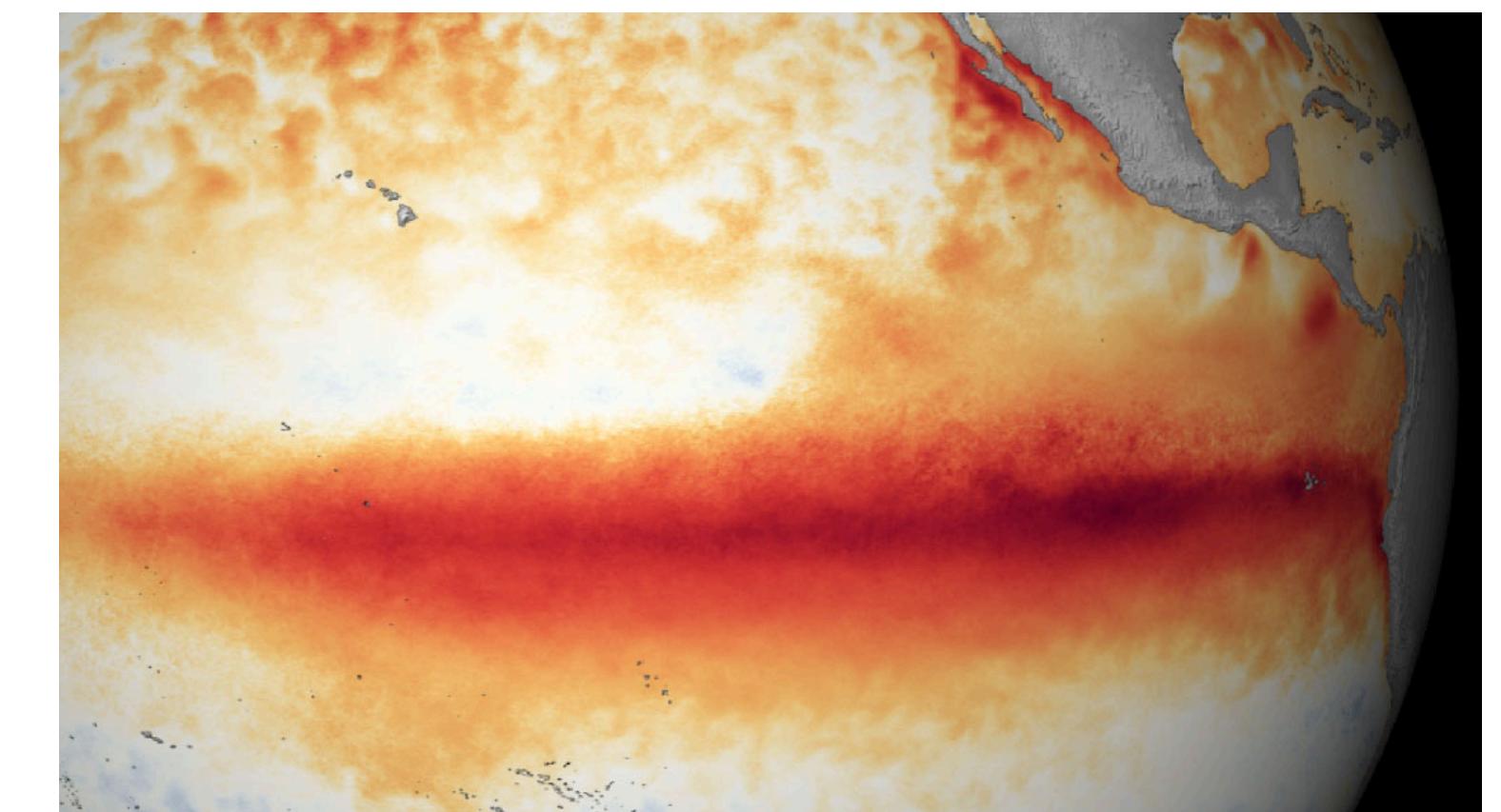
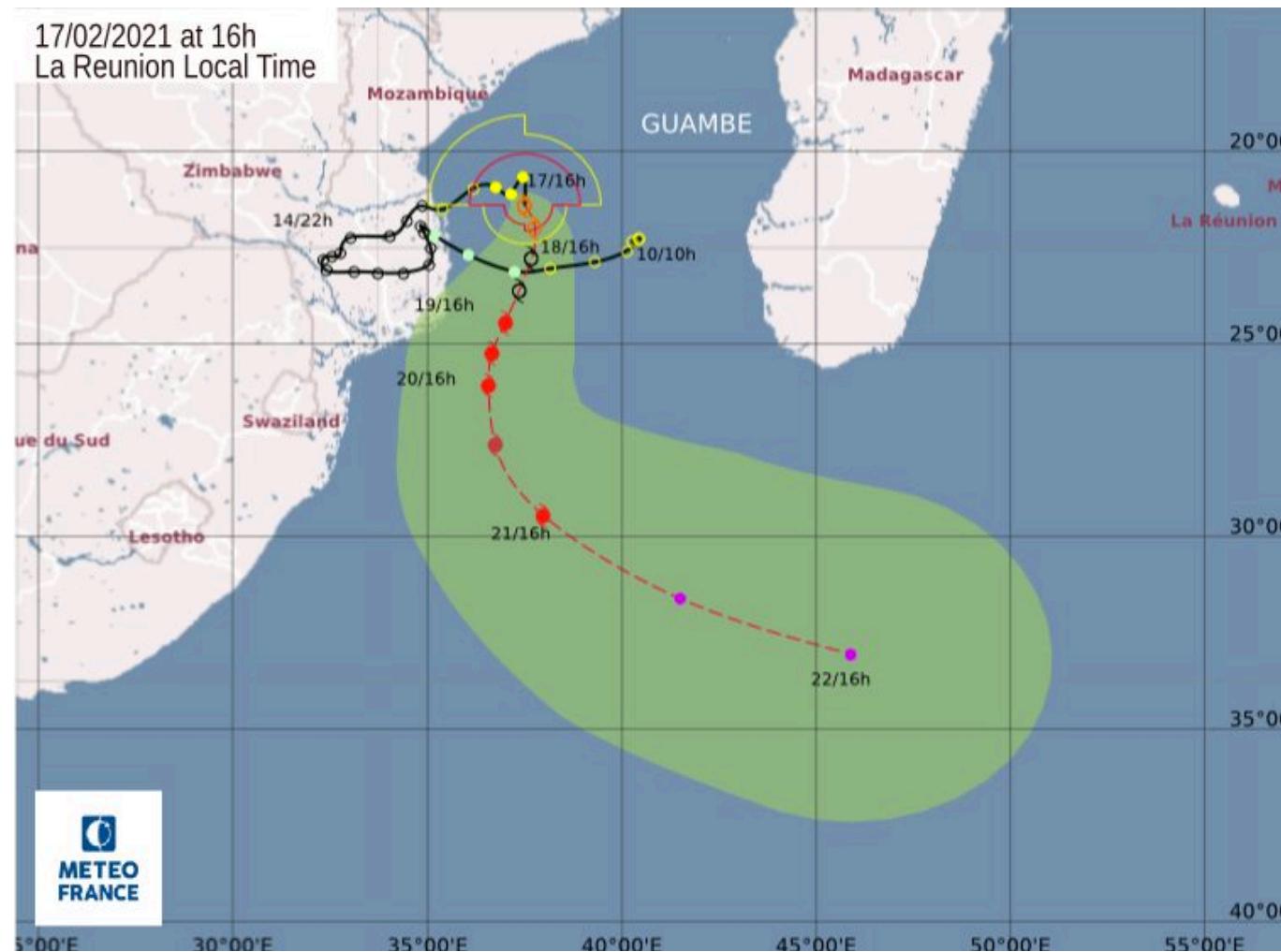


Challenges 1+2+3

Machine Learning with Weather and Climate Data

So Takao (UCL) and Bertrand Chapron (IFREMER)

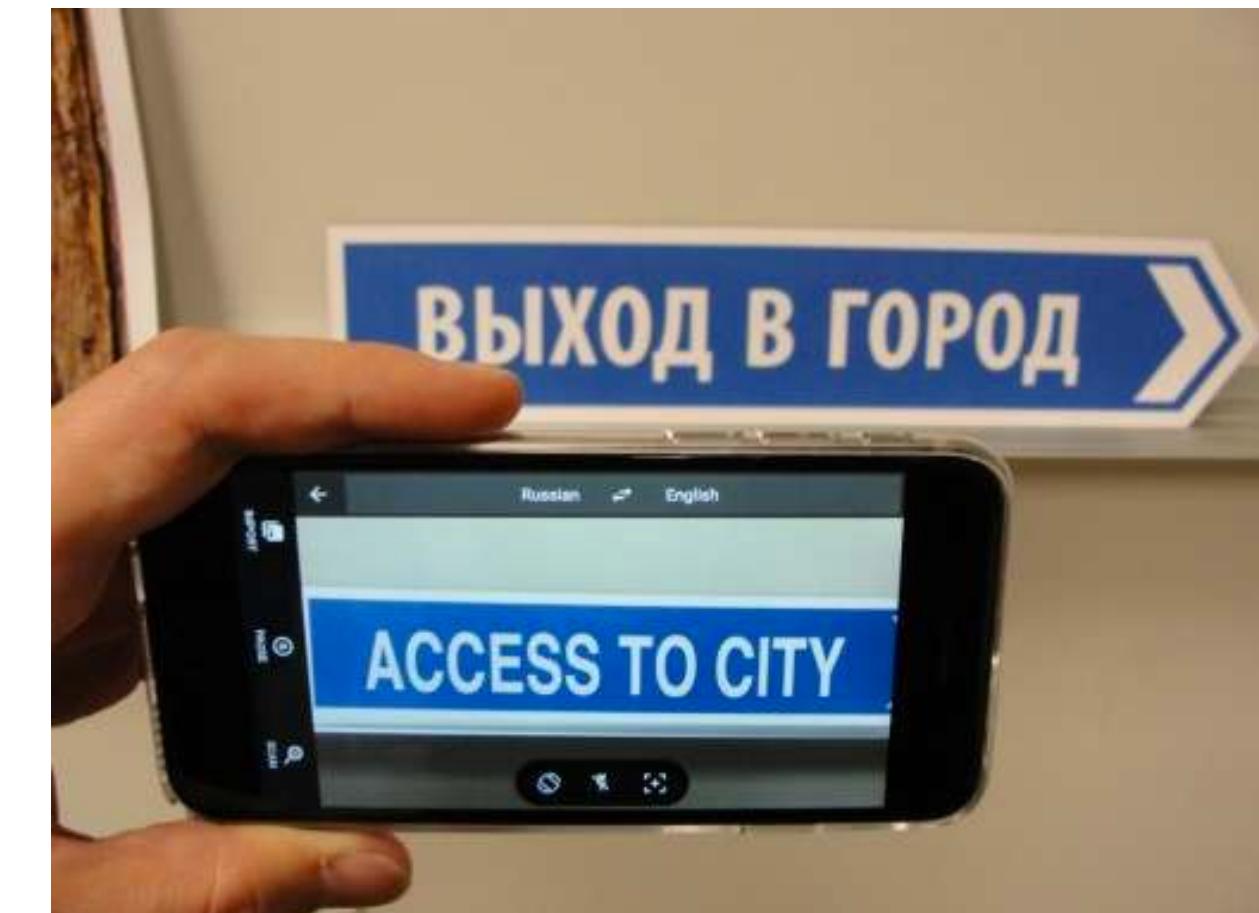
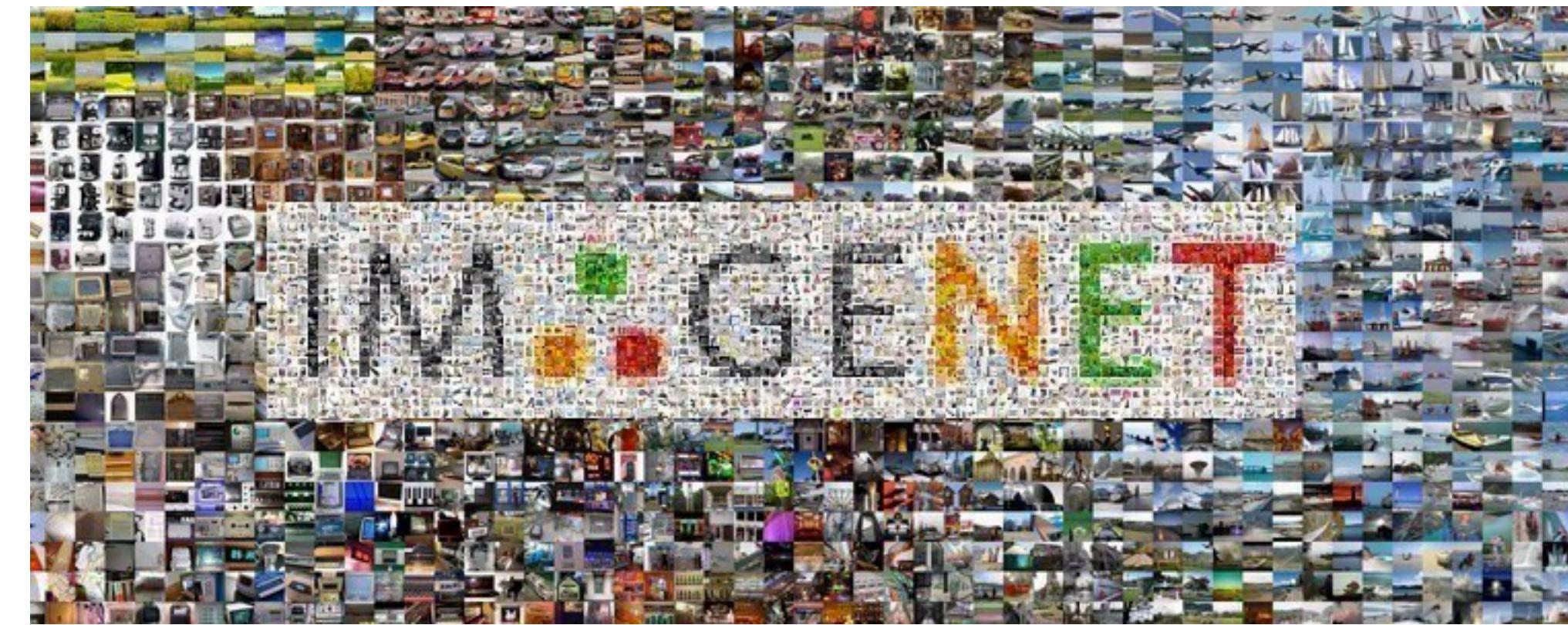
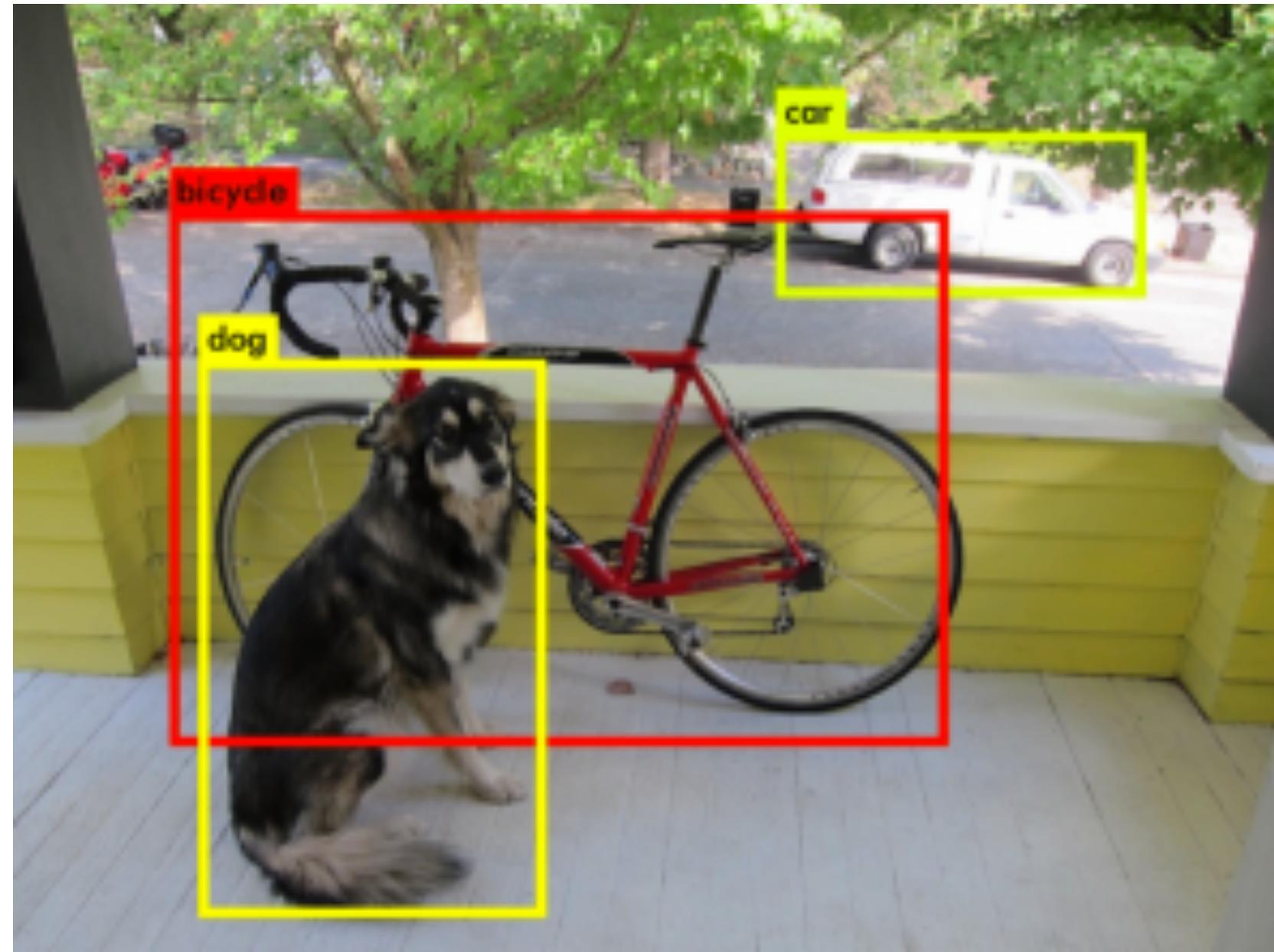
Conventional numerical methods don't work well on certain tasks...



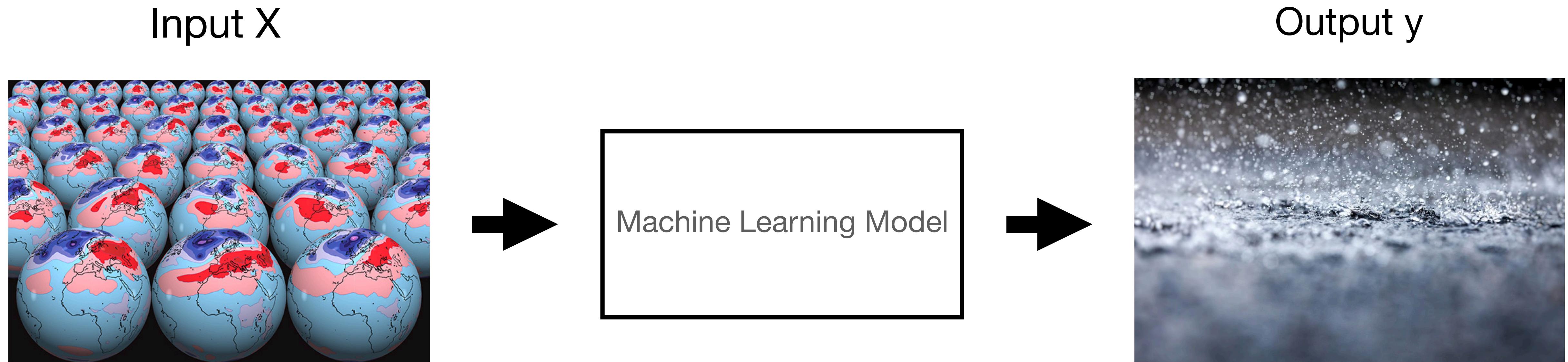
...Can we use Machine Learning to deliver **data-driven** solutions to these problems?

What is machine learning?

Automatic extraction of patterns from large amounts of data



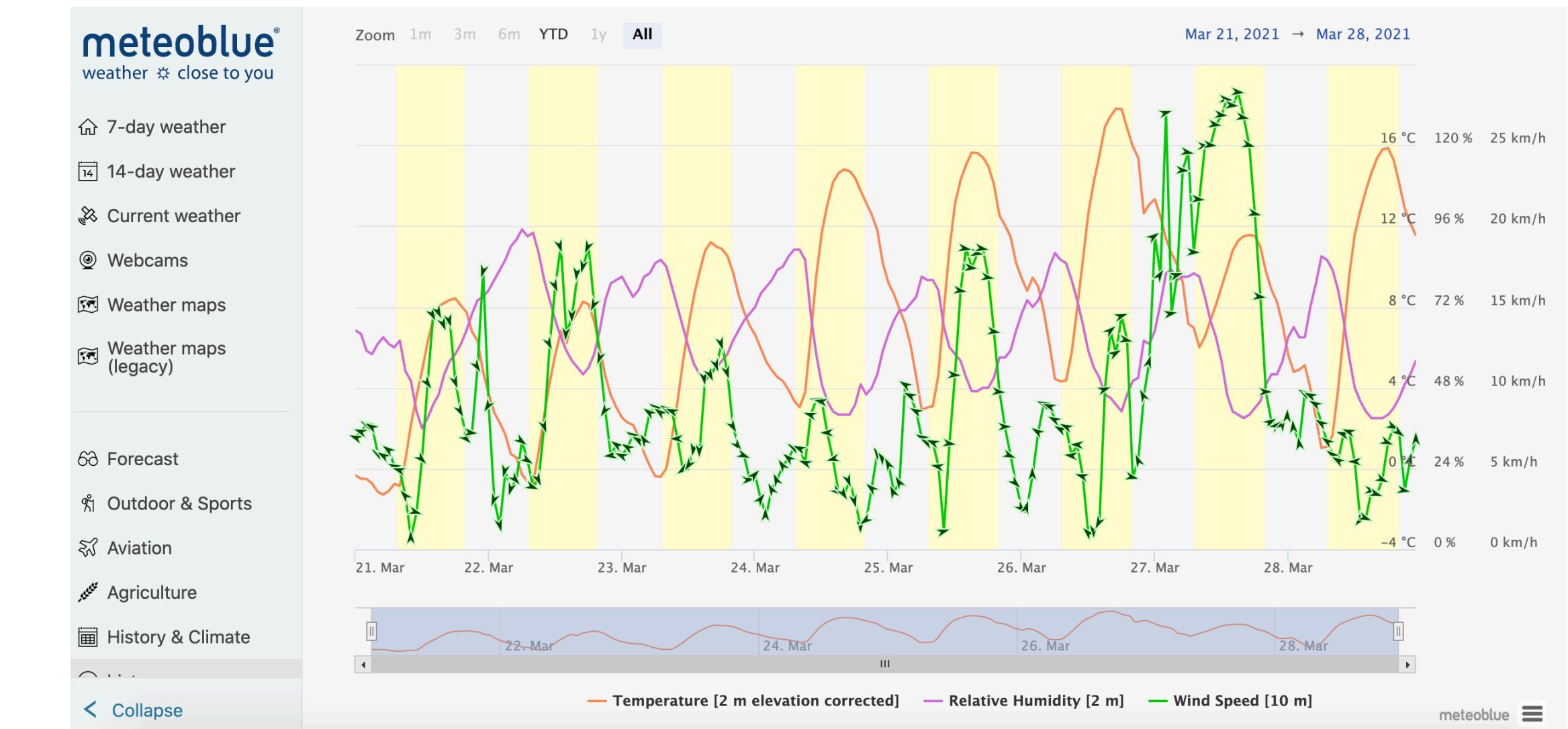
How do we apply this to weather/climate science?



Challenge 1: Predicting the rainfall in Basel, Switzerland

- Data from meteoblue.com

Category of rainfall	Intensity (mm)
Trace	$\leq 3.0 \text{ mm}$
Light rain	= 4.57- 9.64 mm
Moderate rain	= 9.65- 22.34 mm
Moderately heavy rain	= 22.35- 44.19 mm
Heavy rain	= 44.20- 88.90 mm
Very heavy rain	$\geq 89 \text{ mm}$

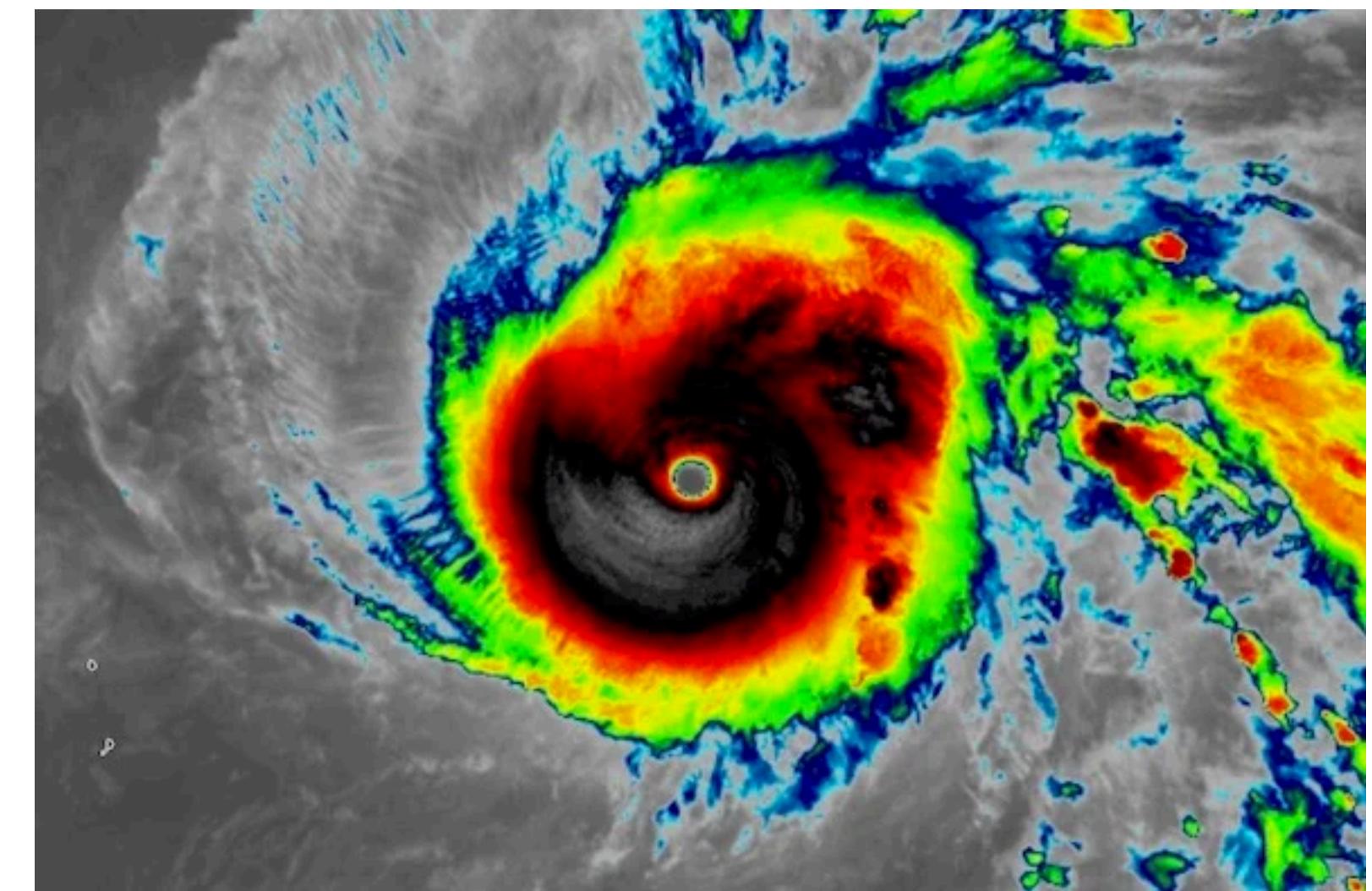


- **Task:** Can you predict the rainfall during the period 03/2020-03/2021 using other variables?
- **Training data:** 03/2010-03/2020

- **Techniques:** Binary/multi-class Classification, Regression, Time-series modelling, etc...
- Beginner-friendly

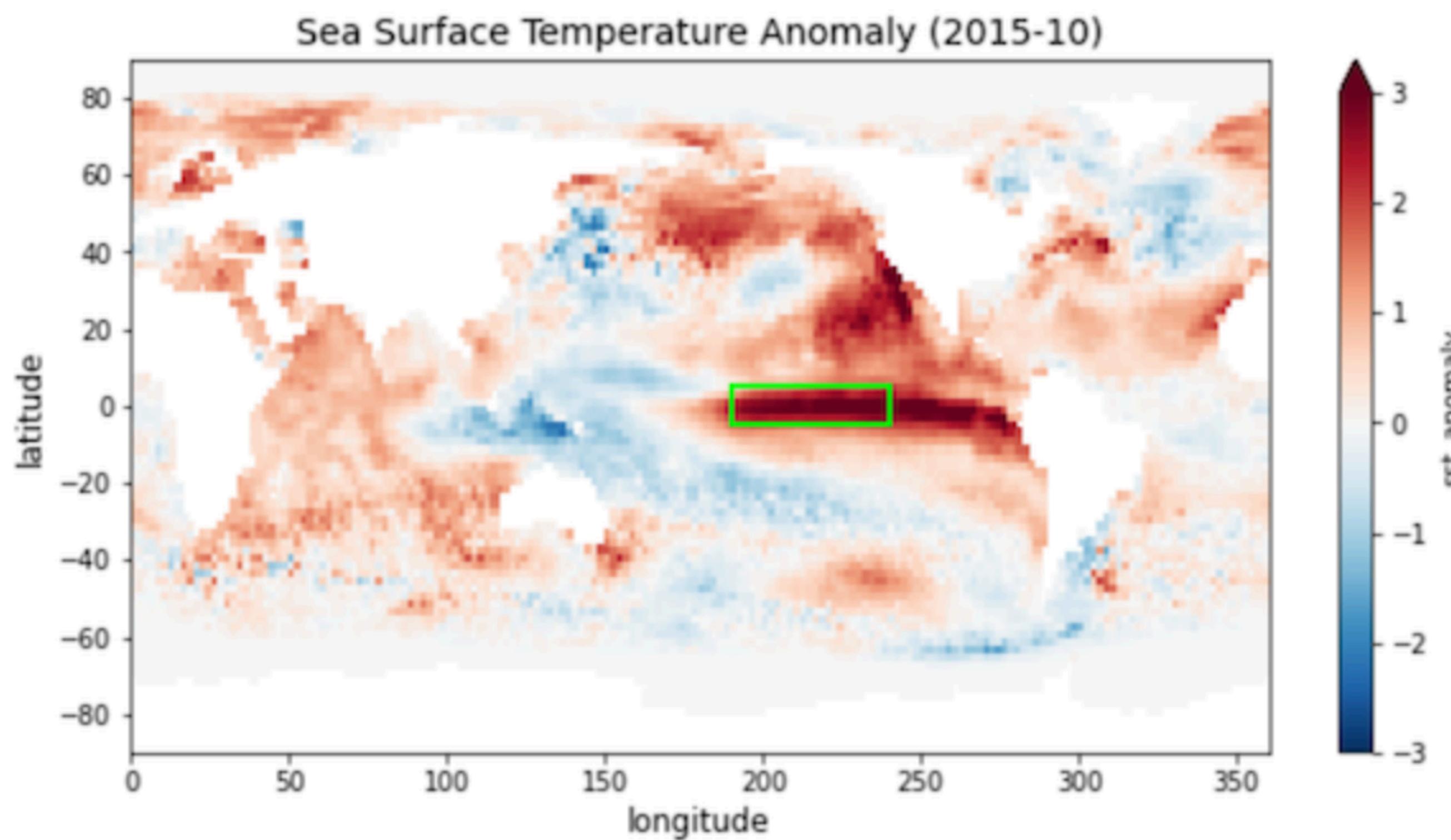
Challenge 2: Predicting the trajectory and intensity of tropical cyclones

- Storm data from US NRL and NOAA
- **Tasks:**
 - Can you predict the highest intensity a cyclone will reach in a given region?
 - How far in advance can you predict this?
 - How many cyclones will there be?
 - How many will reach categories 4 and 5?
- **Techniques:** Regression, classification



Challenge 3: Predicting the occurrence of El-Nino events

El Nino ([Definition](#)): An event where the 5 months running average of the sea surface temperature anomaly in the Nino 3.4 region $5^{\circ}\text{S} \sim 5^{\circ}\text{N}$, $120^{\circ} \sim 170^{\circ}\text{W}$ (equivalently, $-5^{\circ} \sim 5^{\circ}\text{N}$, $190^{\circ} \sim 240^{\circ}\text{E}$) exceeds 0.4°C for periods of 6 months or longer.



Task: Can you predict the occurrence of El Nino events 1-6 months in advance?

Data from various sources available from 1981-2019

Techniques: Timeseries modelling, CNNs

Getting started

- All information about the challenges and datasets are available on Github:

[https://github.com/sotakao/STUOD Hackathon 2021](https://github.com/sotakao/STUOD_Hackathon_2021)

- Preferred methods of communication:

- Zoom
- Slack

(Links can be found in the Hackathon information booklet)

- Who to contact:

Me (So Takao), Bertrand Chapron (aka Canard Duck), Jean-Francois Piolle, Ronan Fablet, Nicolas Reul (for Tropical Cyclones challenge)

- You can use other datasets or even formulate your own problems as long as it's weather/climate related
- Please report any issues/bugs in Github or Slack
- Interactive tutorials available!

- Classification with L63
- Regression with L96
- Handling data with pandas
- Getting started with xarray

The screenshot shows a Jupyter Notebook interface with the title "Part1-Classification". The top bar includes a Python logo, "Logout", "Trusted", and a dropdown menu for "tensorflow_env". The main area features a diagram with a blue box labeled "Regression" and an orange box labeled "Dimensionality reduction". A vertical label "Continuous" is positioned to the left of the boxes. Below the diagram, text explains supervised learning tasks: "In this tutorial, we'll only be concerned with supervised learning tasks, so our data will be *labelled*, which means that it consists of a pair $(X_n, y_n)_{n=1}^N$ of inputs X_n and labels y_n ." An example is given: "Example: The inputs X_n are images of chihuahuas and muffins, and the labels y_n are the tags 'chihuahua' or 'muffin' (yes you read that right)."

Diagram:

```

graph LR
    A[Continuous] --- B[Regression]
    A --- C[Dimensionality reduction]
  
```

Text:

In this tutorial, we'll only be concerned with supervised learning tasks, so our data will be *labelled*, which means that it consists of a pair $(X_n, y_n)_{n=1}^N$ of inputs X_n and labels y_n .

Example: The inputs X_n are images of chihuahuas and muffins, and the labels y_n are the tags 'chihuahua' or 'muffin' (yes you read that right).

Image:

A 3x4 grid of images showing alternating chihuahuas and muffins. To the right of the grid, a large black arrow points from the text "Chihuahua" to the first column of images, and from the text "Muffin" to the second column of images. A small blue letter 'f' is placed above the arrow.

When you present, remember to...

- Include a problem statement
- Indicate what data you chose for training/testing
- What is the baseline that you are comparing against?
- What is the metric that you are using to compare models?
- What are some challenges you faced?
- What next?



Final words

- It's ok if you don't complete, you are here to learn!
- Ask us anything if you are having any difficulties
- Please be respectful of others and work as a team



happy hacking!