



```
In [1]: import numpy as np
import pandas
import matplotlib.pyplot as plt
import sklearn.linear_model
π = np.pi
```

Climate dataset challenge

- What is the rate of temperature increase in Cambridge?
- Are temperatures increasing at a constant rate, or has the increase accelerated?
- How do results compare across the whole of the UK?

Your task is to answer these questions using appropriate linear models, and to produce elegant plots to communicate your findings. Please submit a Jupyter notebook, or a pdf. Include explanations of what your models are, and of what your plots show.

The dataset is from <https://www.metoffice.gov.uk/pub/data/weather/uk/climate/>. Code for retrieving the dataset is given at the bottom.

Are temperatures increasing at a constant rate, or has the increase accelerated?

Don't stop with just one model — keep inventing!

But how shall we choose between all these models?

SIMPLE CONSTANT-RATE INCREASE

$$Y \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t$$

QUADRATIC DEPENDENCE

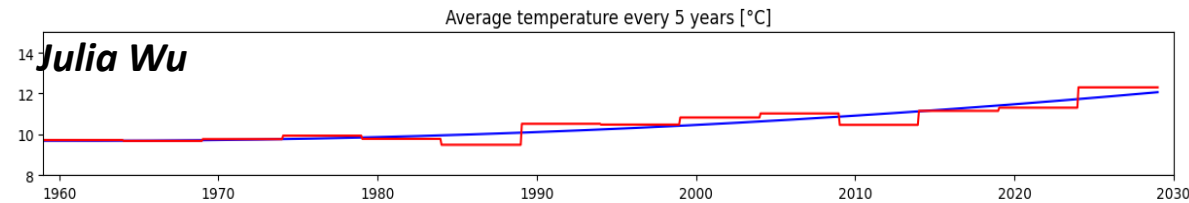
$$Y \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t + \delta t^2$$

STEP FUNCTION

$$Y \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \sum_{b \in \text{bins}} \eta_b 1_{t \in b}$$

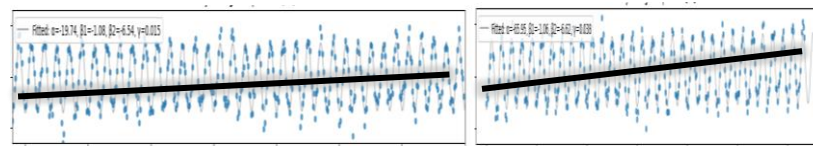
GOING BEYOND LINEAR MODELS

PIECEWISE LINEAR

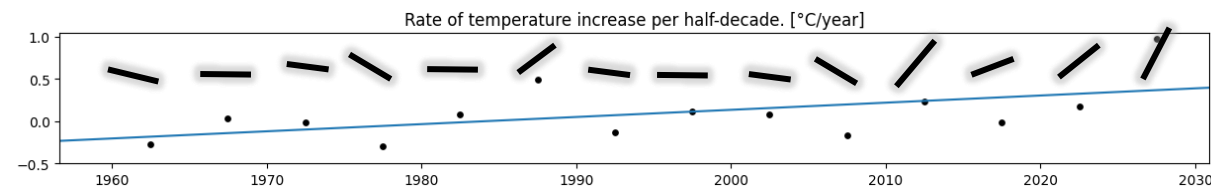


Phil Zhao

polynomial: $\mu = a \sin(2\pi(t+\varphi)) + \gamma t^x + \gamma t + c$
exponential: $\mu = a \sin(2\pi(t+\varphi)) + \gamma e^{\lambda t} + \gamma t + c$



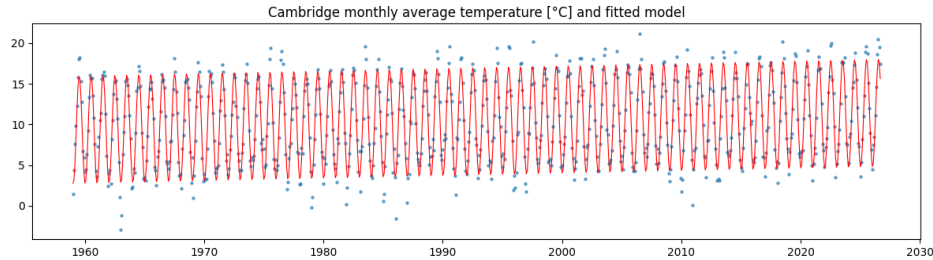
Eric Zhou
Sai Tadepalli



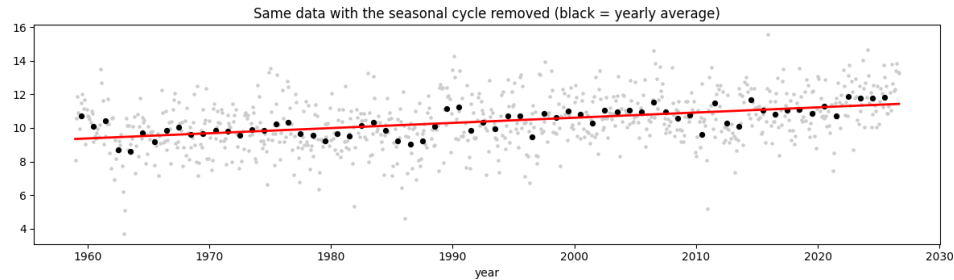
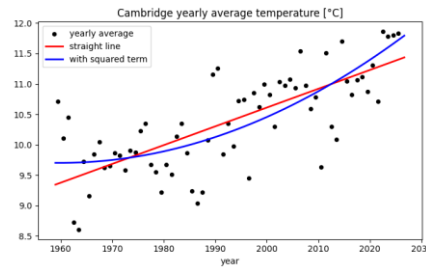
Leanne Pratlong

Are temperatures increasing at a constant rate, or has the increase accelerated?

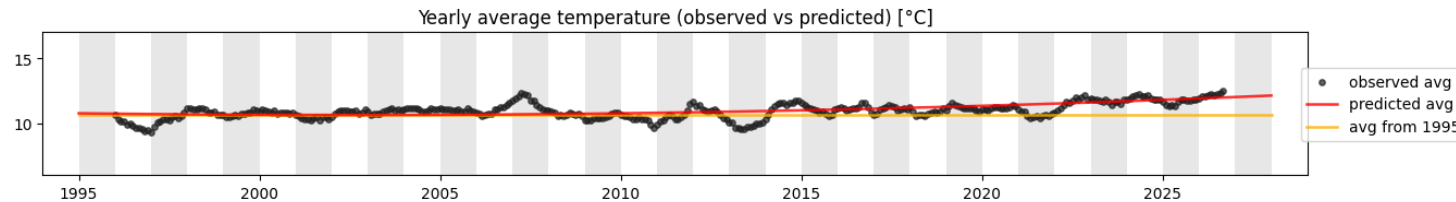
It's great to plot the data, to compare it to the model.
But the data on its own is too cluttered to be intelligible.



To make it readable, we need cleverer visualizations.



Sai Tadepalli



Krzysztof Matusiak

How do results compare across the whole of the UK?

We can do a 'for' loop to look at each location in turn ...

Theo Toma

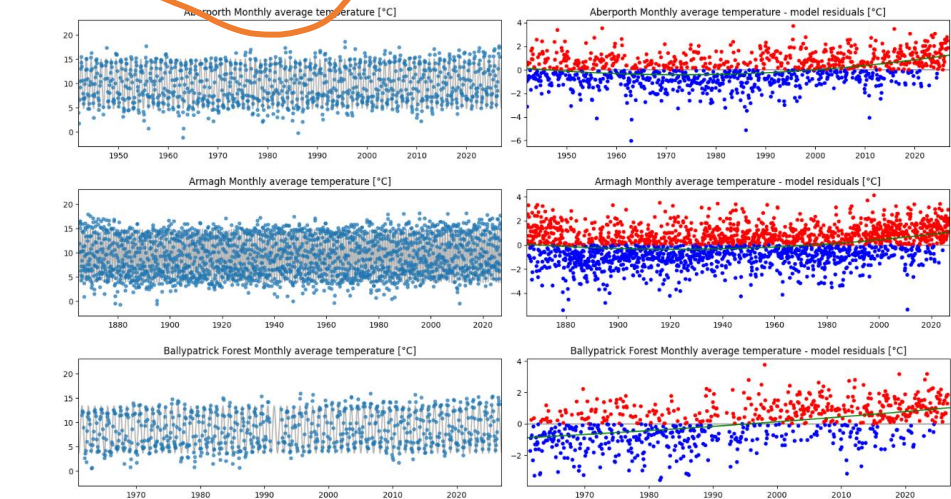
This expresses the idea "the parameters may be different from one station to the next".
It's better to express this within the model, e.g.

$$Temp \approx \alpha_{station} + \gamma_{station}t + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t).$$

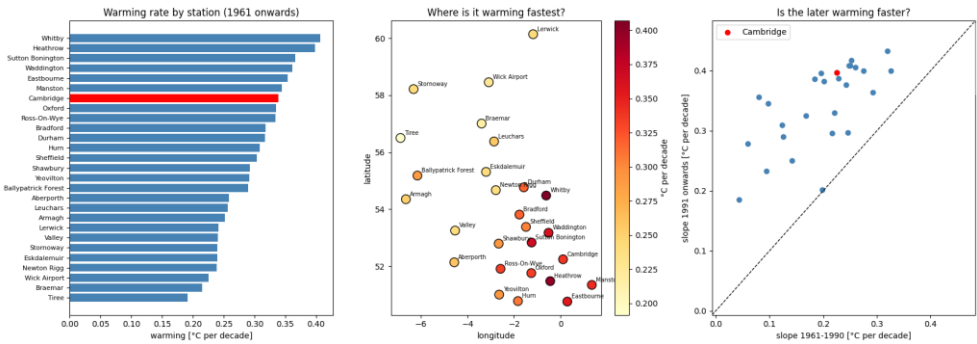
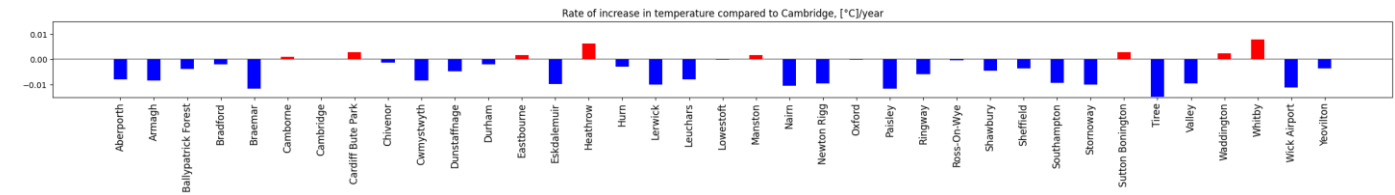
This gives you a bigger design space. It lets you say e.g. 'Maybe the linear bit is different from station to station, but I believe the sinusoidal part is the same' (as my model does).

Sai Tadepalli

Leanne Pratlong



... and communicate what's happening by plotting just the fitted coefficients





```
In [1]: import numpy as np
import pandas
import matplotlib.pyplot as plt
import sklearn.linear_model
π = np.pi
```

Climate dataset challenge

- What is the rate of temperature increase in Cambridge?
- Are temperatures increasing at a constant rate, or has the increase accelerated?
- How do results compare across the whole of the UK?

Your task is to answer these questions using appropriate linear models, and to produce elegant plots to communicate your findings. Please submit a Jupyter notebook, or a pdf. Include explanations of what your models are, and of what your plots show.

The dataset is from <https://www.metoffice.gov.uk/pub/data/weather/uk/climate/>. Code for retrieving the dataset is given at the bottom.

parameter estimation
model choice