

Missing Data Restoration In Power Systems

Yi Hu

yhu28@ncsu.edu

1, 2, 3, 4, 5, , 7, 8, 9, 10

Equation for n^{th} number

$$a(n) = n$$

1, 3, 5, 7, 9, , 13, 15, 17, 19

$$a(n) = 2n - 1$$

1, 4, 9, 16, 25, , 49, 64, 81, 100

$$a(n) = n^2$$

2, 3, 10, 15, 26, , 50, 63, 82, 99

$$a(n) = n^2 - (-1)^n$$

2, 7, 11, 12, 5, , 42, 3, 51, 24

$$a(n) = f(n)$$

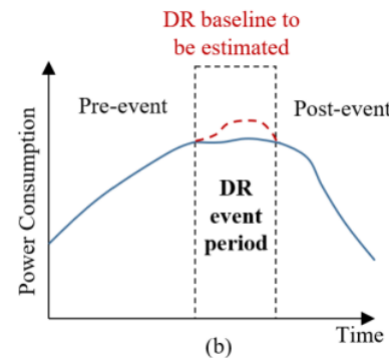
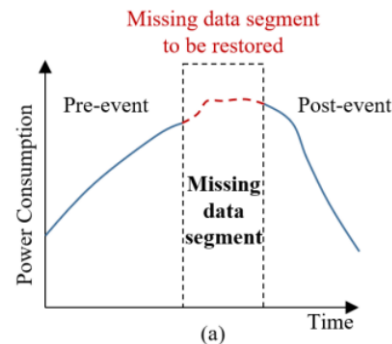
Reasons for missing data in electrical power engineering.

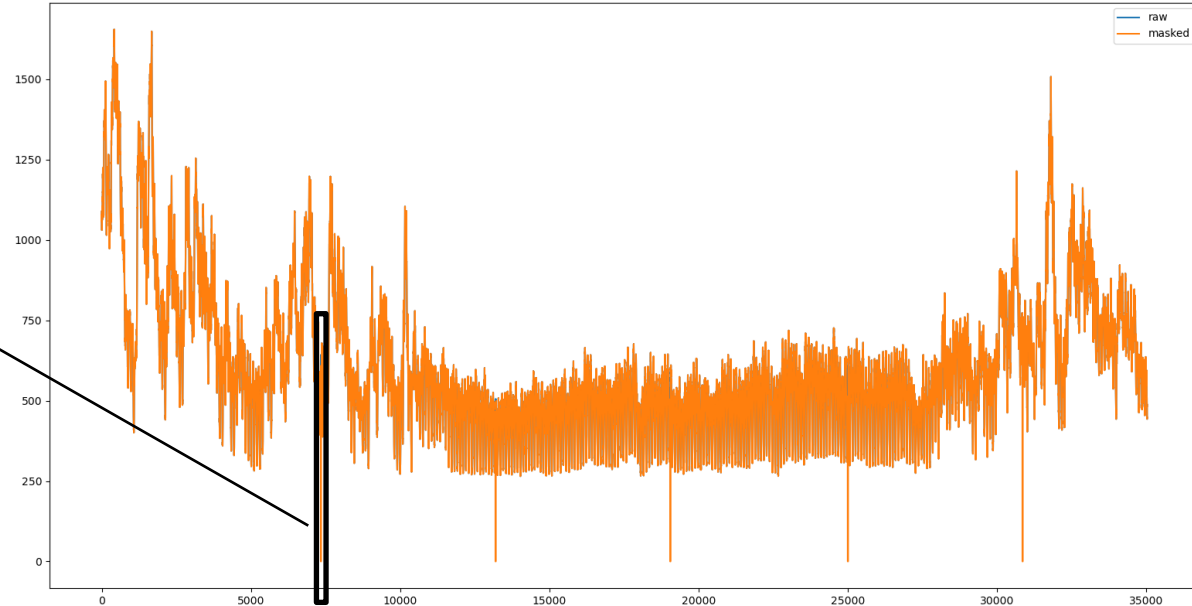
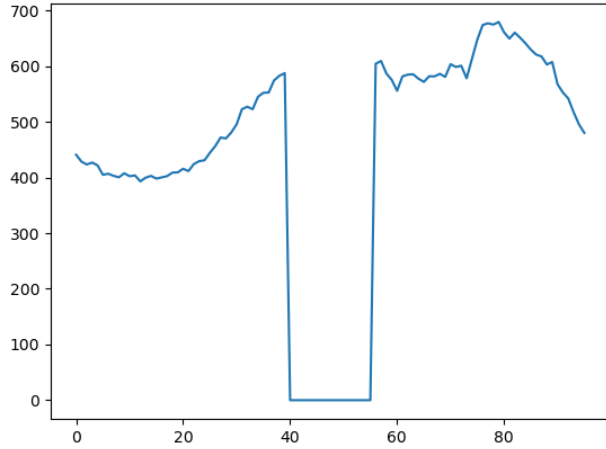
Equipment error

- Smart meter out of order.
- Loss of communication.
- Error in data storage.

Man-made missing data

- Demand Response (DR)
- Conservation Voltage Reduction (CVR)





- Interpolation
- Regression
- Deep Learning
 - Multilayer Perceptron (MLP)
 - GAN-based model
 - Transformer-based model

- Interpolation
- Regression
- Deep Learning
 - Multilayer Perceptron (MLP)
 - GAN-based model
 - Transformer-based model

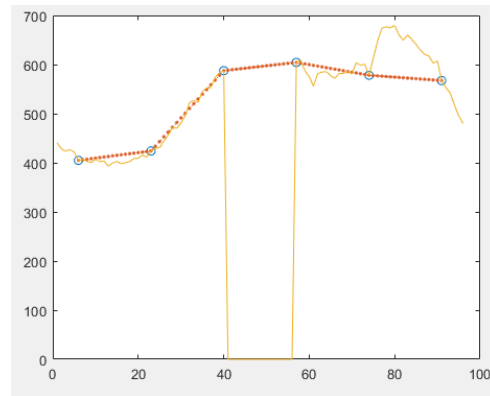
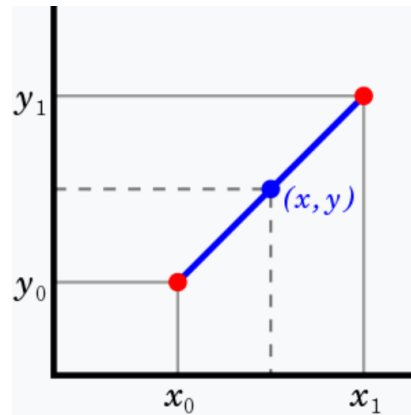
$$y = ax + b$$

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}$$

$$\begin{aligned} y &= y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0} \\ &= \frac{y_0(x_1 - x_0)}{x_1 - x_0} + \frac{y_1(x - x_0) - y_0(x - x_0)}{x_1 - x_0} \\ &= \frac{y_1x - y_1x_0 - y_0x + y_0x_0 + y_0x_1 - y_0x_0}{x_1 - x_0} \\ &= \frac{y_0(x_1 - x) + y_1(x - x_0)}{x_1 - x_0}, \end{aligned}$$

$$= \underbrace{\frac{(y_1 - y_0)}{x_1 - x_0}}_a x + \underbrace{\frac{x_1y_0 - x_0y_1}{x_1 - x_0}}_b$$

https://en.wikipedia.org/wiki/Linear_interpolation



1. Each quadratic goes through two consecutive data points:

$$\begin{aligned} a_1 x_0^2 + b_1 x_0 + c_1 &= f(x_0) \\ a_1 x_1^2 + b_1 x_1 + c_1 &= f(x_1) \\ &\vdots \\ a_i x_{i-1}^2 + b_i x_{i-1} + c_i &= f(x_{i-1}) \\ a_i x_i^2 + b_i x_i + c_i &= f(x_i) \\ &\vdots \\ a_n x_{n-1}^2 + b_n x_{n-1} + c_n &= f(x_{n-1}) \\ a_n x_n^2 + b_n x_n + c_n &= f(x_n) \end{aligned}$$

This condition gives $2n$ equations as there are n quadratics going through two consecutive data points.

2. The first derivatives of two consecutive quadratics are continuous at the common interior points. For example, the derivative of the first quadratic

$$a_1 x^2 + b_1 x + c_1$$

is

$$2a_1 x + b_1$$

The derivative of the second quadratic

$$a_2 x^2 + b_2 x + c_2$$

is

$$2a_2 x + b_2$$

and the two are equal at the common interior point $x = x_1$, giving

$$2a_1 x_1 + b_1 = 2a_2 x_1 + b_2$$

$$2a_1 x_1 + b_1 - 2a_2 x_1 - b_2 = 0$$

Similarly, at the other interior points x_2, \dots, x_{n-1} ,

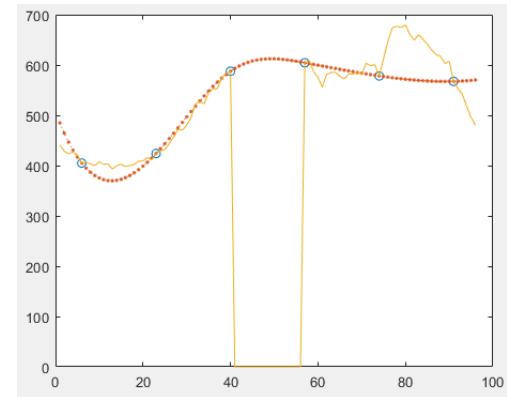
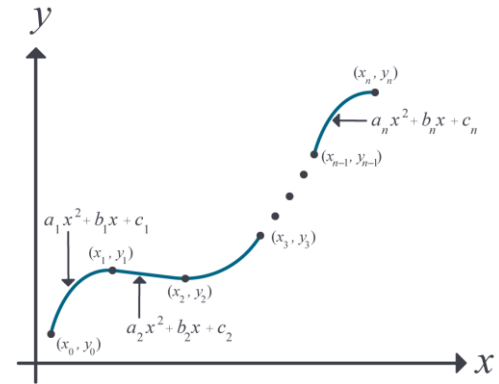
$$2a_2 x_2 + b_2 - 2a_3 x_2 - b_3 = 0$$

\vdots

$$2a_i x_i + b_i - 2a_{i+1} x_i - b_{i+1} = 0$$

\vdots

$$2a_{n-1} x_{n-1} + b_{n-1} - 2a_n x_{n-1} - b_n = 0$$



[https://math.libretexts.org/Under_Construction/Numerical_Methods_with_Applications_\(Kaw\)/5%3A_Interpolation/5.05%3A_Spline_Method_of_Interpolation](https://math.libretexts.org/Under_Construction/Numerical_Methods_with_Applications_(Kaw)/5%3A_Interpolation/5.05%3A_Spline_Method_of_Interpolation)

interp1

1-D data interpolation (table lookup)

R2023b

[collapse all in page](#)

Syntax

```
vq = interp1(x,v,xq)
vq = interp1(x,v,xq,method)
vq = interp1(x,v,xq,method,extrapolation)
```

```
vq = interp1(v,xq)
vq = interp1(v,xq,method)
vq = interp1(v,xq,method,extrapolation)
```

```
pp = interp1(x,v,method,'pp')
```

Description

`vq = interp1(x,v,xq)` returns interpolated values of a 1-D function at specific query points using linear interpolation. Vector `x` contains the sample points, and `v` contains the corresponding values, $v(x)$. Vector `xq` contains the coordinates of the query points.

example

If you have multiple sets of data that are sampled at the same point coordinates, then you can pass `v` as an array. Each column of array `v` contains a different set of 1-D sample values.

`vq = interp1(x,v,xq,method)` specifies an alternative interpolation method: 'linear', 'nearest', 'next', 'previous', 'pchip', 'cubic', 'v5cubic', 'makima', or 'spline'. The default method is 'linear'.

example

`vq = interp1(x,v,xq,method,extrapolation)` specifies a strategy for evaluating points that lie outside the domain of `x`. Set `extrapolation` to 'extrap' when you want to use the `method` algorithm for extrapolation. Alternatively, you can specify a scalar value, in which case, `interp1` returns that value for all points outside the domain of `x`.

example

`vq = interp1(v,xq)` returns interpolated values and assumes a default set of sample point coordinates. The default points are the sequence of numbers from 1 to `n`, where `n` depends on the shape of `v`:

example

- When `v` is a vector, the default points are `1:length(v)`.
- When `v` is an array, the default points are `1:size(v,1)`.

Use this syntax when you are not concerned about the absolute distances between points.

`vq = interp1(v,xq,method)` specifies any of the alternative interpolation methods and uses the default sample points.

`vq = interp1(v,xq,method,extrapolation)` specifies an extrapolation strategy and uses the default sample points.

`pp = interp1(x,v,method,'pp')` returns the piecewise polynomial form of $v(x)$ using the `method` algorithm.

https://www.mathworks.com/help/matlab/ref/interp1.html#btwp6lt-2_1

- Interpolation
- Regression
- Deep Learning
 - Multilayer Perceptron (MLP)
 - GAN-based model
 - Transformer-based model

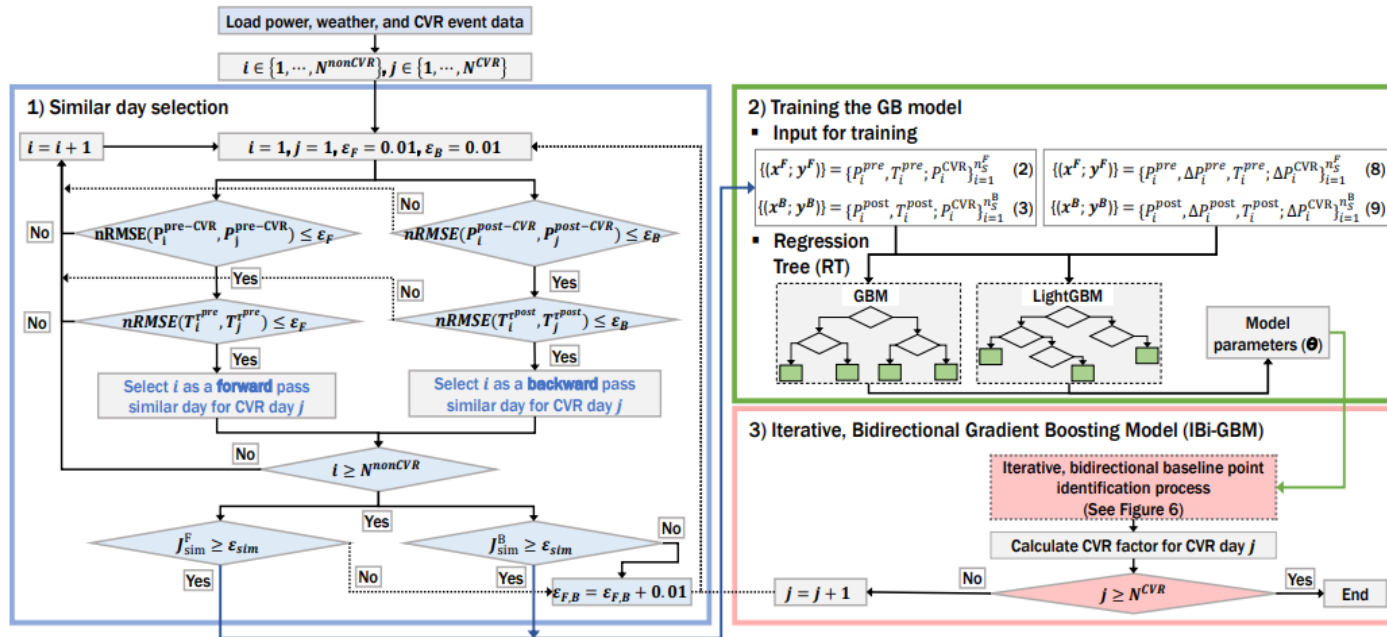
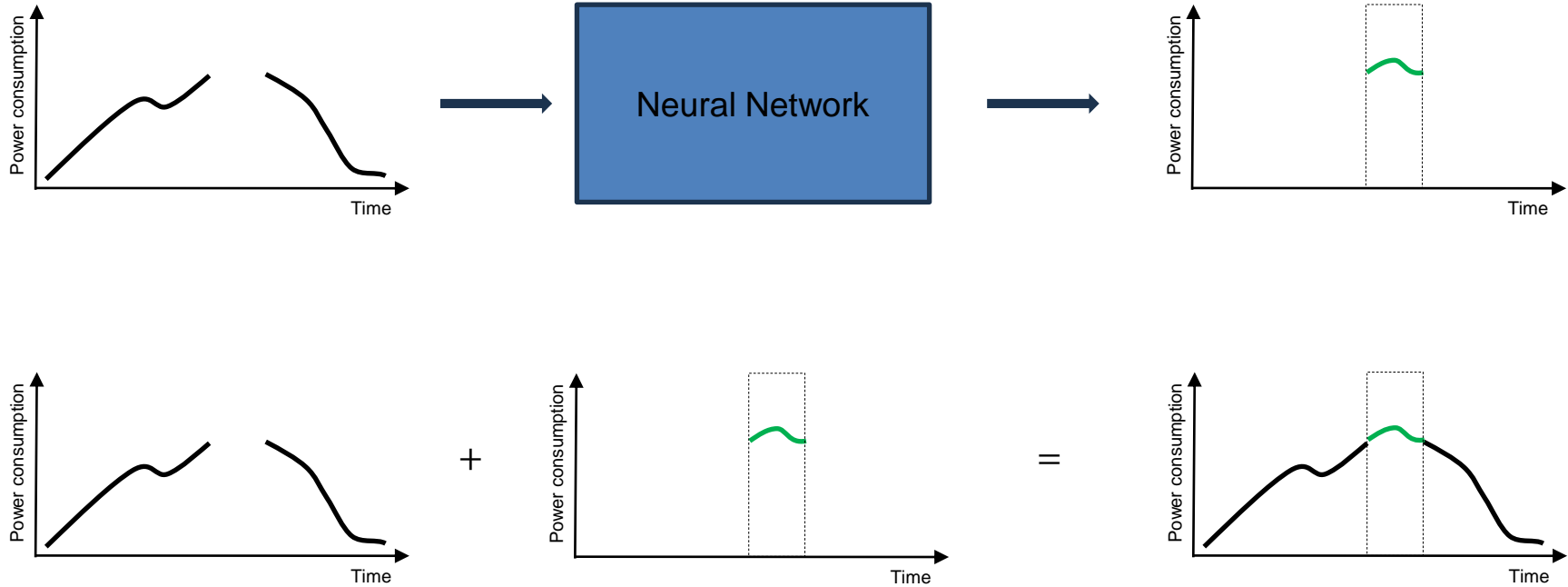


Fig. 2. A three-stage framework proposed for CVR baseline estimation, involving: 1) selecting similar days, 2) training the GB model, and 3) implementing IBi-GBM.

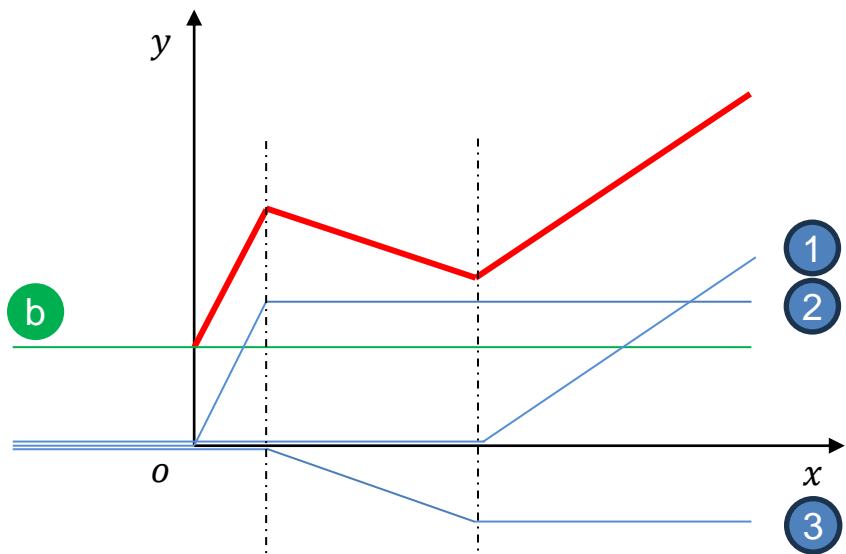
Han Pyo Lee, Lidong Song, Yiyang Li, Ning Lu, Di Wu, PJ Rehm, Matthew Makdad, Edmond Miller, "An Iterative Bidirectional Gradient Boosting Algorithm for CVR Baseline Estimation" 23PESGM0022, *Proc. of 2023 IEEE PES General Meeting*, Available online at: <http://arxiv.org/abs/2211.03733>. 23PESGM0022

- Interpolation
- Regression
- Deep Learning
 - Multilayer Perceptron (MLP)
 - GAN-based model
 - Transformer-based model

$$x \xrightarrow{\quad} f \xrightarrow{\quad} y = f(x)$$



red curve = constant + sum of a set of blue curves



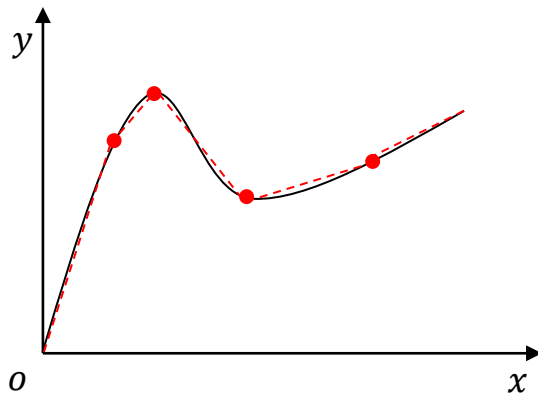
Arbitrary non-linear function:

Hard Sigmoid

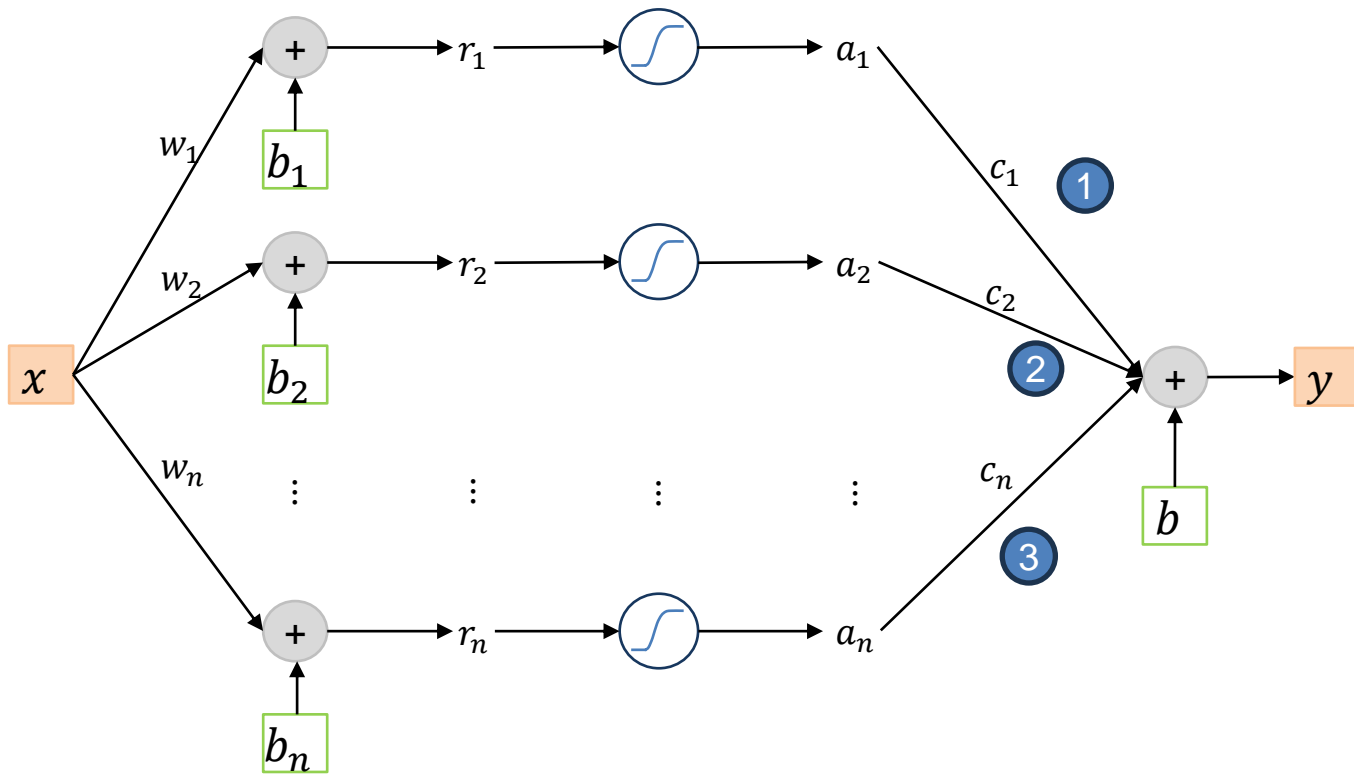


Sigmoid function

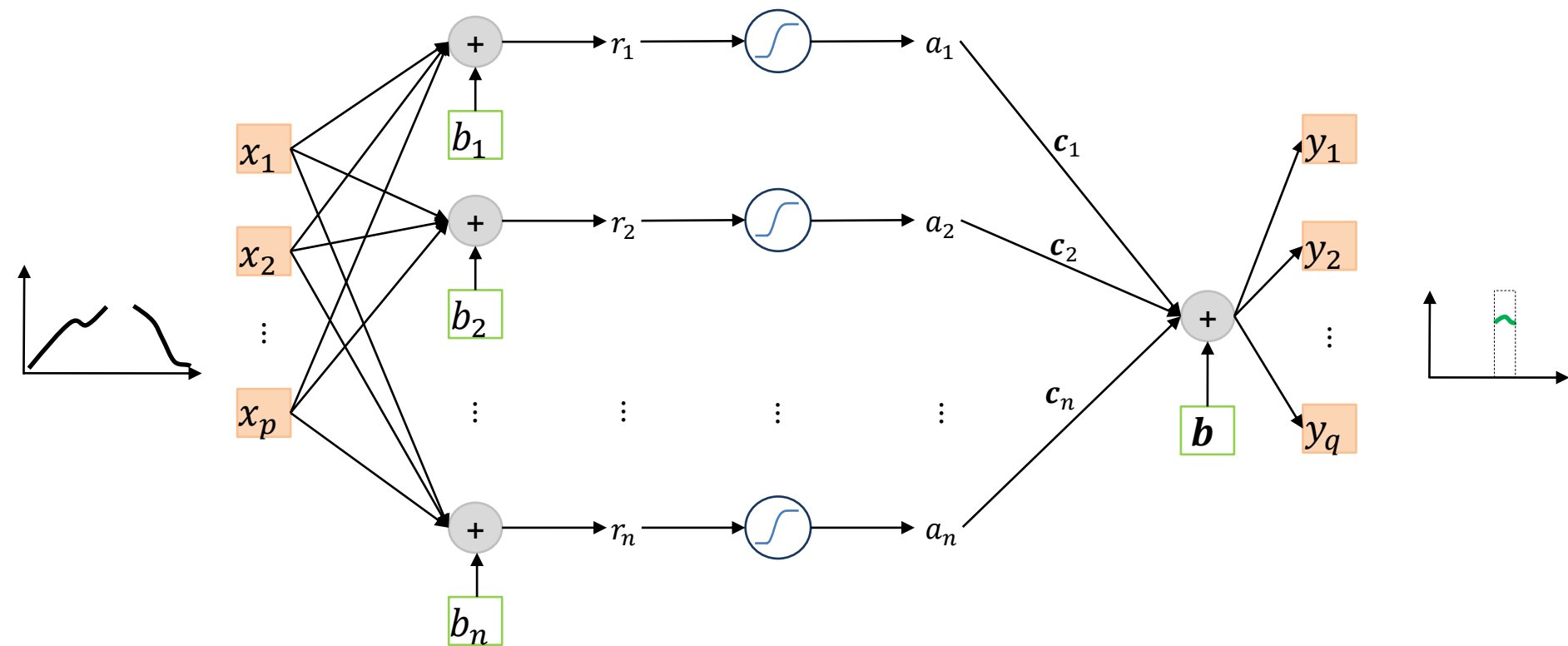
$$y = c \frac{1}{1 + e^{-(wx+b)}} = c \text{ sigmoid}(wx + b)$$



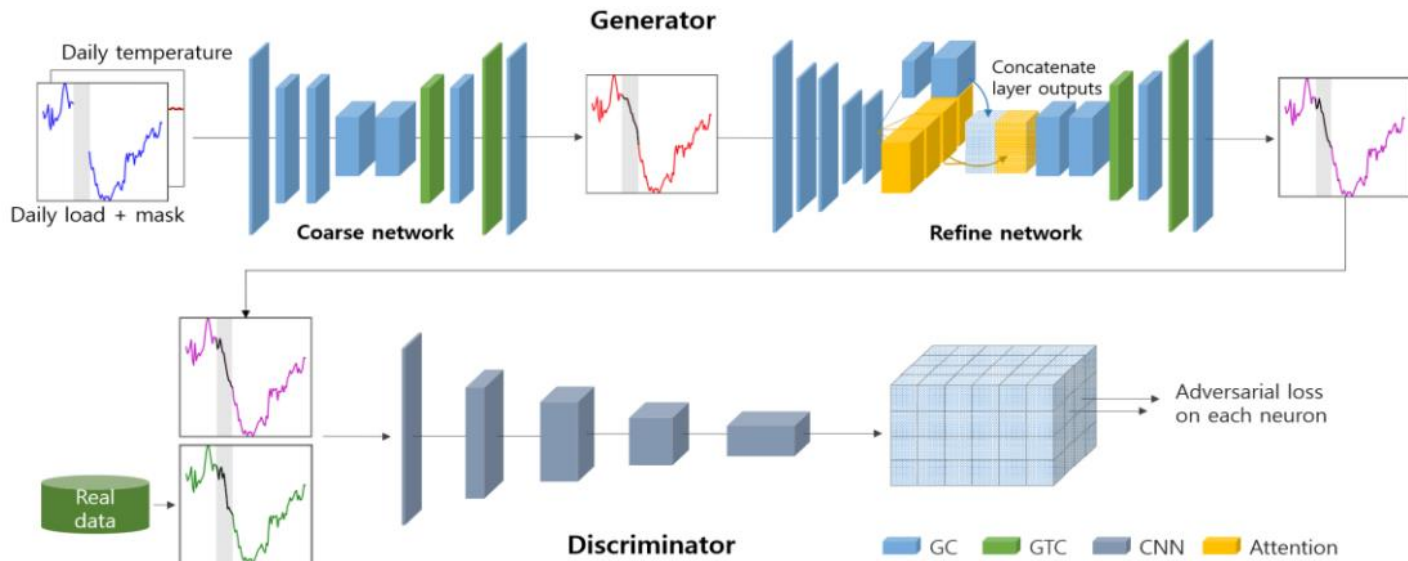
$$f(x) = b + \sum_i c_i \text{ sigmoid}(w_i x + b_i)$$



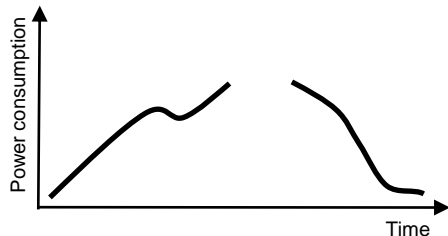
$$f(x) = b + \sum_i c_i \operatorname{sigmoid}(w_i x + b_i) \quad \longrightarrow \quad y = \mathbf{c}^T \sigma(\mathbf{w}x + \mathbf{b}) + b$$



Generative Adversarial Network



Yiyan Li, Lidong Song, Yi Hu, Hanpyo Lee, Di Wu, PJ Rehm, Ning Lu, "Load Profile Inpainting for Missing Load Data Restoration and Baseline Estimation," in *IEEE Transactions on Smart Grid*, doi: 10.1109/TSG.2023.3293188.

**Similar to language:**

- Time sequence data
- Like a missing word filling task
- Transformer based model has dominated Natural Language Processing (NLP) problems.
- Use Mask Language Model

I have lived in ____ for five years, so I can speak fluent French.

I have lived in ____ for five years, so I can speak fluent French.



Transformer
based model



I have lived in France for five years, so I can speak fluent French.

90%

I have lived in Quebec for five years, so I can speak fluent French.

8%

I have lived in England for five years, so I can speak fluent French.

1%

:

I have lived in apple for five years, so I can speak fluent French.

0.01%

Yi

"I have lived in ____ for five years, so I can speak fluent French." Please fill in the blank.



"I have lived in France for five years, so I can speak fluent French."

Yi

Is there another possible answer?

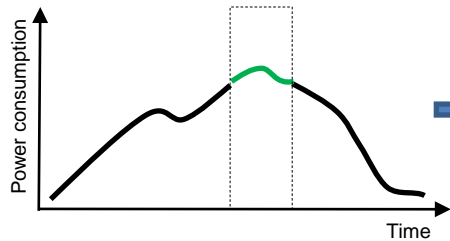


Certainly! Another possible answer could be:

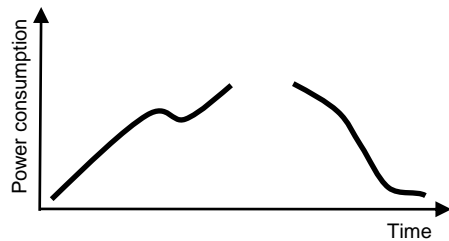
"I have lived in Quebec for five years, so I can speak fluent French."

In this case, Quebec is a province in Canada where French is one of the official languages, and living there for five years would likely provide ample opportunity to learn and become fluent in French.

Training and Testing: Non-CVR days

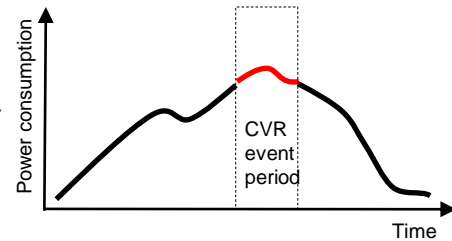


I have lived in **France** for five years, so I can speak fluent French.



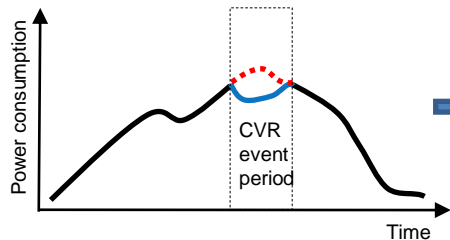
I have lived in ____ for five years, so I can speak fluent French.

Transformer based model

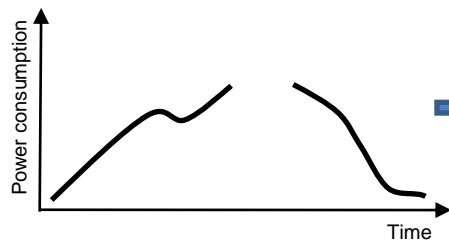


I have lived in **France** for five years, so I can speak fluent French.

Application: CVR days

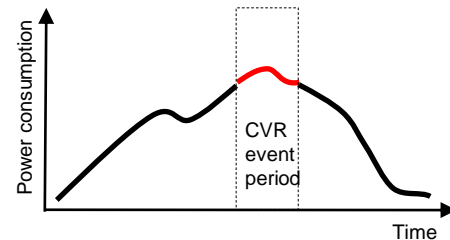


I have lived in **Japan** for five years, so I can speak fluent French.

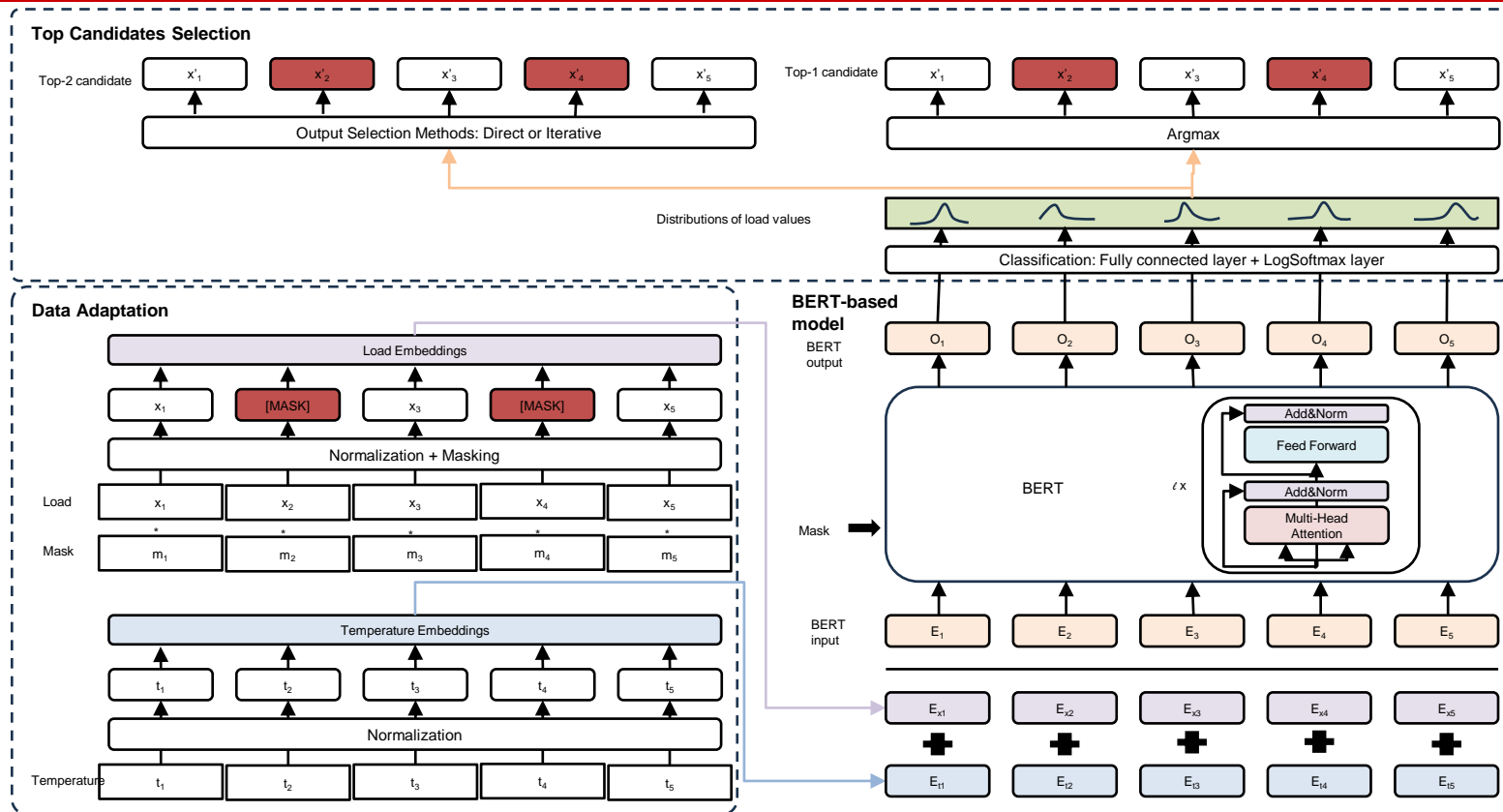


I have lived in ____ for five years, so I can speak fluent French.

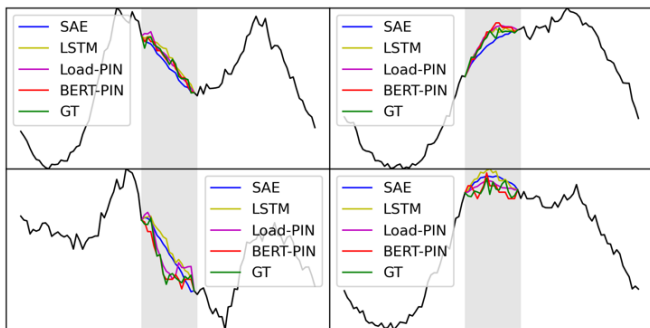
BERT-PIN



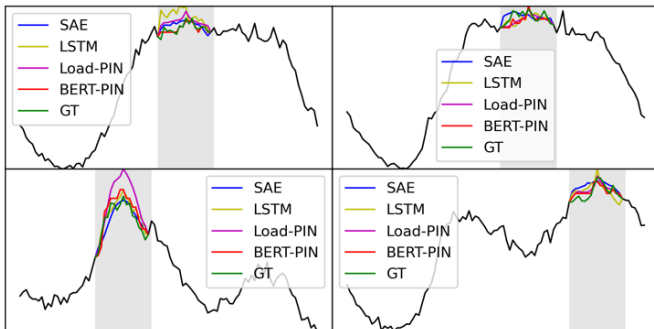
I have lived in **France** for five years, so I can speak fluent French.



Yi Hu, Kai Ye, Hyeonjin Kim, and Ning Lu, "BERT-PIN: A BERT-based Framework for Recovering Missing Data Segments in Time-series Load Profiles", available at <http://arxiv.org/abs/2310.17742>



	SAE (%)	LSTM (%)	Load-PIN (%)	BERT-PIN (%)	Improvement
MPE	2.089	2.144	1.683	1.612	3.74%
RMSE	0.8882	0.9032	0.9067	0.6992	21.27%
PKE	0.7443	0.6281	0.5731	0.5438	5.11%
VLE	0.6235	0.5774	0.5764	0.4815	16.46%
EGYE	1.440	1.516	1.059	0.8879	16.16%
FCE	1.781	1.743	1.452	1.467	-1.03%



	SAE (%)	LSTM (%)	Load-PIN (%)	BERT-PIN (%)	Improvement
MPE	2.231	2.414	1.670	1.523	8.80%
RMSE	1.035	1.112	0.7951	0.7404	6.88%
PKE	1.065	0.8491	0.6183	0.5130	17.10%
VLE	0.8687	0.8991	0.6185	0.5870	5.09%
EGYE	1.525	1.762	1.165	0.8410	27.81%
FCE	2.138	2.046	1.615	1.509	6.56%

Homework 6 Assignment

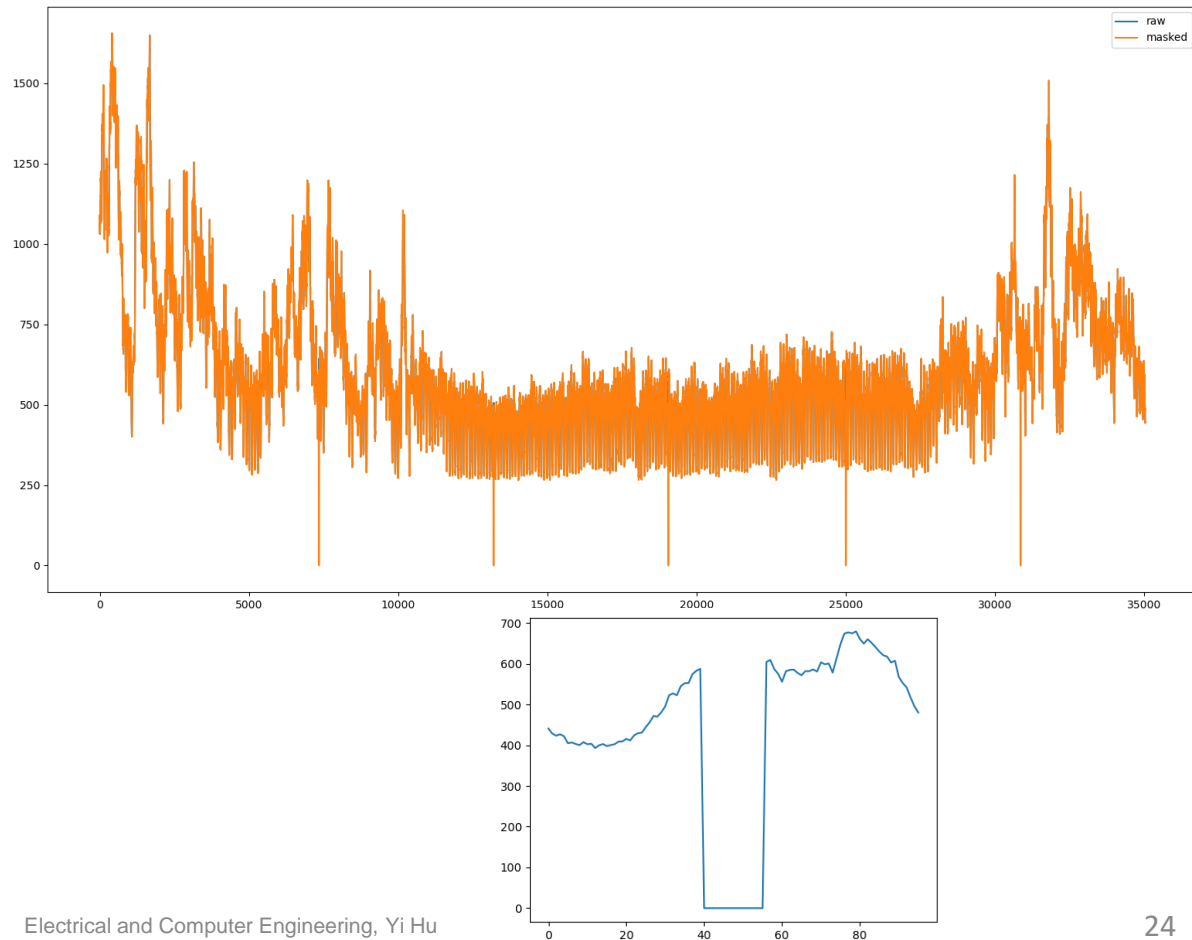
Data description

- 15-min data from 2018-1-1 00:00:00 to 2018-12-31 23:45:00
- 5 days missing data (set to zeros)
- Missing days: 2018-3-18, 2018-5-18, 2018-7-18, 2018-9-18, 2018-11-18
- Missing hours: 10:00-13:45 (4 hours, 16 values)

Task:

Restore all the missing data. Make the error as small as possible.

You can use whatever method.



Data and example code: <https://github.com/hughwln/ECE753-Computational-Methods-for-Power-Systems>

The screenshot shows the GitHub interface for the repository 'ECE753-Computational-Methods-for-Power-Systems'. At the top, the repository name is displayed with a 'Public' badge. To the right are buttons for 'Pin' and 'Unwatch' (with a count of 1). Below this, the 'master' branch is selected, showing '1 Branch' and '0 Tags'. A search bar 'Go to file' and buttons for 'Add file' and '<> Code' are present. The file list shows the following items:

File/Folder	Commit Message	Commit Time
BERT-PIN(Python)	first commitment	5 days ago
data	first commitment	5 days ago
matlab_code	first commitment	5 days ago
README.md	Create README.md	5 days ago
data_plot.PNG	first commitment	5 days ago
result_bert-pin.PNG	first commitment	5 days ago

Below the file list, the 'README' file is expanded, showing the text: 'Homework setup for ECE753 at North Carolina State University.'

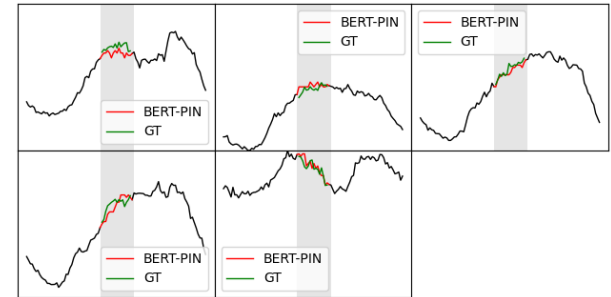
1. Please submit a single **.csv** file with a **single column**. Only include **80 restored missing values** (16 values per day * 5 days) in the order of time.

- Name your file as **“firstname_lastname_result.csv”**.
- Any format or name issue will lead to point reduction.
- Your result will be checked by the script “data/test.m”, error is calculated by MPE.
- An example of data format is given in “matlab_code/output.csv”
- Grading criteria:

```
% mpe < 3%           10/10
% 3% <= mpe < 5%     9/10
% 5% <= mpe < 10%    8/10
% 10% <= mpe < 20%   7/10
% 20% <= mpe         6/10
```

2. Submit a **PDF** file and name as **“firstname_lastname_method.pdf”**

- Description of your methodology with in 100 words.
- Plot restored load profiles in the five days. Use different color in the missing hours. Example is given on the right figure. You don't need to plot the “GT” since you don't have it.



3. (optional) Please submit your source code, if your method needs coding work.

- Put all your code in a folder name as **“firstname_lastname_src”**, and compress to a **zip** file with the same name.
- To reduce the file size, don't include any data file in this part, code only.