Photo by Arun Clarke on Unsplash

The butterfly effect

A Case for Biodiversity using Machine Learning

Capstone Project - Sprint 3 - Apr 15, 2024 - Chloe Cousin

What are we looking at?

6,392,186
butterfly sightings

2001 - 2020

United Kingdom

From: UK Butterfly Monitoring Scheme

Butterfly icon - Alex Muravev | Calendar icon: Iulia Ardeleanu | UK icon: Dolly Holmes

6,392,186
butterfly sightings

2001 - 2020

United Kingdom

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

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**

How?

> An **interactive tool** designed to **facilitate a deeper understanding** of the relationship between our **evolving environment** and its **impact on biodiversity**.

Indicator for Butterfly population

Butterfly sightings
(total of 6,392,186)

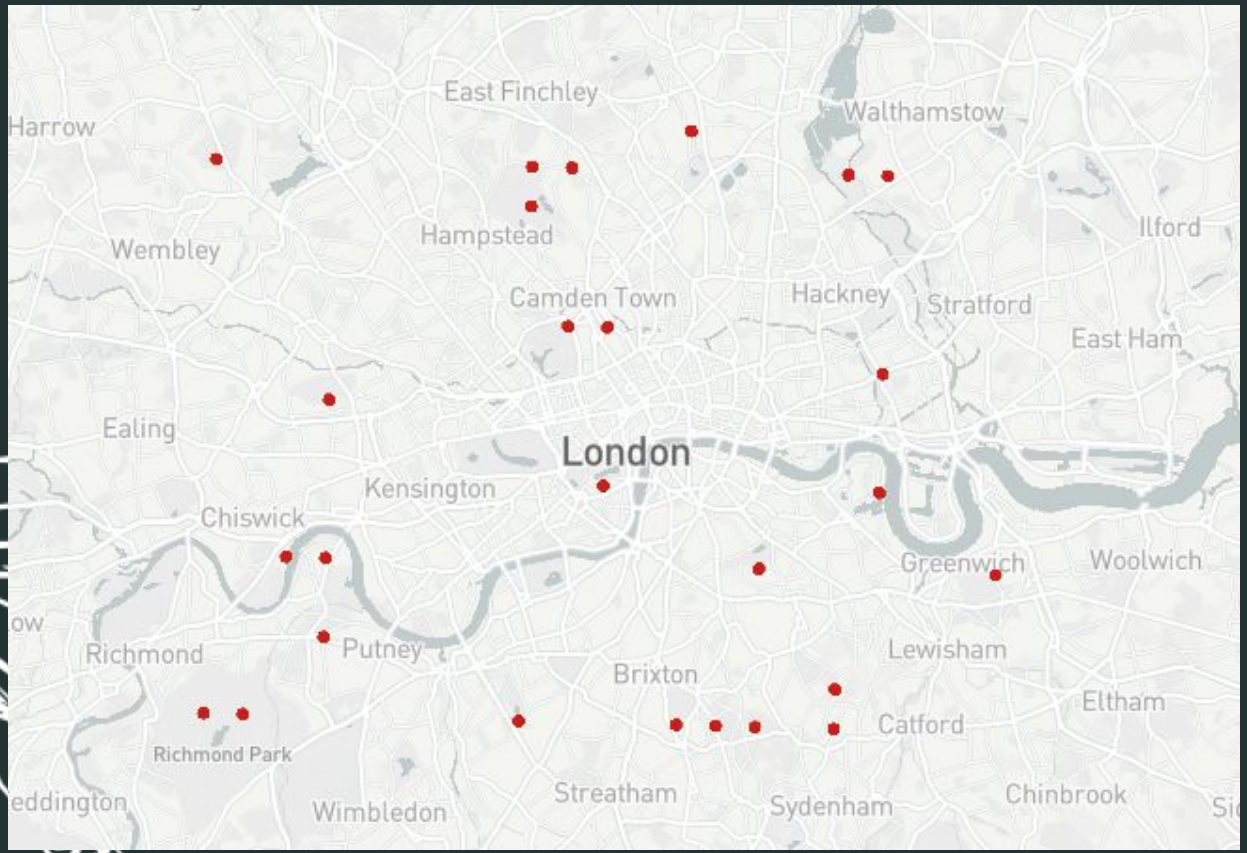
Surveys
(total of 381,276)

>

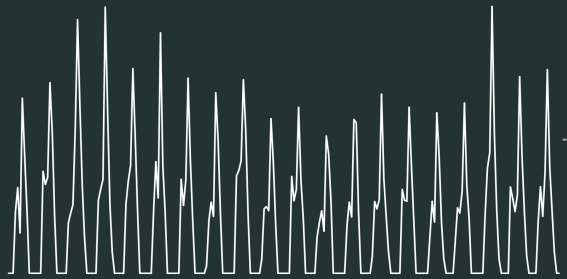
Butterflies seen per Survey
(average of 16.8)

Walking icon: Katrin Scholz

Focus on London Sightings



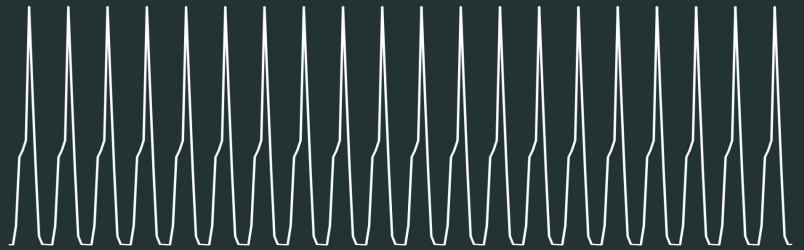
All data



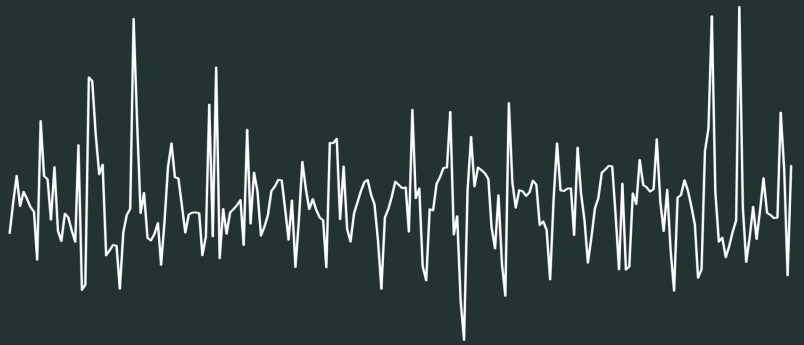
2001 2020

Trend

Seasonality



Residuals



Models

Complexities of Predicting Biodiversity Population Evolution

COLLECTING DATA

- Counts: Warm weather leads to some butterfly species being past their peak in numbers by the time the surveys starts each year

- Population estimation: impossible to account for all butterflies

- Human observation: the methods for collecting data are very thorough at the UK Butterfly Monitoring Scheme, but approximations due to human action are still possible

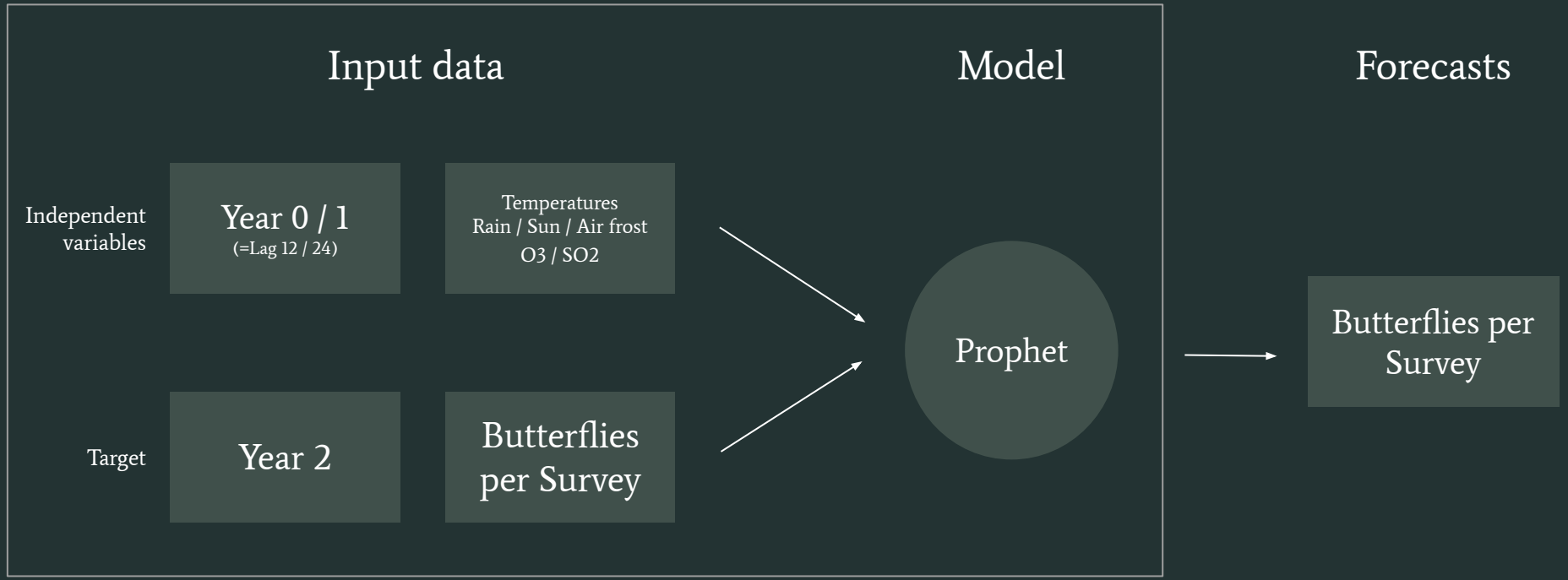
UNDERSTANDING GLOBAL DYNAMIC FROM DATA

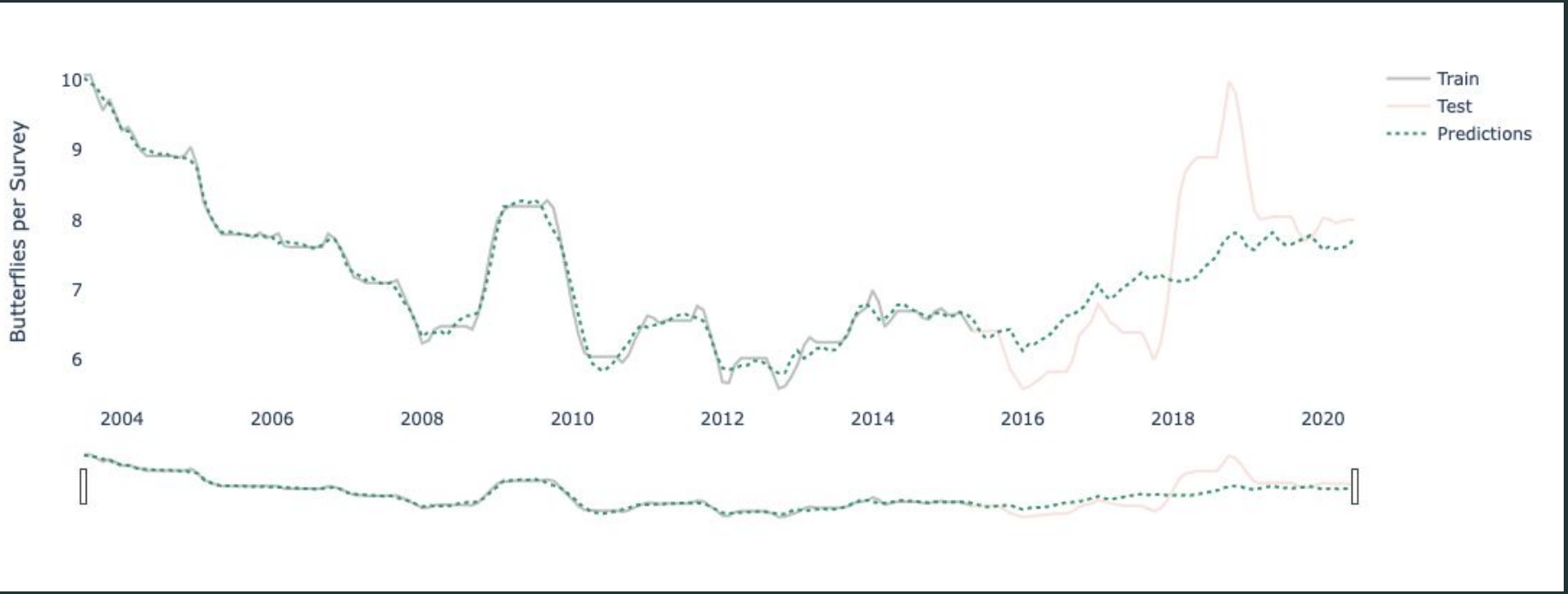
- External factors: not only weather and air quality factors, but also conservation efforts, population migration, plants density...

Models results

	Univariate models		Multivariate models	
	ARIMA	Prophet	Prophet	Prophet
Input data	-	-	All variables Lags = 12	Some variables Lags = 12
MAPE train	1.7%	5.2%	1.3%	1.2%
MAPE test	24.4%	11.4%	19.9%	8.8%
Delta MAPE train/test	22.7%	6.2%	18.6%	7.6%
Notes		Predictions line somewhat “flat”: no variation in the last 8 years		

Best data / model combination





MAPE Train set: 1.24%

MAPE Test set: 8.81%

Forecasts

Model filters

Temperature variation

2%

−30% 30%

O3 variation

27%

−30% 30%

SO2 variation

−19%

−30% 30%

Data Sources

Butterfly sightings data: UKBMS

Weather data: Met Office

Air Quality data: London Air

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

Butterflies population evolution estimation trend in London, UK

Train set MAPE: 1.23% | Test set MAPE: 8.93% | Difference in accuracy between train and test sets: 7.7%

> Streamlit app

Thanks!