# Weather prediction analysis

***George Petropoulos, Thijs Spoor & Skipper van den Brekel***

**Project overview:**

This project focuses on weather prediction using time series analysis techniques. The goal is to analyse historical weather data and predict future trends by using three methodologies:

1. Linear Regression Model (LRM): A statistical approach to understand and model the relationship between variables.
2. Vector Autoregression (VAR): A multivariate model to capture the relationship between multiple time series.
3. Fourier Transform (FT): A mathematical method to decompose time series data into frequencies.

By using these techniques separately and combining them, we want to improve the accuracy of weather predictions and explore seasonal patterns and trend in the weather data. The three hypothesis we’d like to test are:

1. Vector Autoregression does not improve the Linear Regression Model in weather prediction.
2. Variables exhibiting seasonal periodicity do not have statistically lower forecasting errors than variables without such patterns.
3. Increasing the sample size does not have an effect on the predicting power of a Vector Autoregression.

**Prerequisites:**

Make sure the following tools and libraries are installed:

* Python 3
* Pandas
* NumPy
* Matplotlib
* Scipy
* Math

**Data source:**

<https://www.kaggle.com/datasets/thedevastator/weather-prediction>

The cleaned up data can either be downloaded from the git repository or created with the original dataset using cleanup weather data.

**Repository overview:**

* Archive: unused or already rewritten code
* Data: original dataset + cleaned up data csv files
* Figures: plots mostly used during the presentation
* LRM\_Data: LRM prediction data used in hypothesis testing
* Presentation: contains the presentation
* Testing\_Figures: contains the results of the hypothesis tests
* cleanup\_weather\_data: Separates the data per city + add python datetime
* weather\_visualisation: initial visualisation and correlation of data
* weather\_var: our VAR model and initial fitting/testing
* weather\_fourier\_prediction: our Fourier analysis and failed attempt at prediction
* lrm\_prediction: our LRM functions and produces lrm presentation figures
* fourier\_plot: creates the figures used for fourier explanation
* correlation: creates the correlation figure used in VAR explanation
* Hypothesis\_x: the code used in hypothesis testing / testing figures creation
* LRM\_hypothesis\_2: creates the LRM\_data csv-files

**Hypothesis Results:**

H0: Vector Autoregression does not improve the Linear Regression Model in weather prediction. (Hypothesis\_1.ipynb), (Hypothesis\_1\_DM\_results)

Test: Diebold-Mariano test

Fail to reject: pressure, temp\_mean, and temp\_min

Reject: humidity, sunshine, global\_radiation

H0: Seasonality does not make a variable more predictable in the VAR framework (in terms of lower RMSE). (Hypothesis\_2.ipynb), (H2\_trainings, H2\_boxplots)

Test: Welsch’s Two Sample t-test

Fail to reject: Non-seasonality grouped VAR systems had lower RMSE values than seasonality grouped VAR systems.

H0: Increasing the sample size does not have an effect on the predicting power of a Vector Autoregression. (Hypothesis\_3.ipynb), (H3\_DM\_Test\_Heatmap)

Test: Diebold-Mariano test

Reject: Humidity, Pressure, Global Radiation

Fail to Reject: Temp\_mean, Temp\_min, Temp\_max