Advanced Data Science with IBM Specialization

Capstone project:

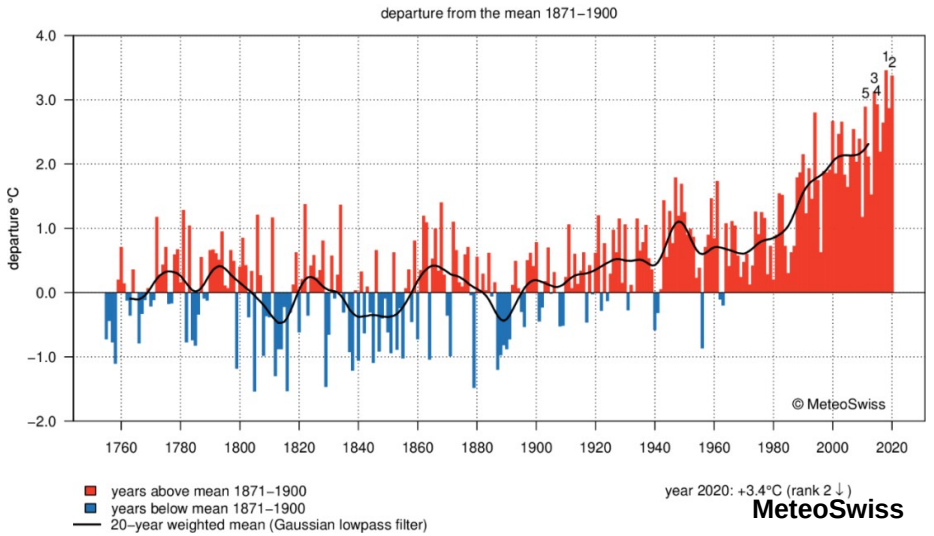
Analyze, model and forecast extreme weather events in Basel (Switzerland, CH)

July, 2021

Edoardo Farina

Introduction

Climate change is, nowadays, a fact, with an increasing temperature recorded world-wide with respect to the period 1900-1980.



Increasing temperature does not only affect Arctic ice melting or glaciers in the Alps, but induces deep upheavals in the climate.

In particular, the first aspect we are already facing is the large increase of extreme events (heat wave, dry periods, or floods)

The project aims at:

- analyze the temperature, precipitations and weather in Basel (CH) in last 35 years, to evaluate the increase of extreme events
- Provide a ML framework to reproduce the past events and forecast the future trend

Data Source

Meteorological data have been obtained from https://www.meteoblue.com/

It provides .csv/.xls files including relevant information for the desired location such as:

- Temperature
- Precipitation
- Humidity
- Wind

Data are provided hourly.

In our case, data starting from 1985 are exported as csv file in two independent datasets

Project structure

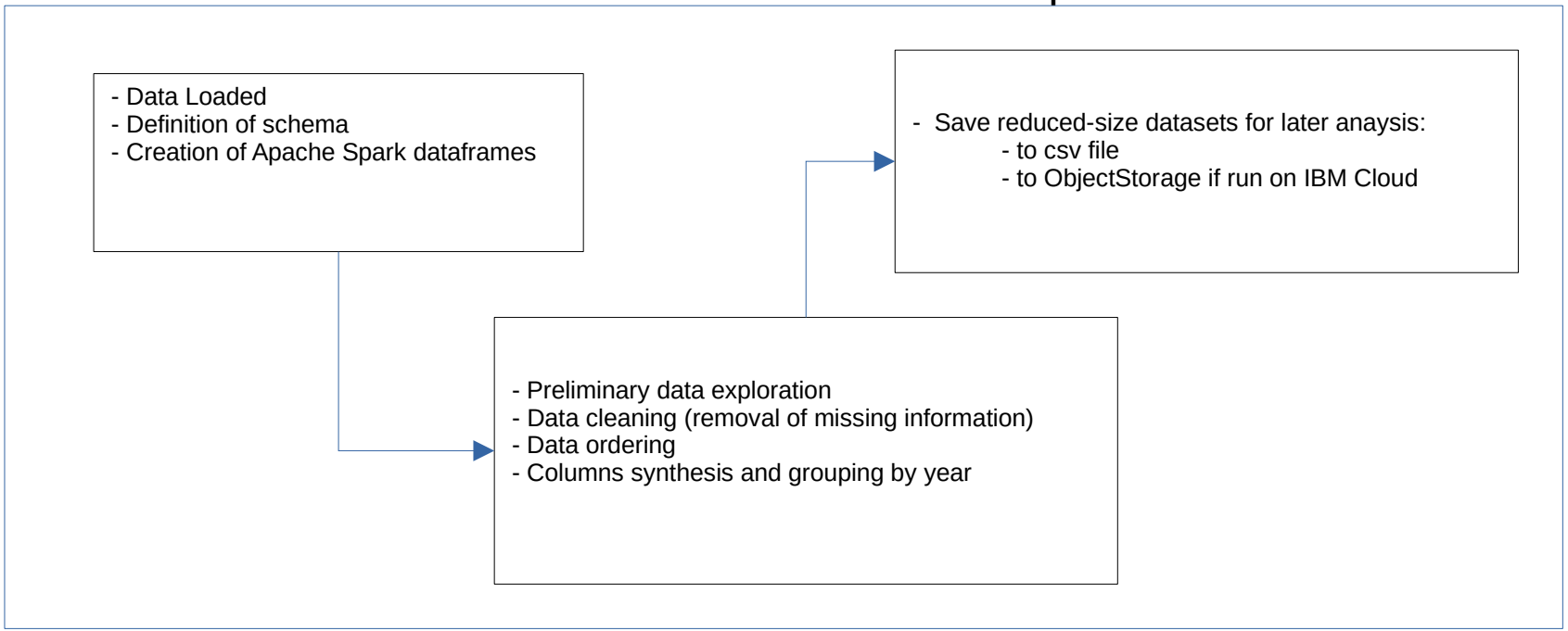
Based on Jupyter-notebooks:

1) **ETL**: Data extraction, preliminary analysis, transformation
2) **EDA**: Data analysis and visualization
3) **ML** Model definition and training
4) **ML** Model evaluation
5) **ML** Model II definition, training, evaluation and forecast

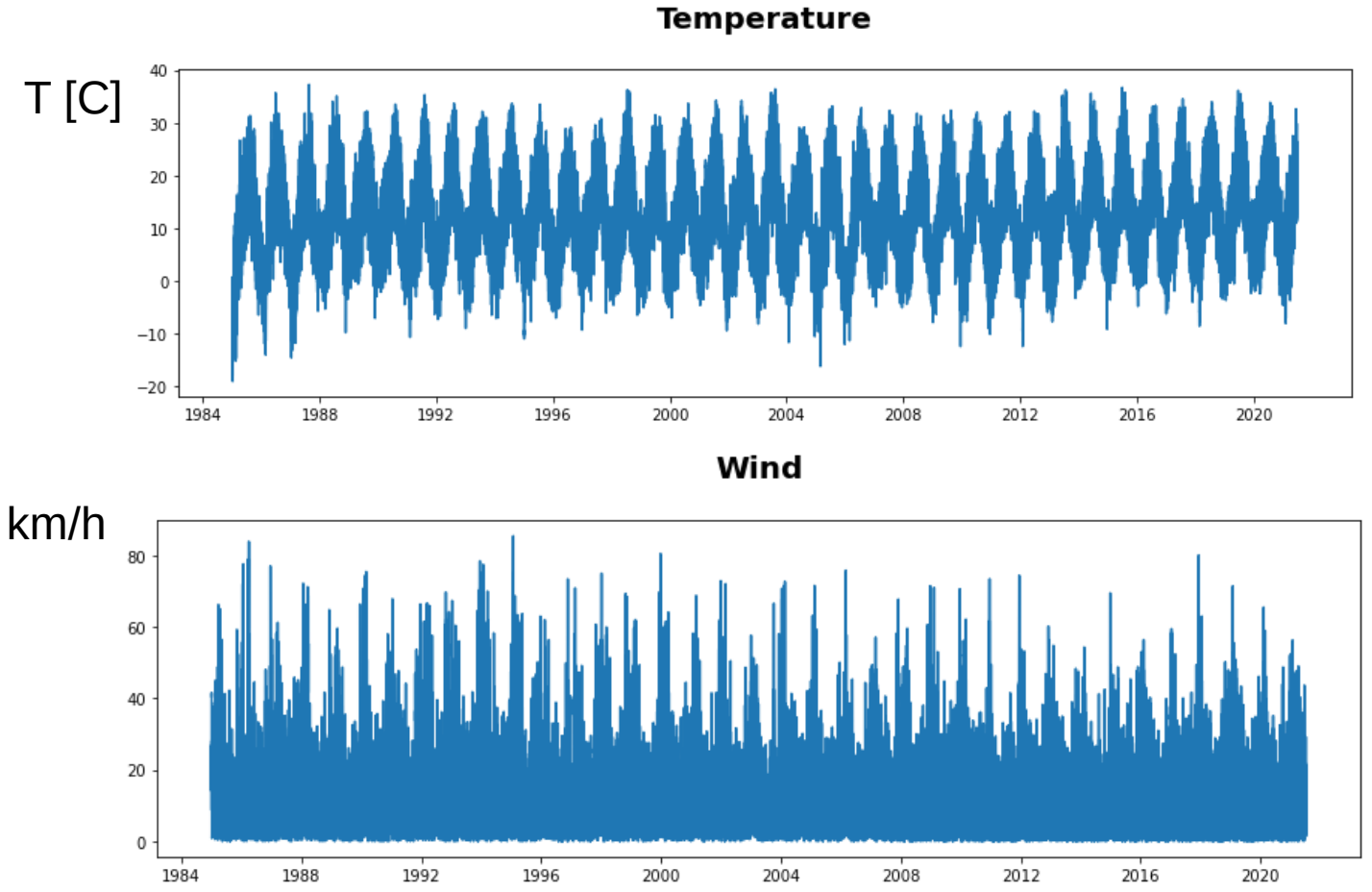
ETL

Extract – Transform - Load

Jupiter-notebook – I



ETL


Temperature
T [C]
40
30
20
10
0
−10
−20
1984
1988
1992
1996
2000
2004
2008
2012
2016
2020
Wind
km/h
80
60
40
20
0
1984
1988
1992
1996
2000
2004
2008
2012
2016
2020

ETL

Per-day maximum, average, minimum temperature, humidity and precipitation intensity are computed and saved

```
df_clean.printSchema()
```

```
root
 |-- Day: date (nullable = true)
 |-- MaxT: float (nullable = true)
 |-- MinT: float (nullable = true)
 |-- AvgT: double (nullable = true)
 |-- MaxWind: float (nullable = true)
 |-- AvgWind: double (nullable = true)
 |-- MaxRain: float (nullable = true)
 |-- MinRain: float (nullable = true)
 |-- AvgRain: double (nullable = true)
 |-- MaxHU: float (nullable = true)
 |-- MinHU: float (nullable = true)
 |-- AvgHU: double (nullable = true)
```

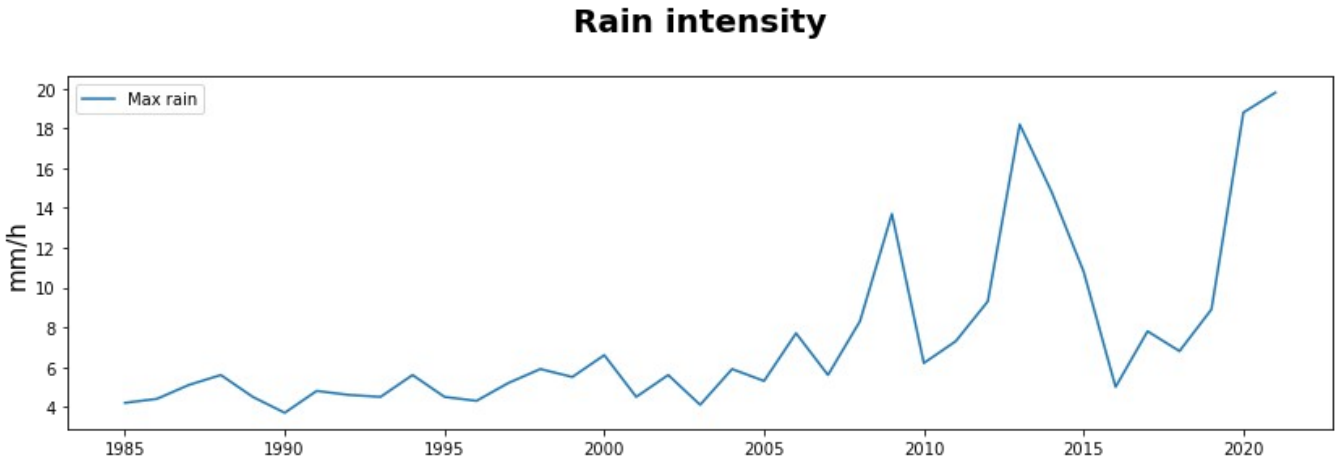
```
print("In total: {} years data".format(df_clean.count()/365))
```

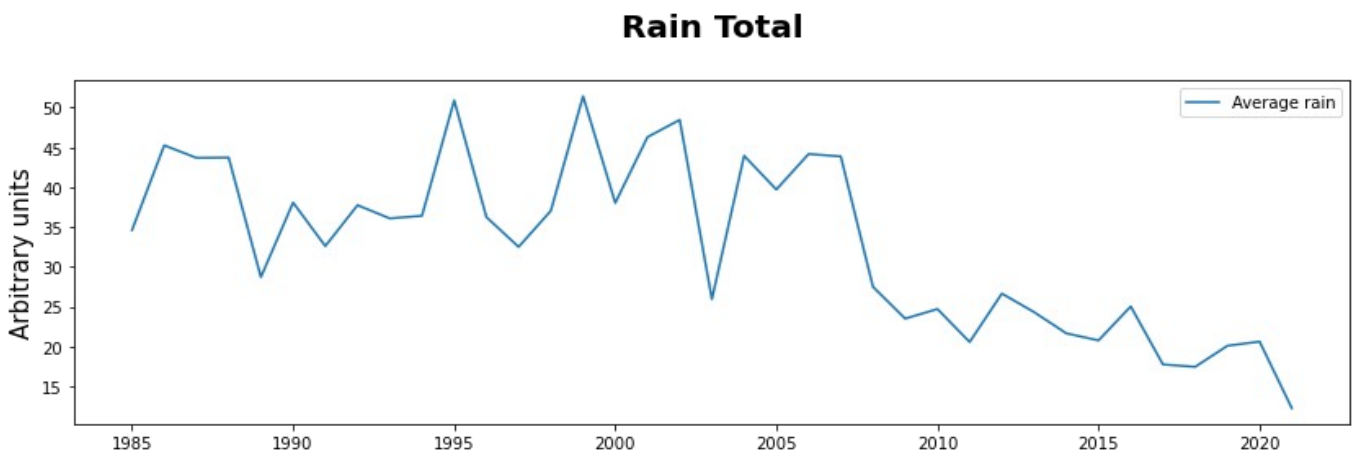
```
In total: 36.53972602739726 years data
```

EDA

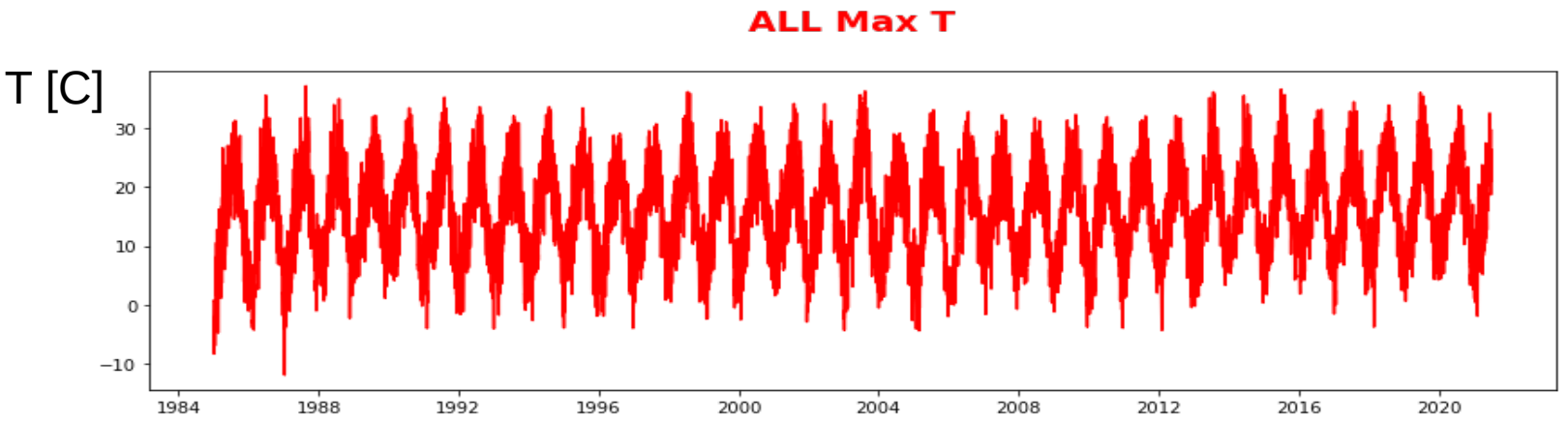
Preliminary analysis of the data shows a large increase of the maximum rain intensity as a function of the year, while a decrease of the average precipitation

It rains less, but more violently

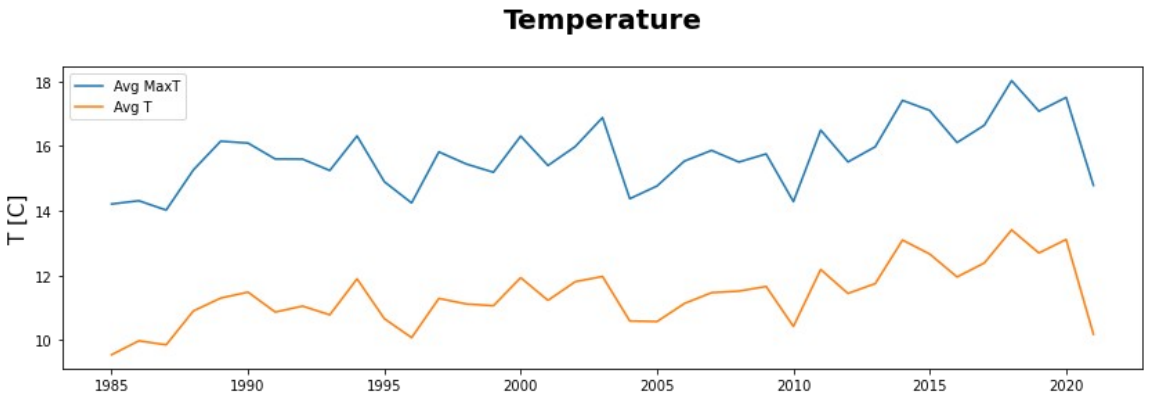




EDA



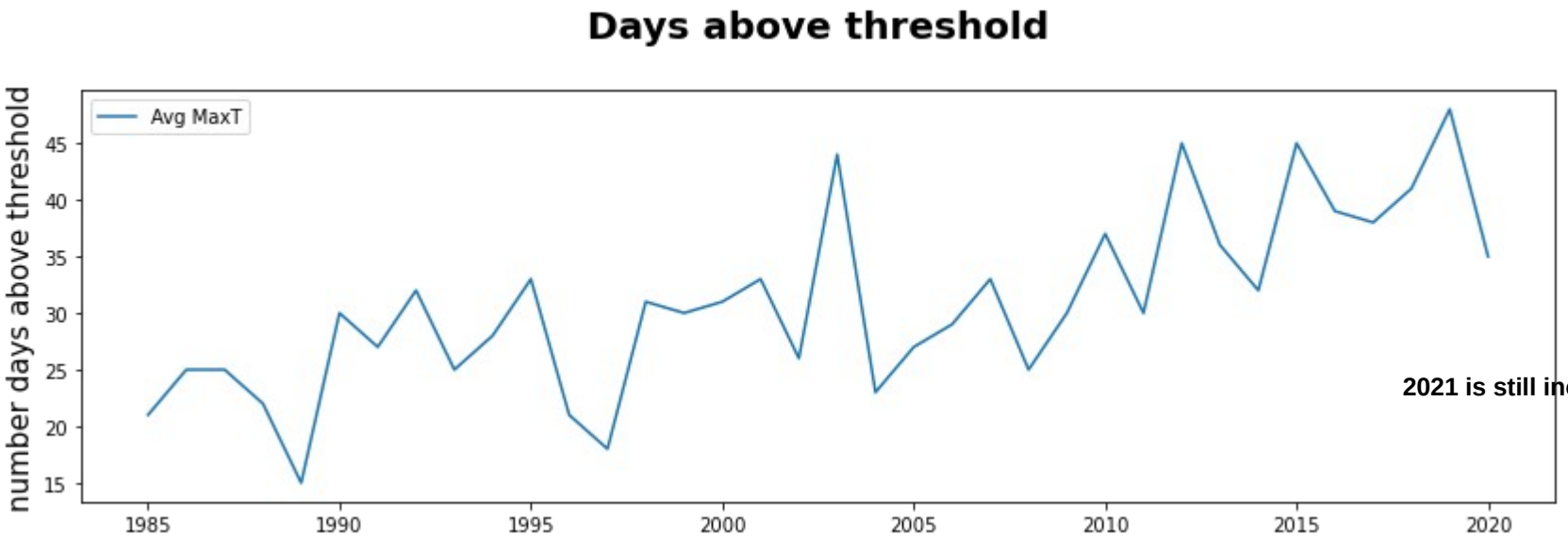
Both maximum temperature and average temperature are increasing over the years



Data engineering

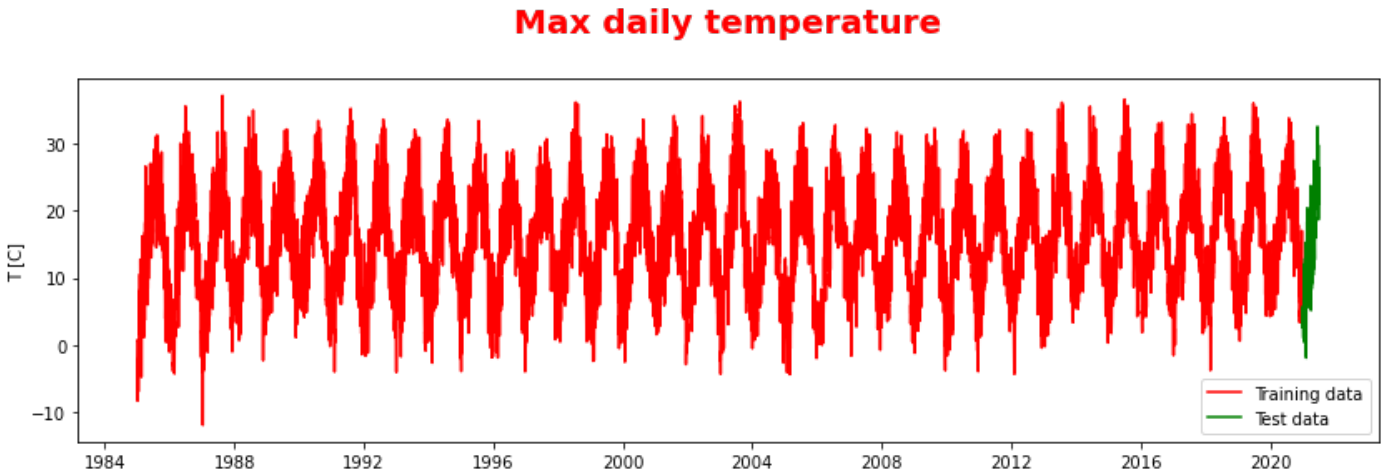
An immediate and tangible sign of climate change is the increase of violent events, heat waves, storms, floods

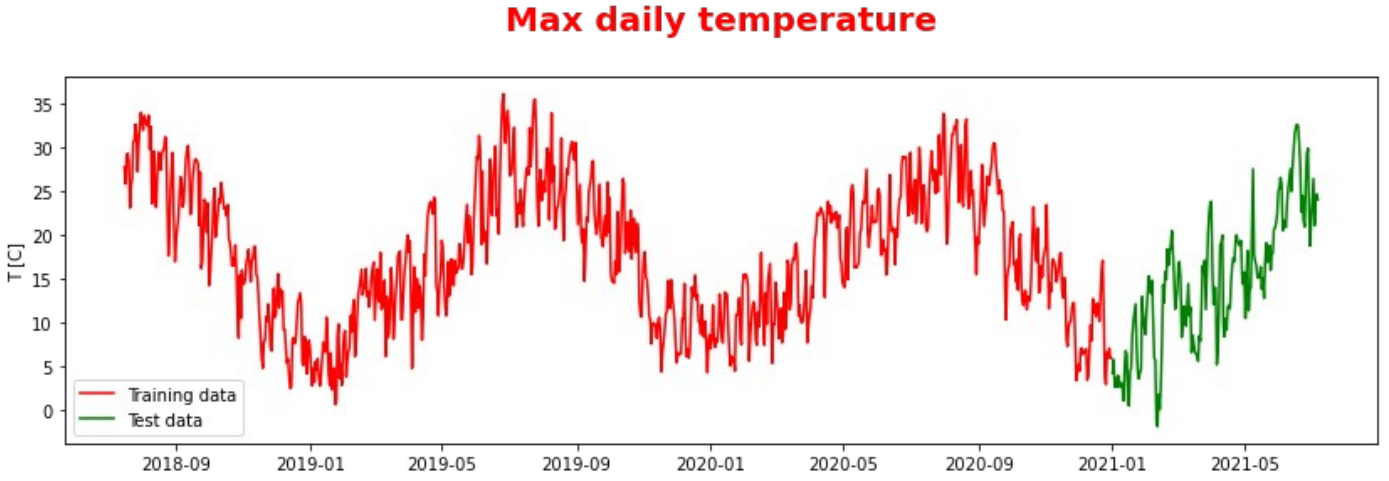
Above threshold events are defined as:
- temperature above 30C
- Hourly precipitation above 2mm/h



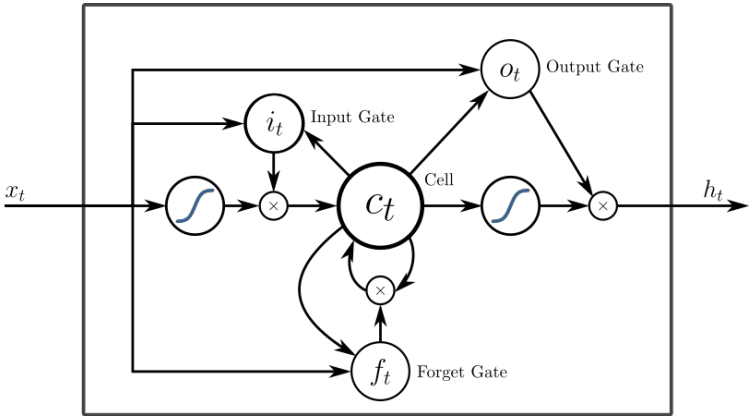
2021 is still incomplete

RNN - LSTM model





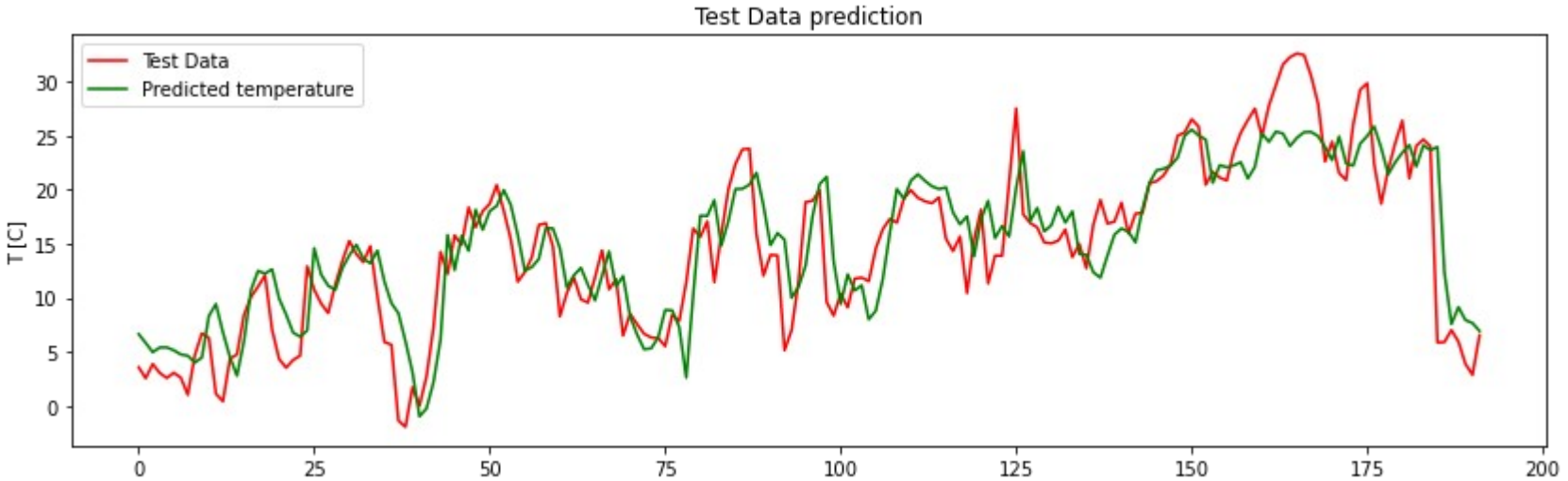
RNN - LSTM model



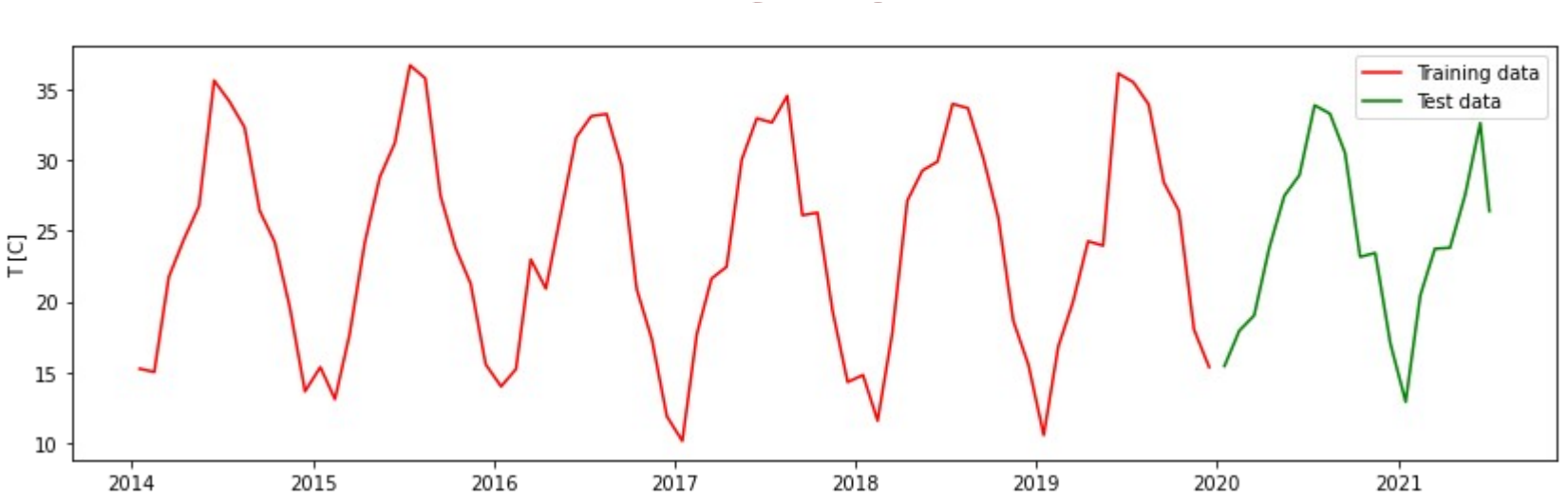

```
# Input
inputs = Input(batch_shape=(batch_size,timesteps,1))

# Layer 1: LSTM
lstm_1 = LSTM(12,
              activation='tanh',
              recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(inputs)
# Layer 2: LSTM
lstm_2 = LSTM(32,
              activation='tanh',
              recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(lstm_1)
# Layer 2: LSTM
lstm_3 = LSTM(64,
              activation='tanh',
              recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(lstm_2)

lstm_4 = LSTM(32,

              stateful=True,
              return_sequences=True)(lstm_3)
lstm_5 = LSTM(12,
              stateful=True,
              return_sequences=False)(lstm_4)



# Output
output = Dense(units = 1)(lstm_5)

# Sticking all layers into a Model
regressor = Model(inputs=inputs, outputs = output)

#adam is fast starting off and then gets slower and more precise
regressor.compile(optimizer='adam', loss = mean_squared_error)
```

RNN - LSTM model

RMSE: 3.929
Loss: 26.60%

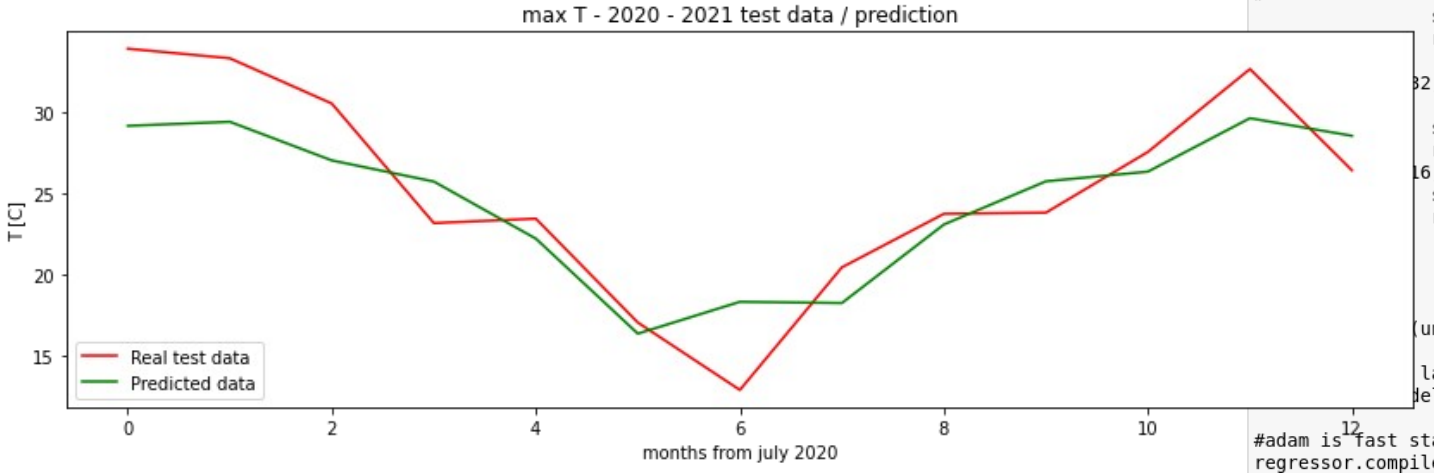


RNN - LSTM model 2 – monthly maximum



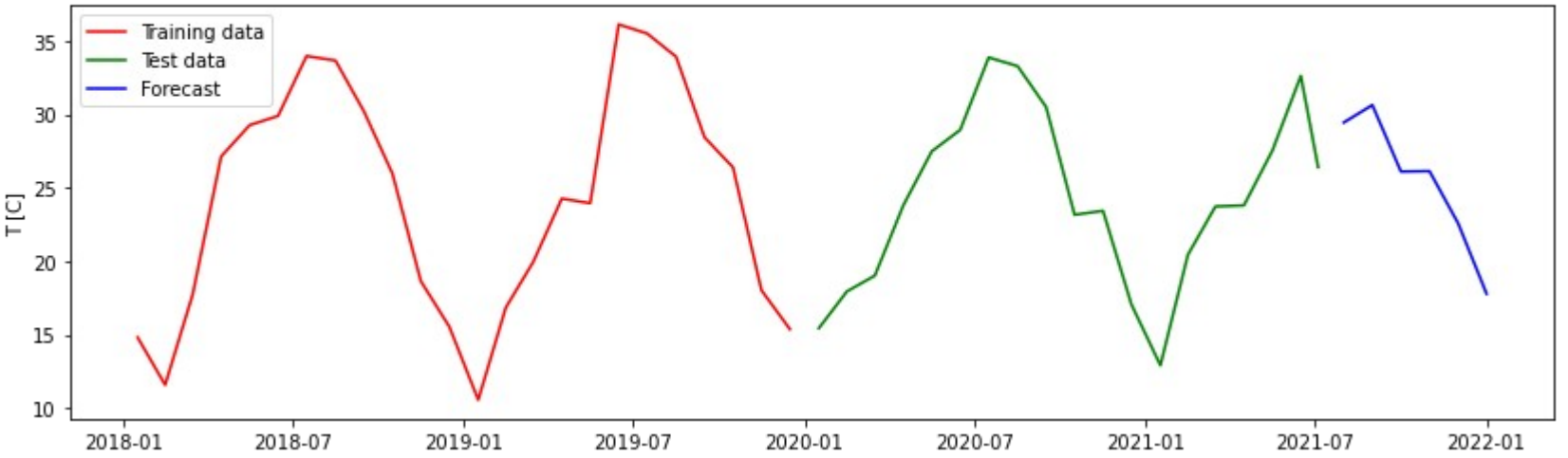
RNN - LSTM model 2 – monthly maximum

Good performance is achieved on unseen data!
RMSE: 24.89
Loss: 14.22%




```
# Input
inputs = Input(batch_shape=(batch_size,timesteps,1))

# Layer 1: LSTM
lstm_1 = LSTM(16,
#               activation='tanh',
#               recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(inputs)
# Layer 2: LSTM
lstm_2 = LSTM(32,
#               activation='tanh',
#               recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(lstm_1)
# Layer 2: LSTM
lstm_3 = LSTM(64,
#               activation='tanh',
#               recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(lstm_2)
lstm_4 = LSTM(64,
#               activation='tanh',
#               recurrent_activation='sigmoid',
              stateful=True,
              return_sequences=True)(lstm_3)

                                    32,

              stateful=True,
              return_sequences=True)(lstm_4)
                                    16,
              stateful=True,
              return_sequences=True)(lstm_5)




                                    (units = 1)(lstm_6)

                                     layers into a Model
                                    del(inputs=inputs, outputs = output)

#adam is fast starting off and then gets slower and more precise
regressor.compile(optimizer='adam', loss = mean squared error)
```

RNN - LSTM model 2 – forecasting….



Conclusions

- The analysis of the principal meteorological variables collected since 1984 in the city of Basel (CH) has been performed

- A clear trend in the precipitation and temperature can be observed as a function of the years. It rains less, but floods are more frequent. Both average and maximum temperature increased since 1984

- extreme events have increased of almost factor 2 since 1984 and will continue to increase

- A LSTM model has been developed to model the data trend, and works satisfactory on the 2021-previously-unseen data

- Further optimization of the model are foreseen: parameters tuning, longer training, different activation functions