Air Quality Measurement in Poland

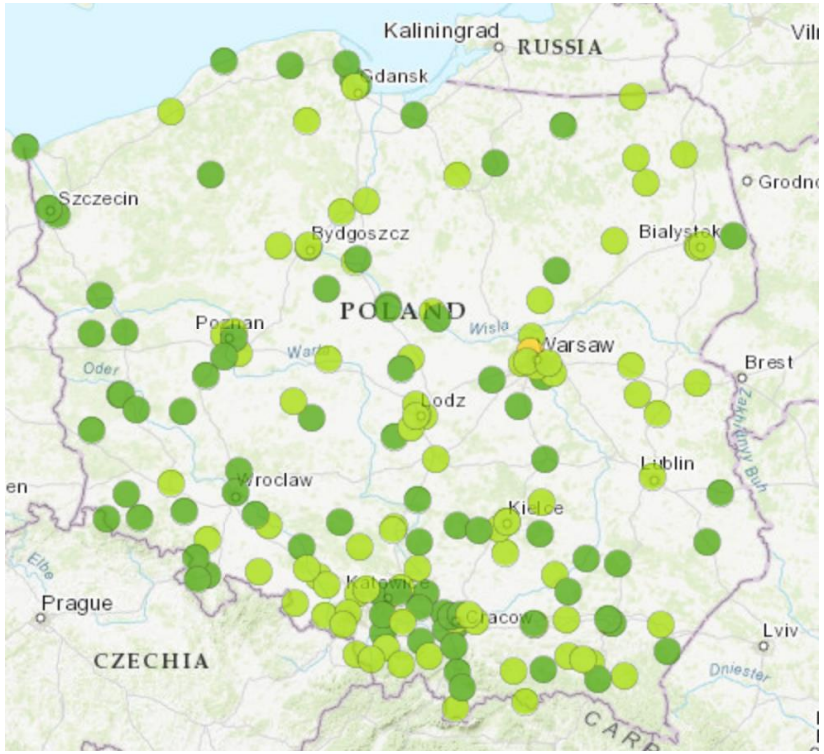
Tobias Kaps, Niklas Lange, Svea Worms, Niklas Theis, Sarah Flohr

Content

- Area of interest
- Collected data
- Research Questions
- Proposed model
- Data visualization
- Conclusion

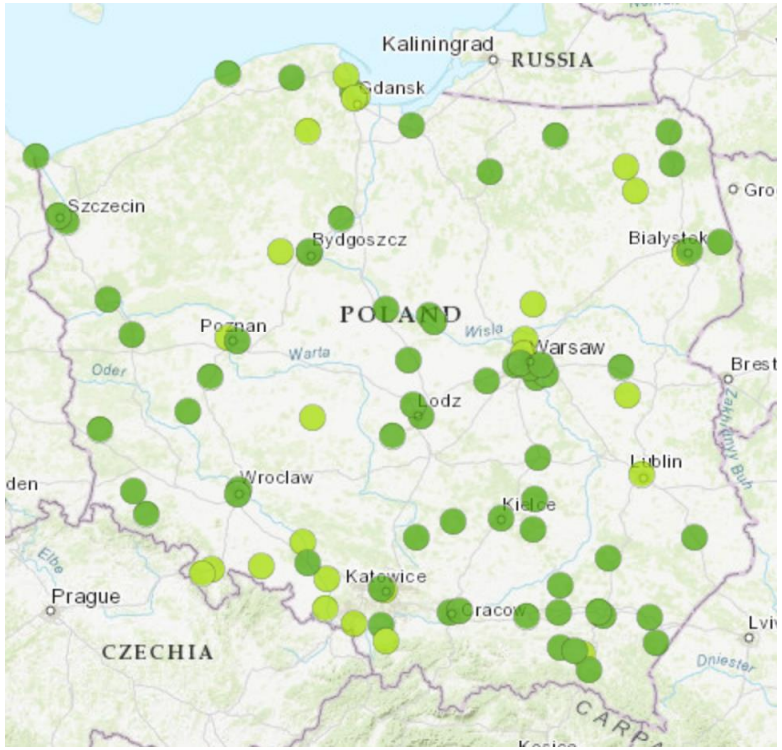
Area of Interest

PM_{10}



0 - 20 µg/m^3	Very good
20.1 - 50 µg/m^3	Good
50.1 - 80 µg/m^3	Moderate
80.1 - 110 µg/m^3	Sufficient
110.1 - 150 µg/m^3	Bad
> 150 µg/m^3	Very bad

$PM_{2.5}$



0 - 13 µg/m^3	Very good
13.1 - 35 µg/m^3	Good
35.1 - 55 µg/m^3	Moderate
55.1 - 75 µg/m^3	Sufficient
75.1 - 110 µg/m^3	Bad
> 110 µg/m^3	Very bad

Collected Data

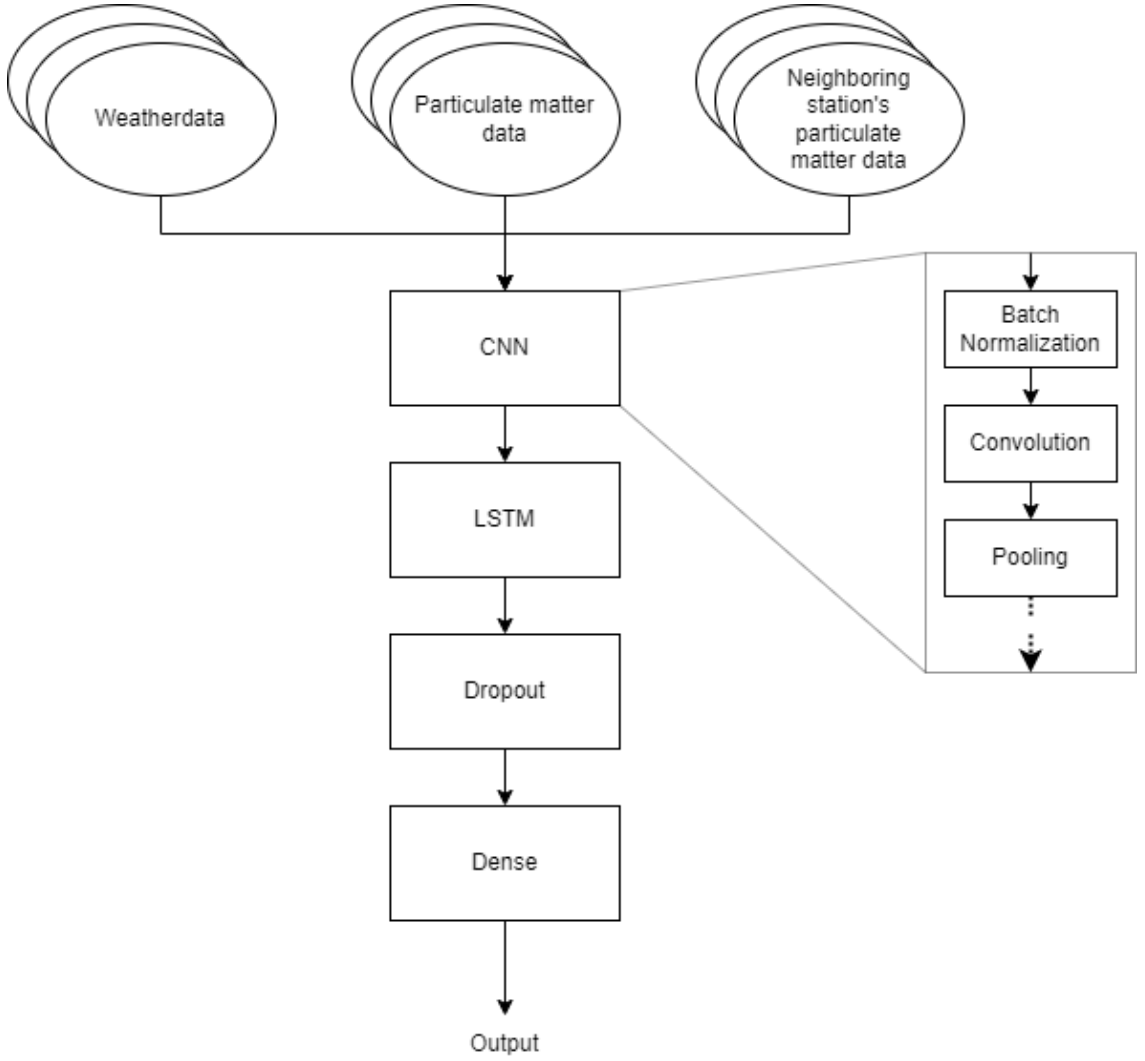
- Particulate Matter:
  - PM_{10}
  - $PM_{2.5}$
- Weather data:
  - Temperature
  - Humidity
  - Wind direction
  - Windspeed
  - Precipitation
- Time period considered:
  - 2018 - 2022

Research Questions

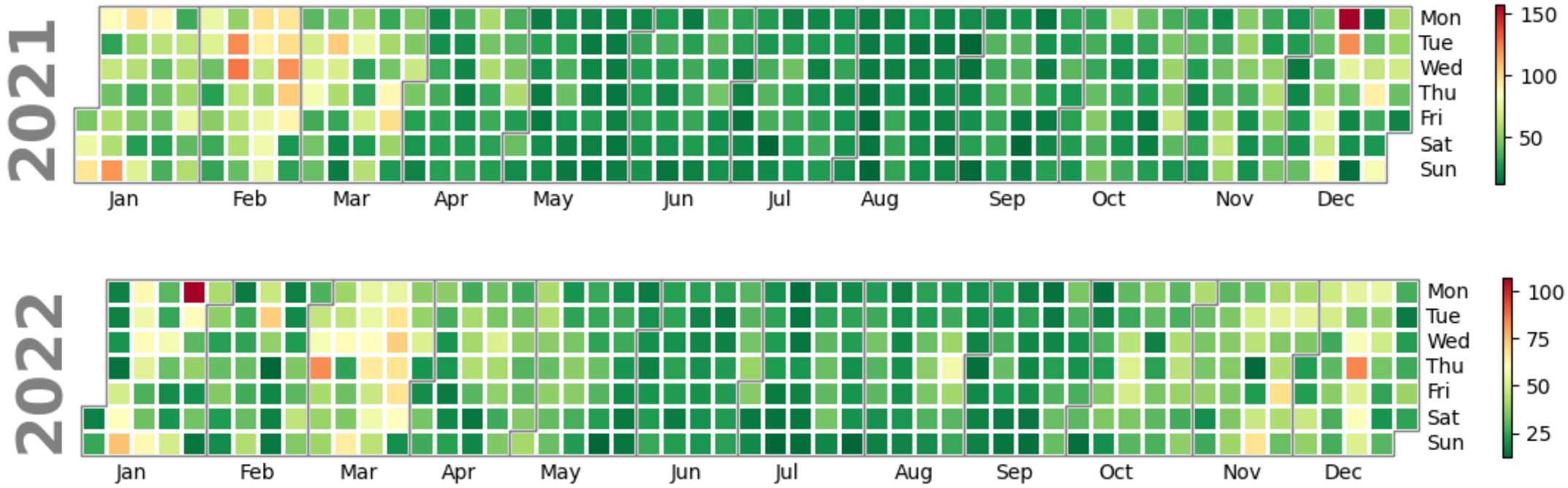
- Is it possible for a neural network to generate an hourly forecast of particulate matter concentration over the next 14 days using a combination of CNN and LSTM?
  - Is it possible to predict the PM_{10} value with a MAE below 10?
  - Is it possible to predict the $PM_{2.5}$ value with a MAE below 10?
  - Is there a relationship between PM_{10} and $PM_{2.5}$ such that $PM_{2.5}$ can be predicted using the model for PM_{10}?
  - How do our forecasts compare with those of the Polish Environmental Protection Agency? (for one day)
  - Does it make sense to combine stations into areas, so that the predictive power remains, is the same or improved compared to individual stations?

Model

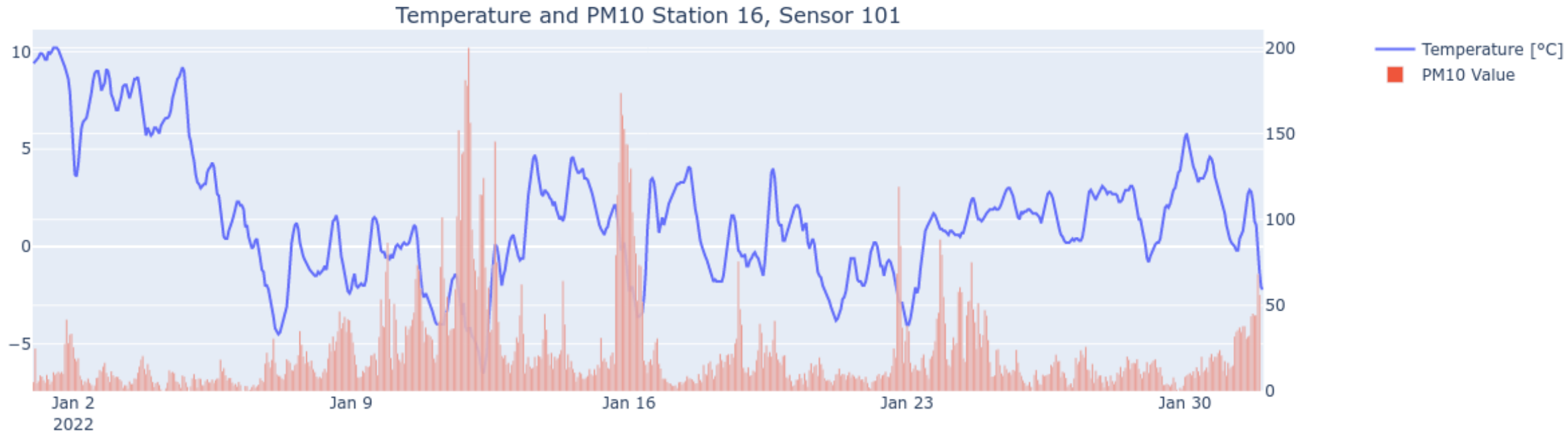
- Supervised ML problem
- CNN-LSTM combination for solving the regression problem
  - Multi-Step Forecasting (PM_{10})
- Train data and test data 1 year
  - Total 5 years collected
- Inclusion of 3 adjacent stations
  - Weighting according to distance
- Creation and Training in tensorflow



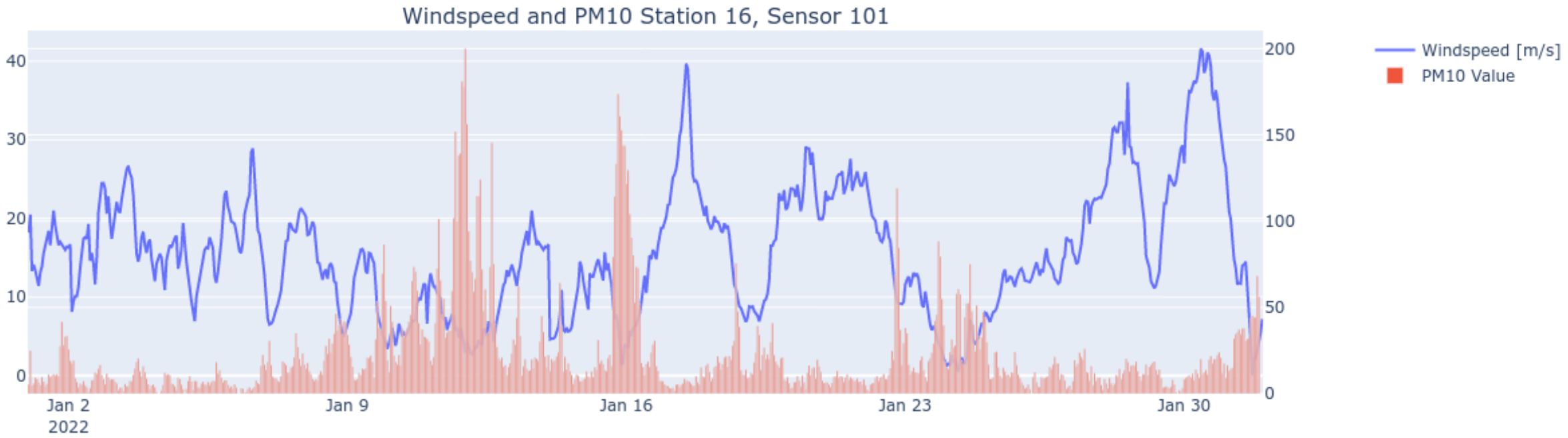
Heatmaps for PM_{10} daily average value



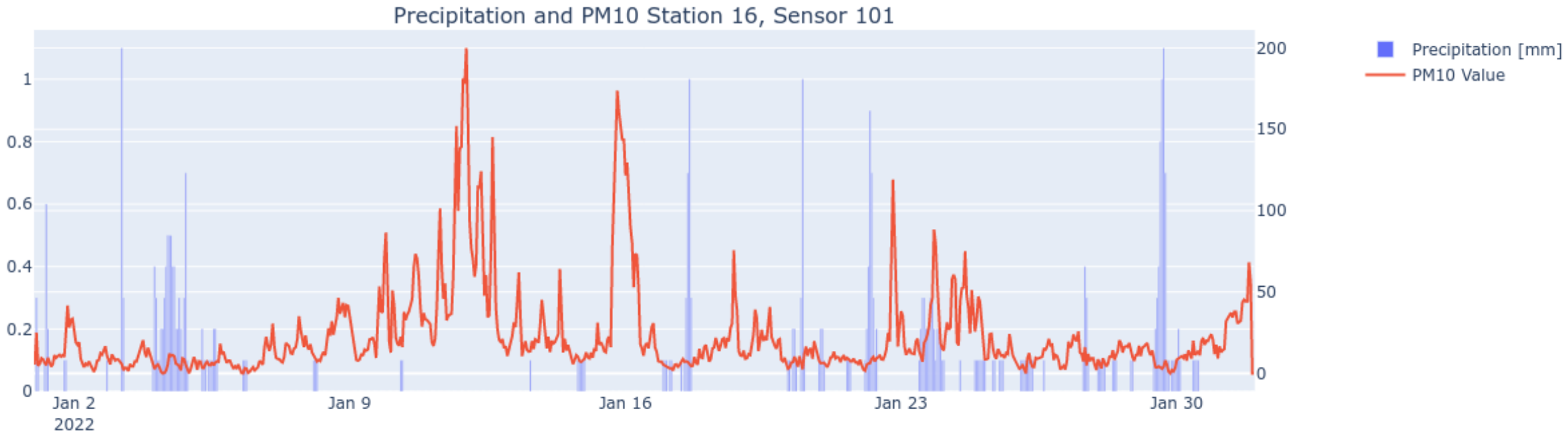
Data insights



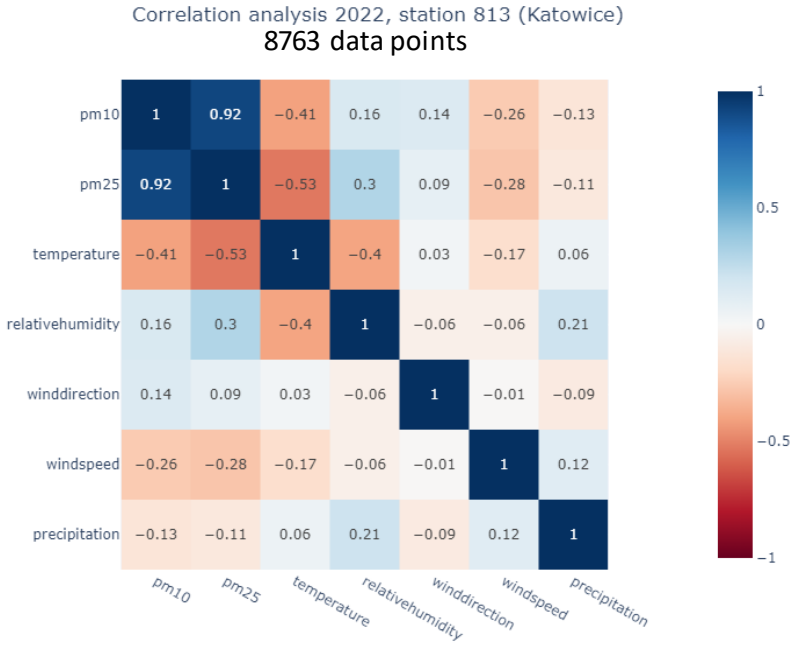
Data insights

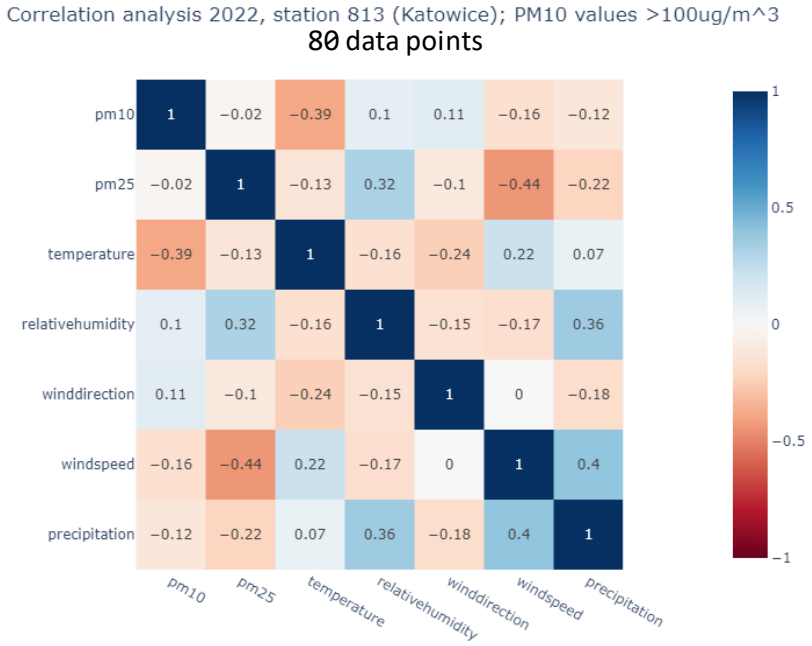


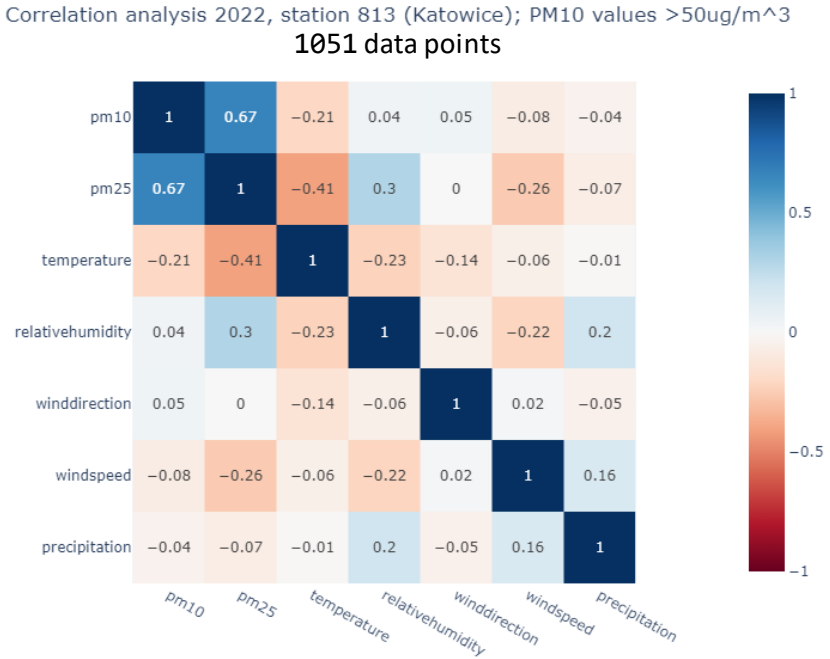
Data insights

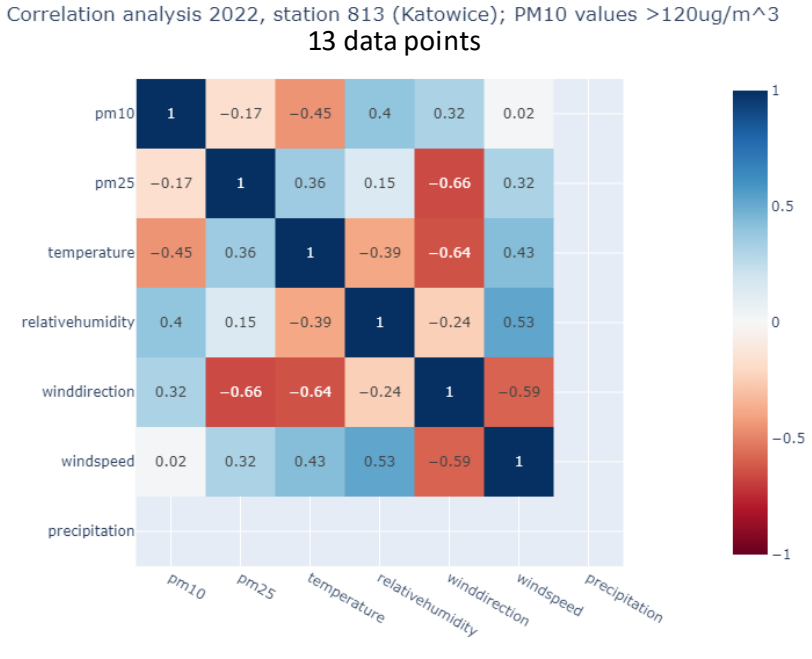


Correlation Analysis









Conclusion

- Although we would have expected otherwise, we did not find any (strong) correlation between the weather data and the particulate matter pollution.
- Therefore, we will try to feed as much data as possible into the neural network to achieve a learning effect.
- There is still a need to determine how to handle extreme outliers and predict days with particularly high pollution levels