Photo by Arun Clarke on Unsplash

The Butterfly Effect

A Case for Biodiversity using Machine Learning

What do we want to predict?

Using a **Time Series model**,

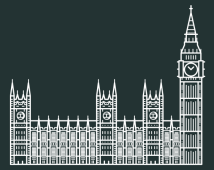
an **estimation of the butterfly population evolution** over the next years

In the context of **climate change**

What are we looking at?

65,000
butterfly sightings

2001 - 2020



London

From: UK Butterfly Monitoring Scheme

Butterfly icon - Alex Muravev | Calendar icon: Iulia Ardeleanu | UK icon: Lewis K-T

Dataset 1 is merged with:

Weather data

- Temperatures
- Rain
- Sun
- Air Frost

From: Met Office

Air Quality data

- Ground Level Ozone (O3)
- Nitrogen Dioxide (NO2)
- Sulphur Dioxide (SO2)
- Particles (PM10)
- Carbon Monoxide (CO)

From: London Air Quality Network

Weather icon: ratnaningsih nining | Air Quality icon: nakals

How to build the model?

Data Cleaning

Butterfly pop indicator

- Check for anomalies
- Fill Missing values
- Merge all datasets

>

Data Transformation

Butterfly population indicator

- From 65K Butterfly sightings
- To 1 data point / month: **Butterfly seen per Survey**

Stationarity

- Differencing
- Dickey Fuller test

Time Series Decomposition

- Trend
- Seasonality
- Residuals

Scaling

- MinMaxScaler

Lags

- 12/24 months
- => Impact of external factors on the next generations of butterflies

>

Data Selection

External Factors

- Granger Causality Test
- Cointegration

Preprocessing steps

(Depending on the model)

Data Cleaning

Butterfly pop indicator

- Check for anomalies
- Fill Missing values
- Merge all datasets

>

Data Transformation

Butterfly population indicator

- From 65K Butterfly sightings
- To 1 data point / month: **Butterfly seen per Survey**

Stationarity

- Differencing
- Dickey Fuller test

Time Series Decomposition

- **Trend**
- **Seasonality**
- **Residuals**

Scaling

- MinMaxScaler

Lags

- 12/24 months
- => Impact of external factors on the next generations of butterflies

>

Data Selection

External Factors

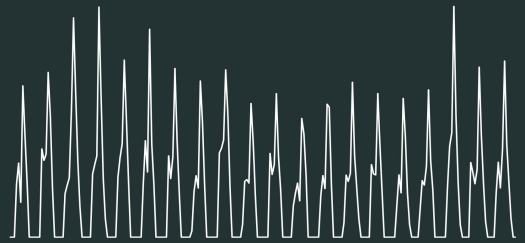
- Granger Causality Test
- Cointegration

Preprocessing steps

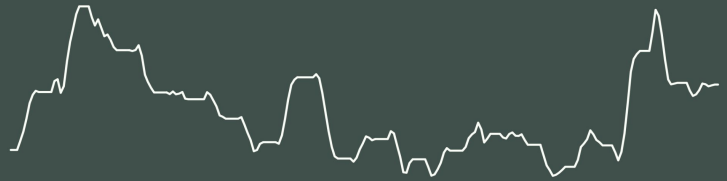
(Depending on the model)

1| Time Series decomposition

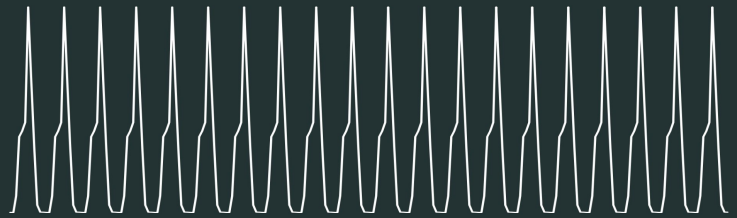
Original data



Trend



Seasonality



Residuals

1| Time Series decomposition

Trend

2| Monthly lags

Original data

	Butterflies per Survey	Temperature - No Lag
July 2005	7.8	16.3
...		
July 2006	7.6	15.5
...		
July 2007	7.1	15.6

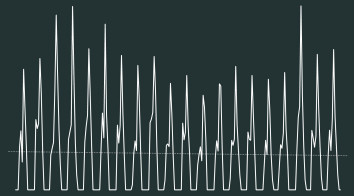
>

Lagged data for the model

Butterflies per Survey	Temperature - Lag 12
7.8	-
7.6	16.3
7.1	15.5

Model & Forecasts

Univariate models

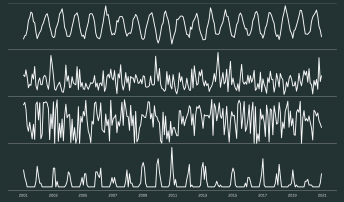


1 variable impact the predictions

> Butterflies per Survey

SARIMA / Prophet

Multivariate models

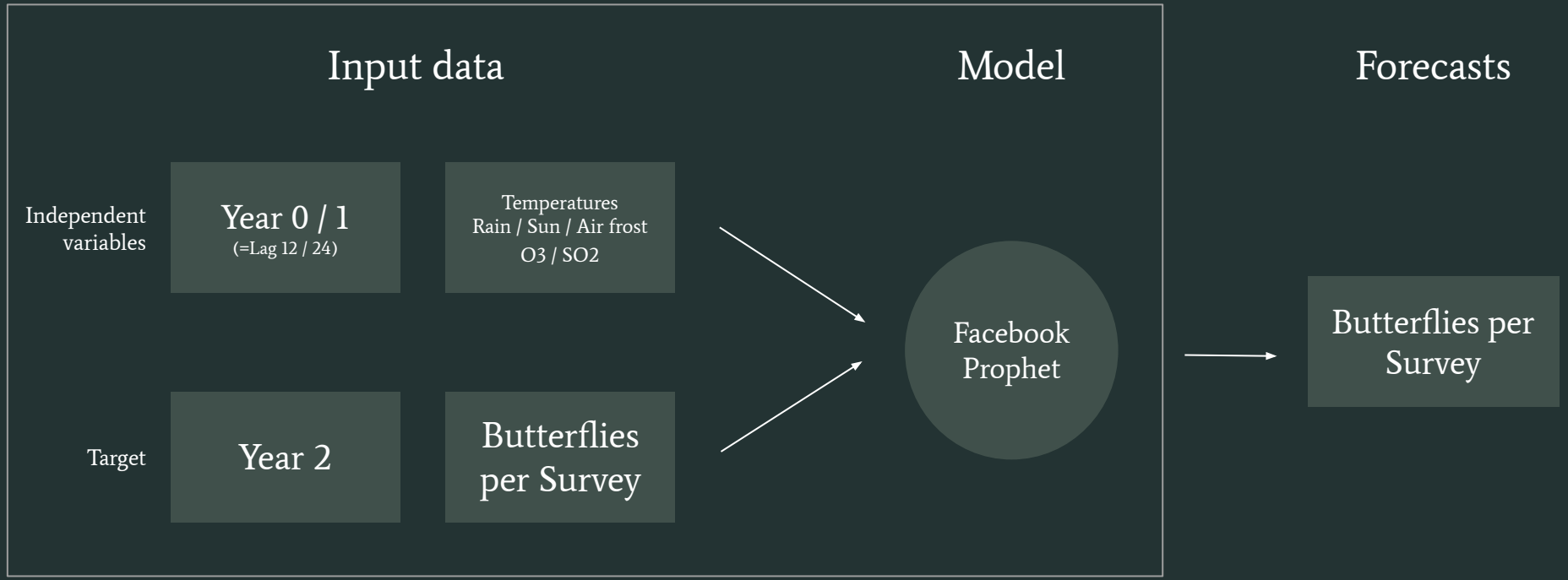


Several variables impact the predictions

> Butterflies per Survey + weather / air quality data

VAR / Prophet

Best data / model combination



A multivariate Time Series model aimed to predict the Butterfly Population Evolution amid climate change (a focus around London, UK) - using Facebook Prophet model

The Butterfly Effect

Model filters

Temperature variation

1%

−30% 30%

O3 variation

1%

−30% 30%

SO2 variation

1%

−30% 30%

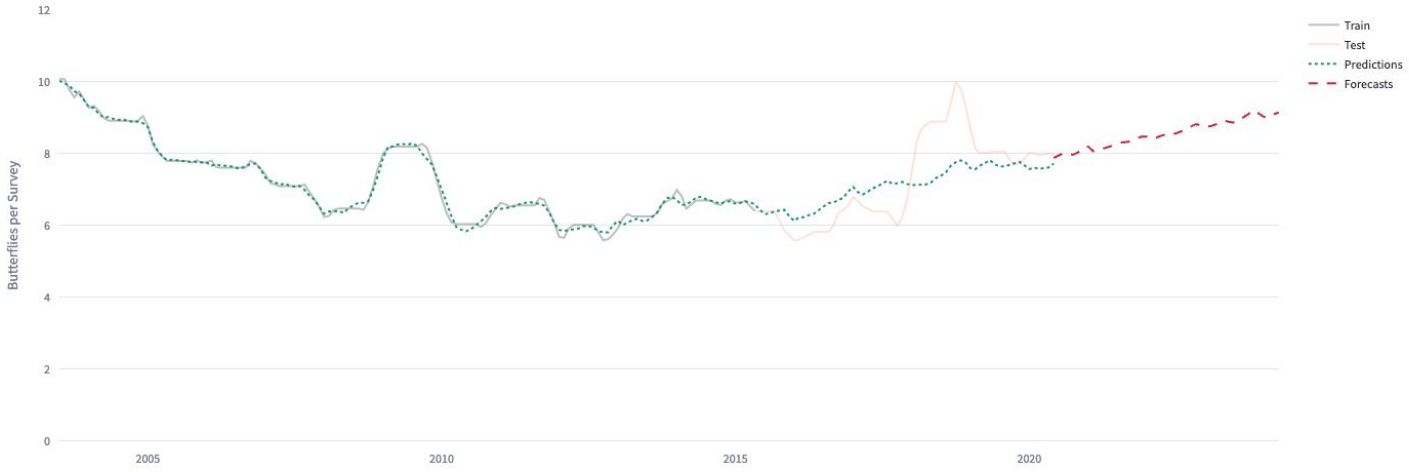
Data Sources

Butterfly sightings data: UKBMS

Weather data: Met Office

Air Quality data: London Air

Butterflies population evolution estimation trend in London, UK



> butterflies.streamlit.app

Next steps

> Extend to more locations

> Break down per species or butterflies group
(using specific traits, for example: wingspan or number of host plants)

Thank you!

chloe.m.cousin@gmail.com

linkedin.com/in/chloe-m-cousin/