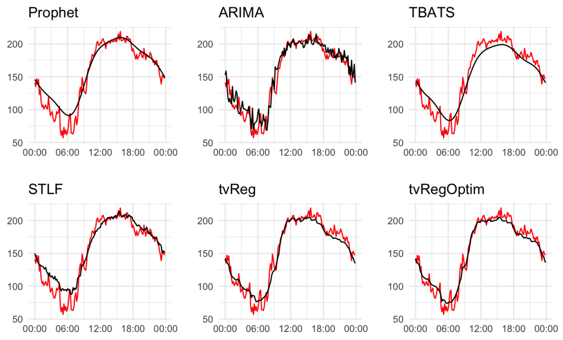
- We employ the models to forecast the future consumption



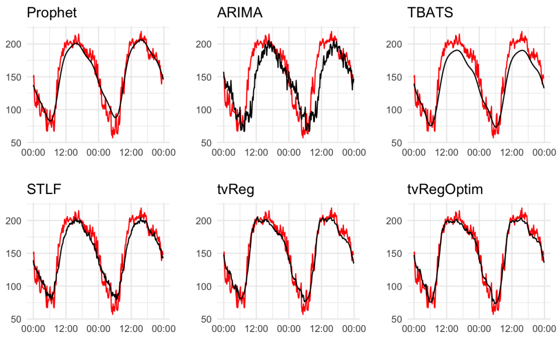
24-hour ahead forecast results (red: actual, black: predicted)

Results

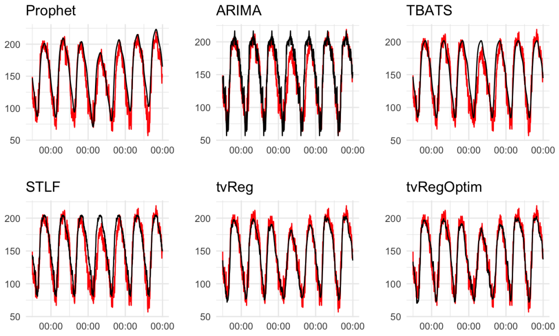
- For comparison, we perform forecast using other forecasting models:
    - Prophet, ARIMA, TBATS, STLF
- (Example) Forecast results for Library building



1-day ahead forecast results



2-day ahead forecast results



7-day ahead forecast results

*tvReg: TVC model with bandwidth of 1, tvRegOptim: TVC model with optimal bandwidth

Results

- Time series cross-validation with rolling window for Library Building
- Forecast horizon $h = 24$ hours

	Root Mean Square Error (RMSE)					
Window	Prophet	TBATS	ARIMA	STLF	tvReg	tvRegOptim
1	25.611	20.365	26.065	31.555	12.556	12.130
2	28.528	16.671	27.206	25.744	9.581	9.897
3	25.617	20.308	16.771	20.098	10.743	11.127
4	11.551	14.860	30.391	14.084	10.997	10.399
5	14.119	13.318	13.217	14.176	9.988	9.122
Mean	21.085	17.104	22.730	21.131	**10.773**	**10.535**

- Functional regression models produce excellent forecast with the lowest error

References

[1] Wang, J. L., Chiou, J. M., & Müller, H. G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3, 257-295.

[2] Ramsay, J., and Silverman, B. (2005). *Functional data analysis 2nd Ed*. Springer. New York.

[3] Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757-779.

[4] Wu, H., & Zhang, J. T. (2006). Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches. John Wiley & Sons.

Thank you