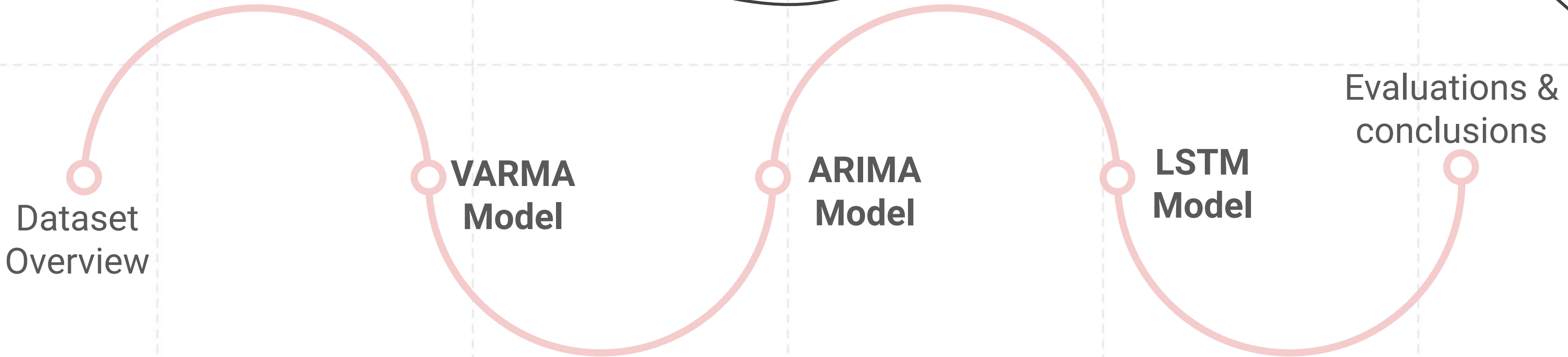
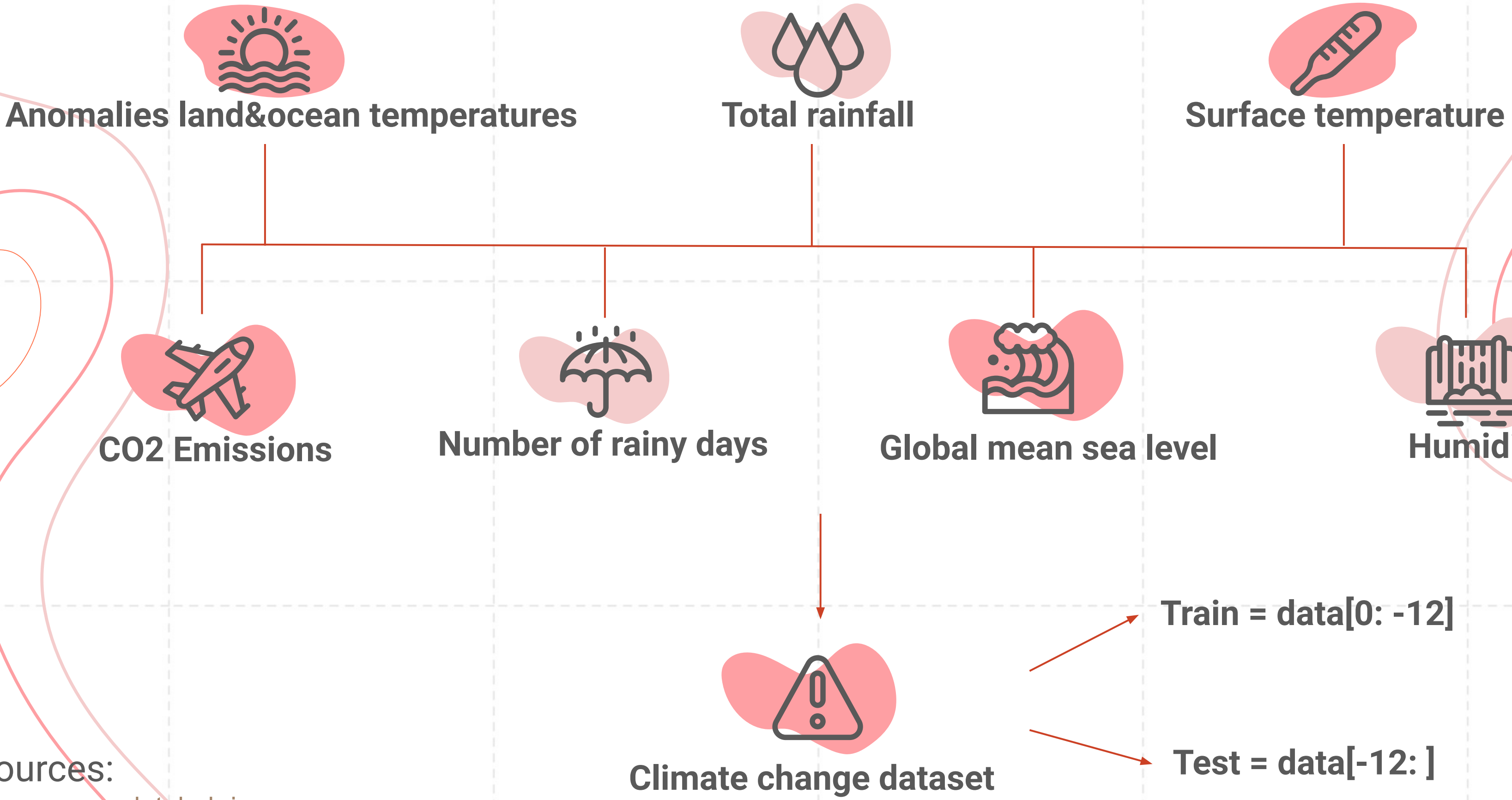
Climate change time series analysis and forecasting

Stefania Sferragatta
Matricola 1948081

Big Data for Official Statistics
AY 2021-2022

Project flow

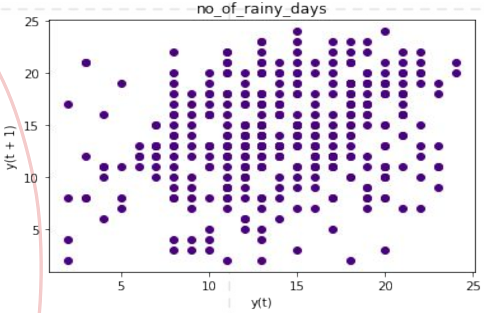


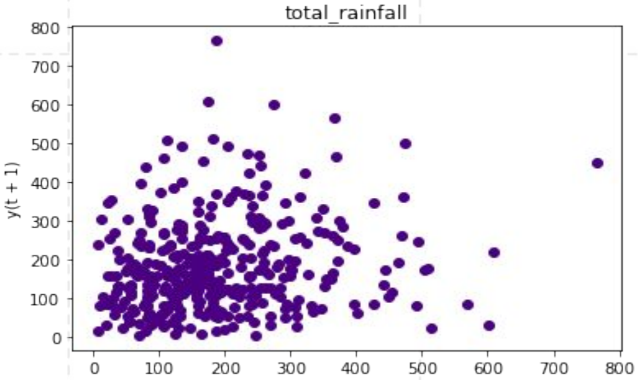


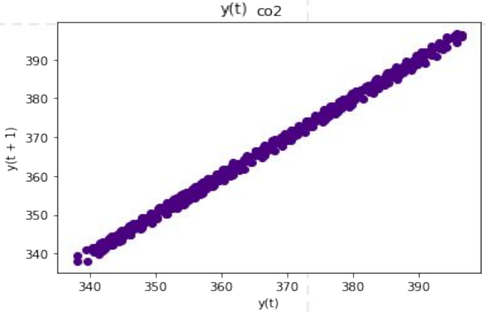
Sources:

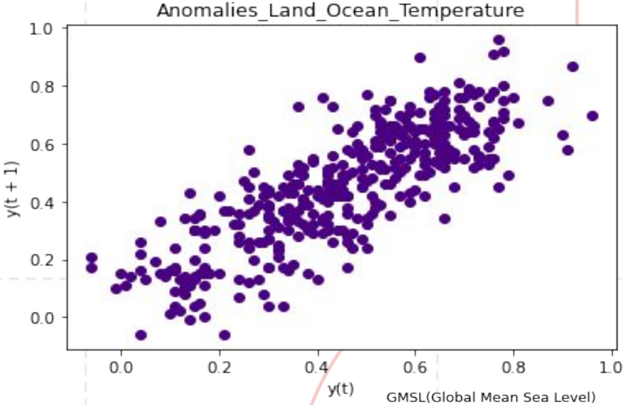
- www.datahub.io
- www.data.gov.sg

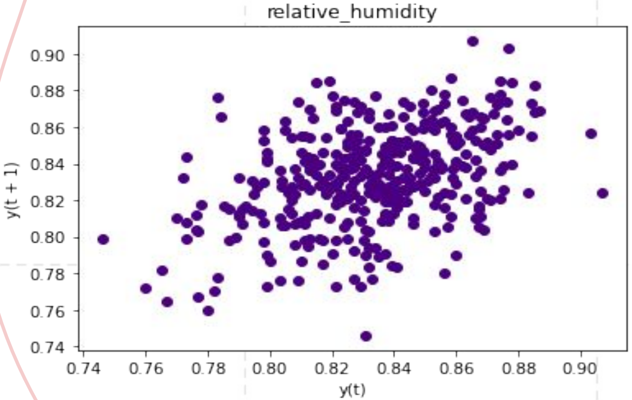
Lag scatter plots

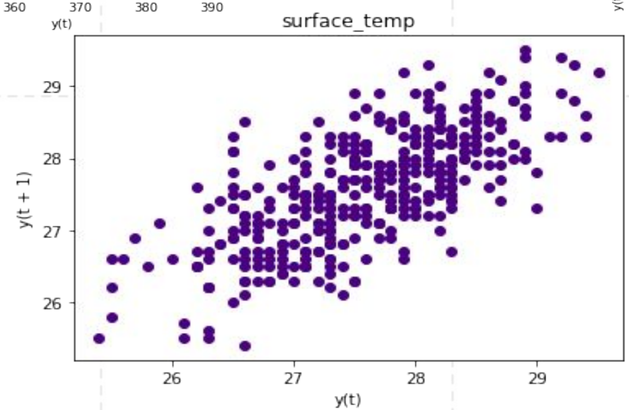


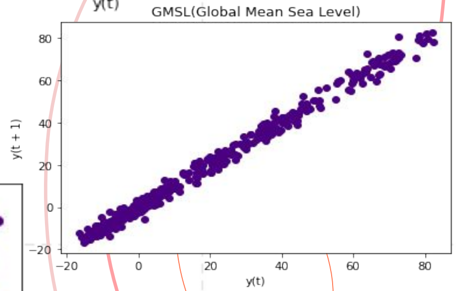




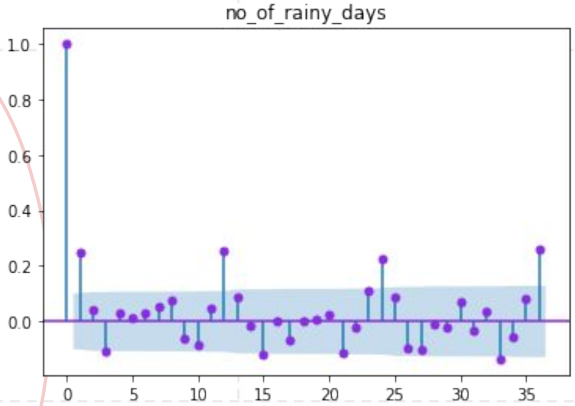


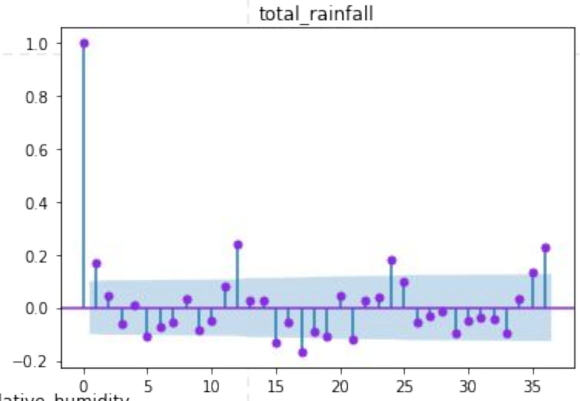


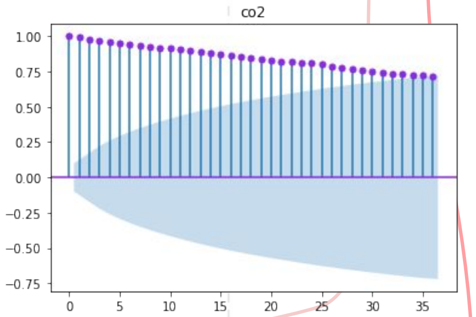


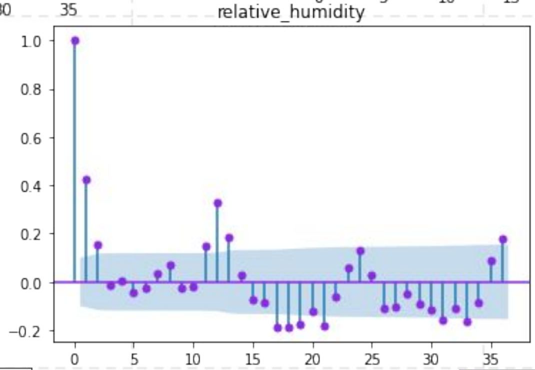


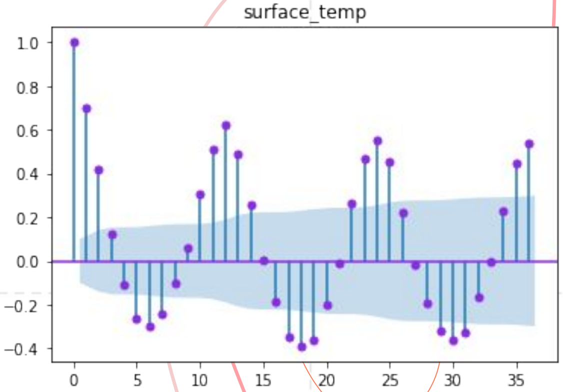
Autocorrelation

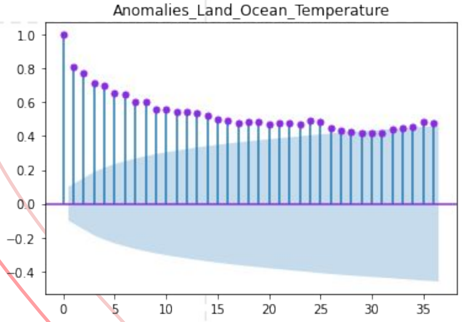


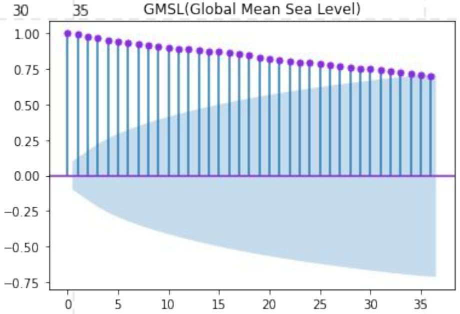




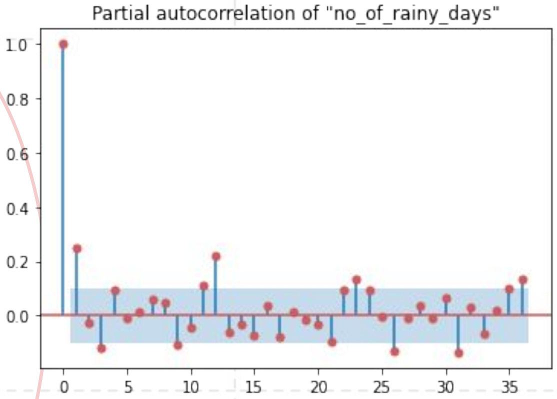


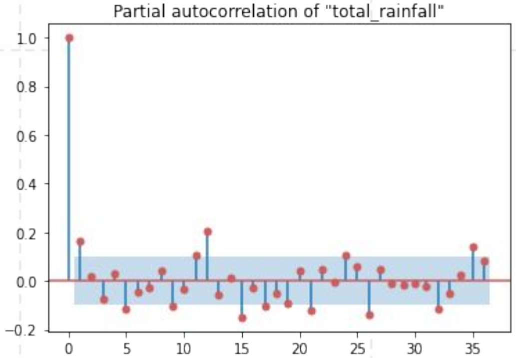


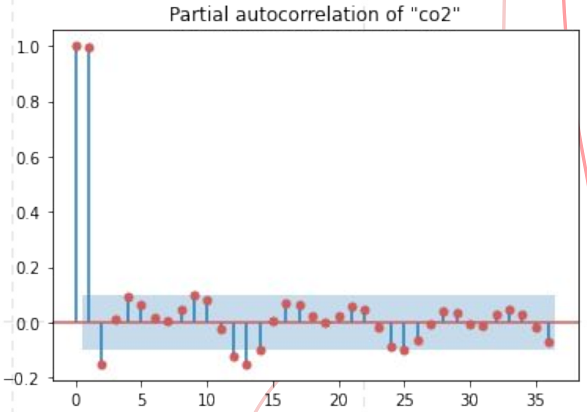


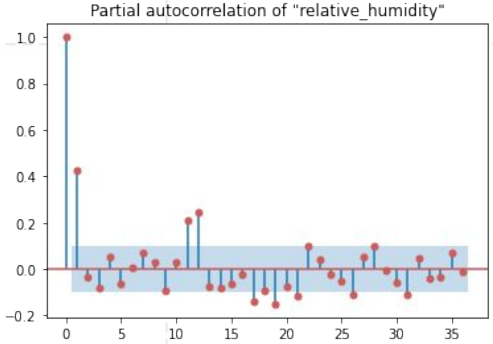


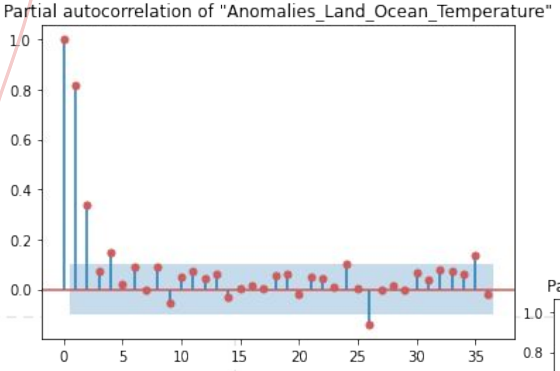
Partial autocorrelation

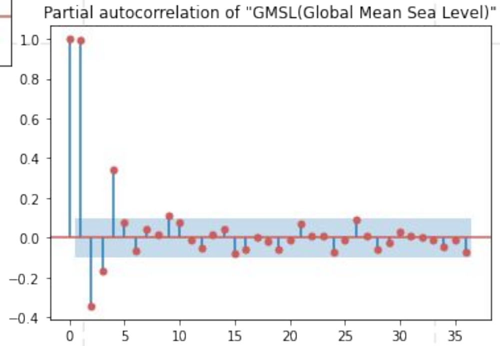


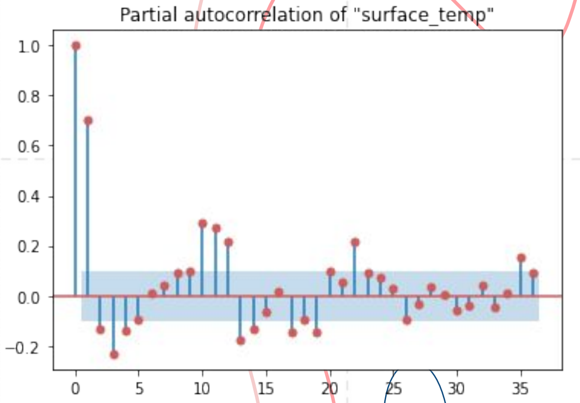








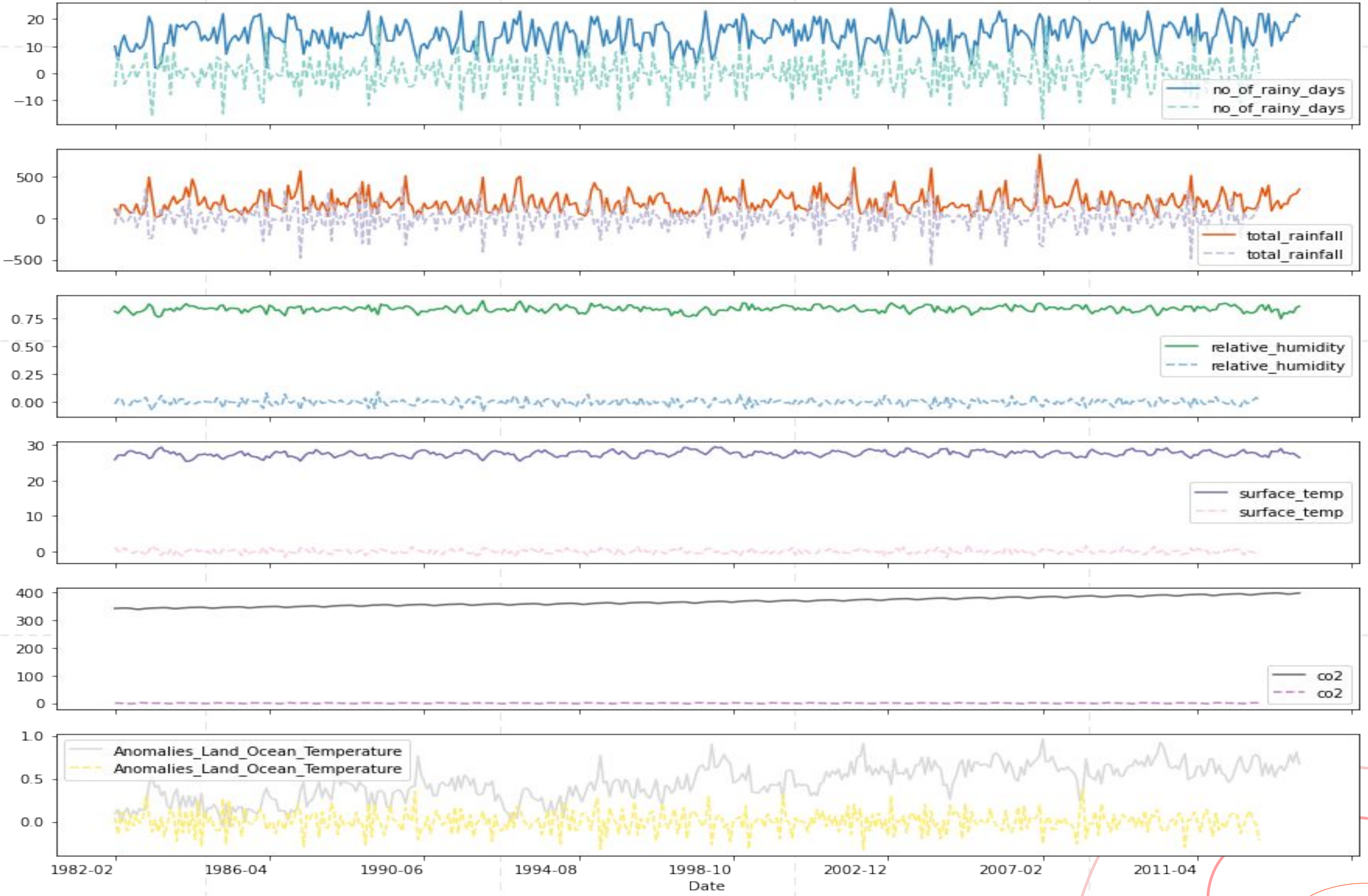




Steps of the VARMA model

1. Causality investigation using **Granger Causality** test
2. Test for stationarity using **Adfuller** test
3. Test for autocorrelation using **Durbin-Watson statistics** test
4. Fit the model and predict

STATIONARITY PLOT



VARMA MODEL FIT

```
model = VAR(data_differenced)
x = model.select_order(maxlags=20)
x.summary()
```

```
# train the VAR model of selected order (p=12)
p=12
model_fitted = model.fit(p)
model_fitted.summary()
```

Result:

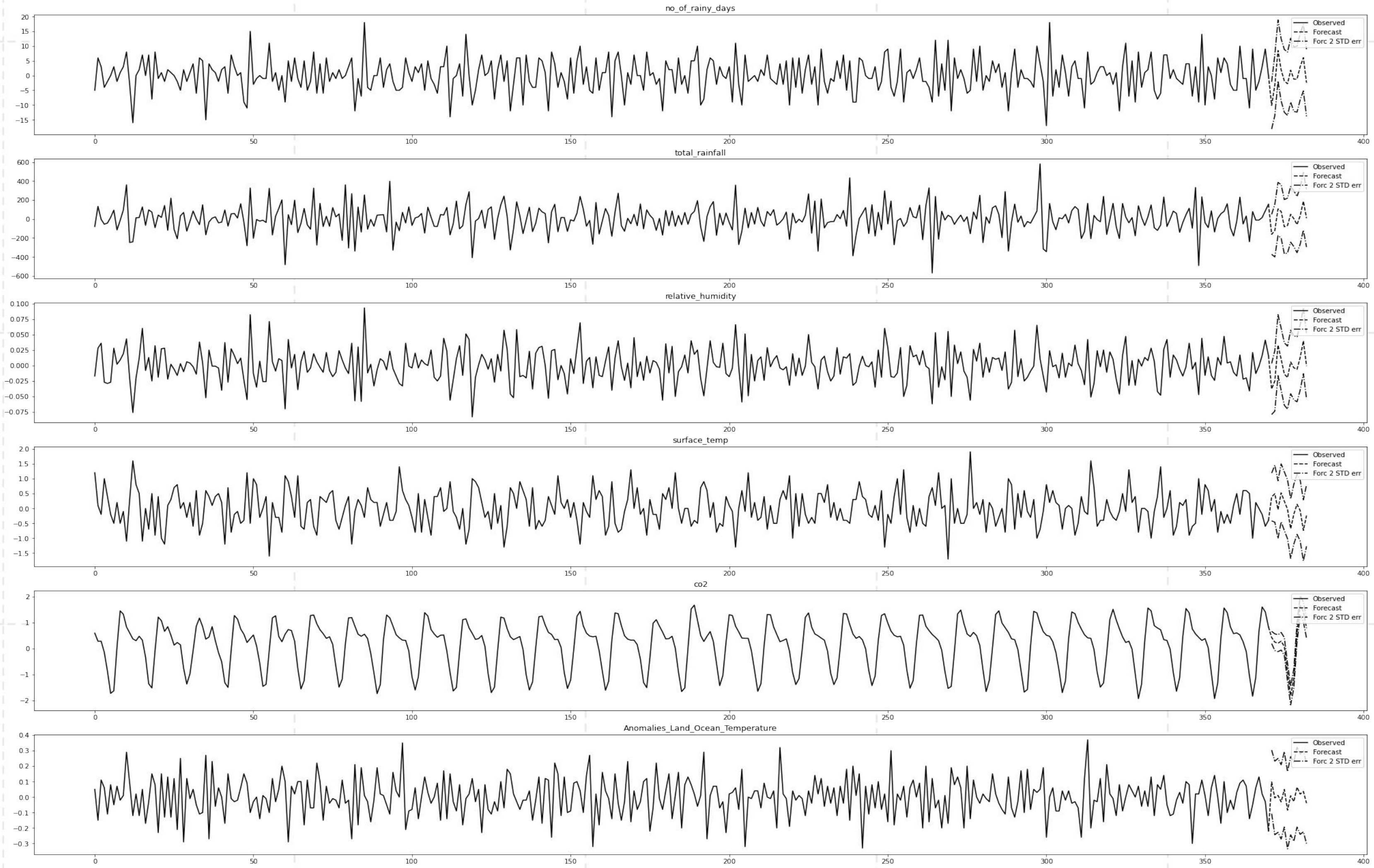
- Each table is related to one feature treating it as the variable of interest;
- In each table is used as possible regressor every lag from 1 to 12 for all the 7 features, and we should consider the regressor with the smallest p-value;
- In the end there is a confusion matrix that shows the correlation between the variables.

PREDICTION

```
lag = model_fitted.k_ar #12

# input data for forecasting
lagged_Values = data_differenced.values[-lag:] # specify the 'initial value' for the forecast

# Forecast
pred = model_fitted.forecast(y= lagged_Values,steps= n_obs)
df_pred = (pd.DataFrame(pred, index=test.index, columns=test.columns + '_pred'))
df_pred
```

```
result = invert_transformation(train, df_pred)
```

FUTURE FORCAST



ARIMA Model

Climate change dataset

70%

30%

(p,q,d)

Number of the lag observations decided according to the PACF plot

Degree of differencing=1

Size of the Moving Average decided according to the ACF plot

LSTM Model with Keras

Transform the dataset to a supervised learning problem

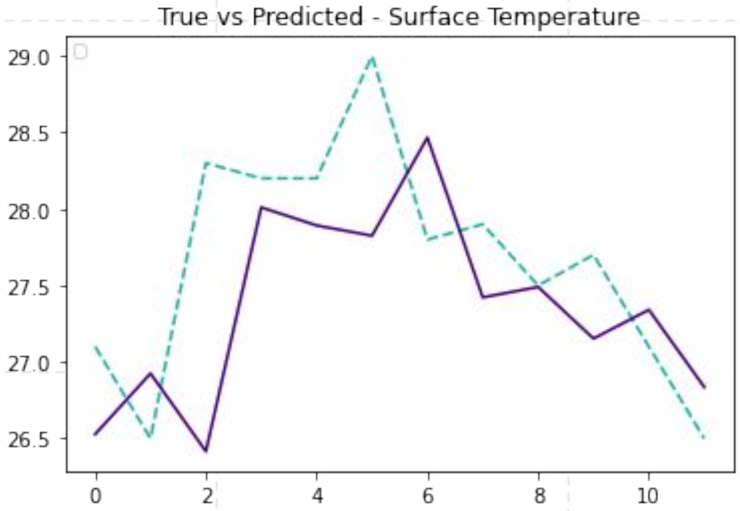
Make the data stationary

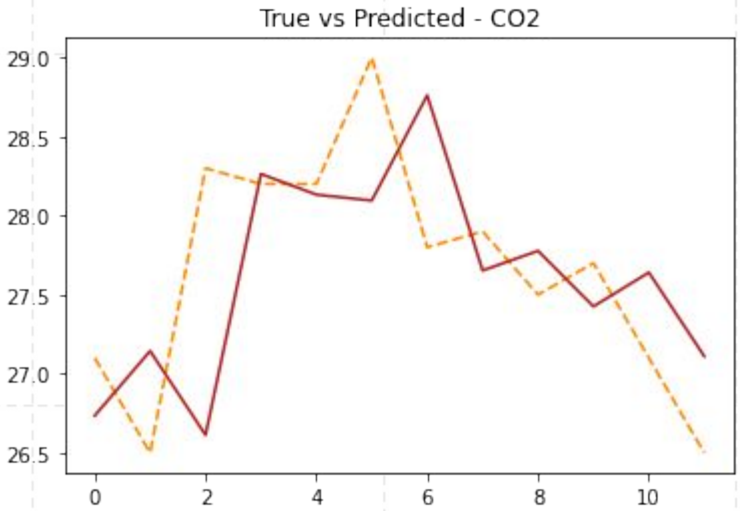
Normalize the data

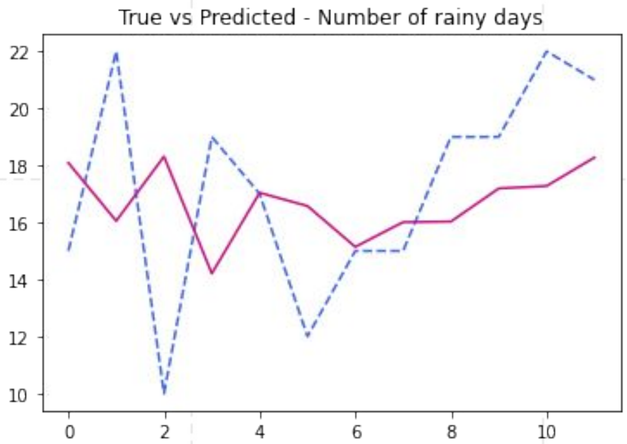
Fit the model

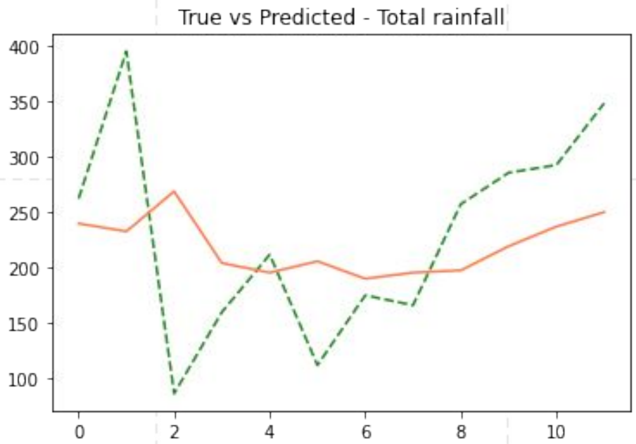
Visualize the predictions

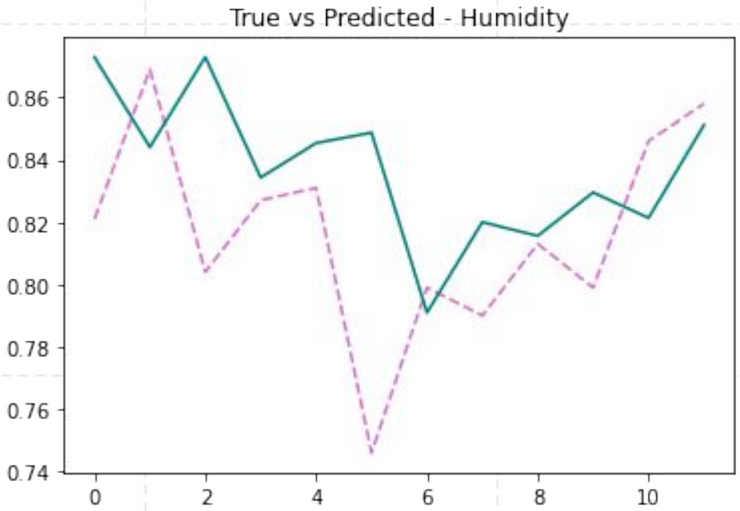
```
repeats = 10
error_scores = list()
for r in range(repeats):
    # fit the model
    lstm_model = fit_lstm(train_scaled, 1, 10, 4)
    # forecast the entire training dataset to build up state for forecasting
    train_reshaped = train_scaled[:, 0].reshape(len(train_scaled), 1, 1)
    lstm_model.predict(train_reshaped, batch_size=1)
    # walk-forward validation on the test data
    predictions = list()
    for i in range(len(test_scaled)):
        # make one-step forecast
        X, y = test_scaled[i, 0:-1], test_scaled[i, -1]
        yhat = forecast_lstm(lstm_model, 1, X) #yhat = y
        # invert scaling
        yhat = invert_scale(scaler, X, yhat)
        # invert differencing
        yhat = inverse_difference(raw_values, yhat, len(test_scaled)+1-i)
        # store forecast
        predictions.append(yhat)

    # report performance
    rmse = sqrt(mean_squared_error(raw_values[-12:], predictions))
    error_scores.append(rmse)
```

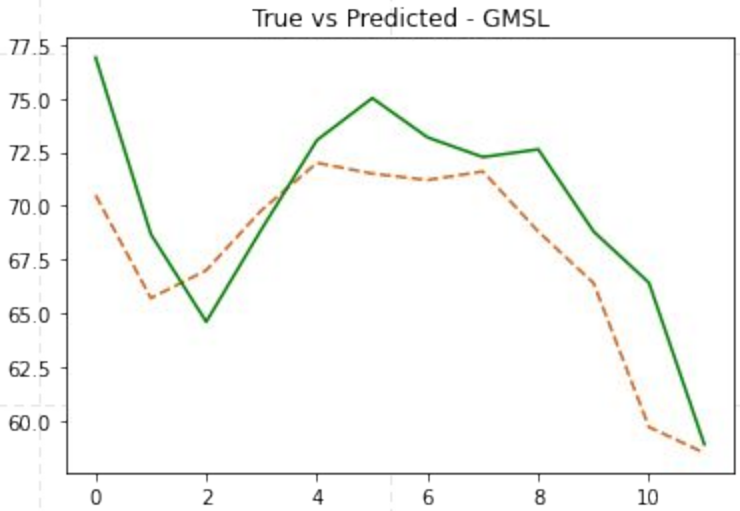

True vs Predicted - Surface Temperature
29.0
28.5
28.0
27.5
27.0
26.5
0
2
4
6
8
10

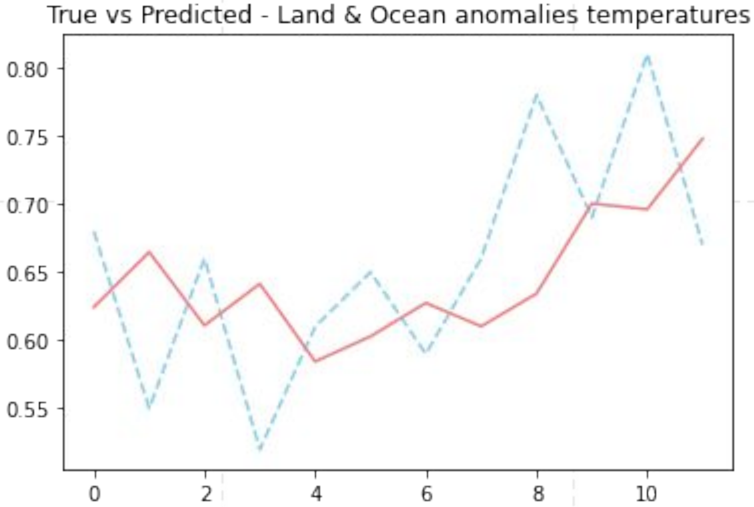

True vs Predicted - CO2
29.0
28.5
28.0
27.5
27.0
26.5
0
2
4
6
8
10


True vs Predicted - Number of rainy days
22
20
18
16
14
12
10
0
2
4
6
8
10


True vs Predicted - Total rainfall
400
350
300
250
200
150
100
0
2
4
6
8
10


True vs Predicted - Humidity
0.86
0.84
0.82
0.80
0.78
0.76
0.74
0
2
4
6
8
10


True vs Predicted - GMSL
77.5
75.0
72.5
70.0
67.5
65.0
62.5
60.0
0
2
4
6
8
10


True vs Predicted - Land & Ocean anomalies temperatures
0.80
0.75
0.70
0.65
0.60
0.55
0
2
4
6
8
10

EVALUATIONS

Number of rainy days	ARIMA	VARMA	LSTM
RMSE	4.132	18.353	4.101

Total rainfall	ARIMA	VARMA	LSTM
RMSE	114.650	119.033	99.787

Relative humidity	ARIMA	VARMA	LSTM
RMSE	0.026	0.034	0.037

Surface tempertature	ARIMA	VARMA	LSTM
RMSE	1.422	0.478	0.793

CO2	ARIMA	VARMA	LSTM
RMSE	9.892	1.001	0.805

Anomalies Land&Ocean temperature	ARIMA	VARMA	LSTM
RMSE	0.120	0.088	0.081

GMSL	ARIMA	VARMA	LSTM
RMSE	29.974	-	3.319

Conclusions

- ➔ Analysed a time series dataset using three different methods;
- ➔ Highlighted the differences between them;
- ➔ Present the main steps done to compute the predictions
- ➔ Improving may be provided tuning the parameters before fit the model

- ➔ Useful for organizations to be aware of the actual climate change
- ➔ Starting point for future in-depth analyses to try to edge out what is happening.

Thank you for the attention!

Stefania Sferragatta
Matricola 1948081