Probabilistic Air Quality Forecasting with Bayesian Inference

Course: ML703 – Probabilistic and Statistical Inference

Team Members: Ayesha Alhammadi, Muhra AlMahri

Term: Spring 2025

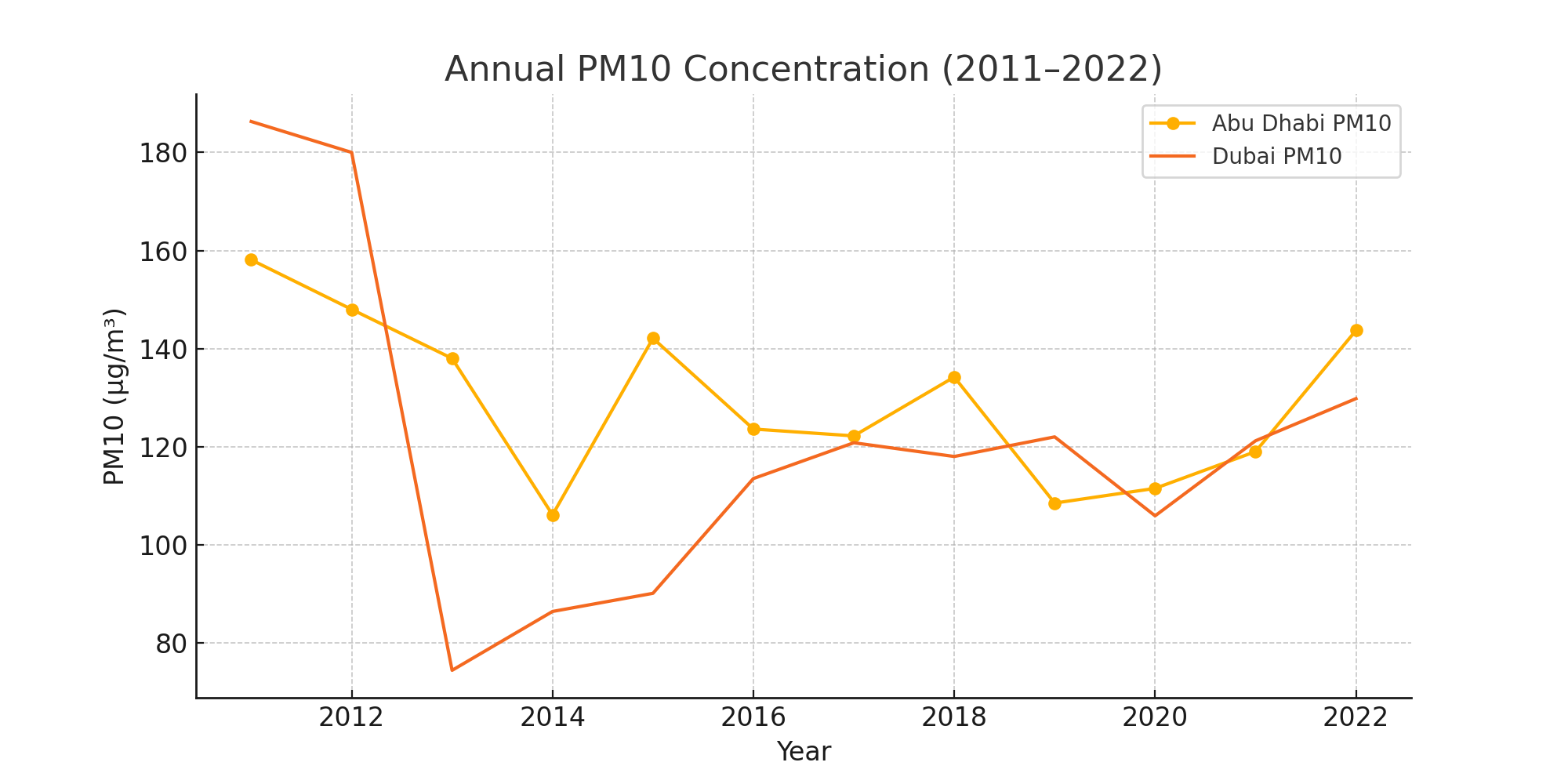
# 1. Introduction

Air pollution is a growing concern in urban environments due to its direct impact on public health, climate, and sustainability. In the UAE, cities like Abu Dhabi and Dubai frequently experience high levels of PM10 pollution (particulate matter with diameter <10 microns), often influenced by climate conditions and urbanization. This project explores time series forecasting of PM10 pollution using both classical and probabilistic models to provide not only predictions but also quantification of uncertainty.

# 2. Dataset

The dataset was obtained from the UAE Federal Competitiveness and Statistics Center (FCSC), titled: "Average of Air Pollutants Concentration by Monitoring Station (2011–2022)." Yearly PM10 concentrations were extracted for Abu Dhabi and Dubai, averaged across monitoring stations, and interpolated where missing.

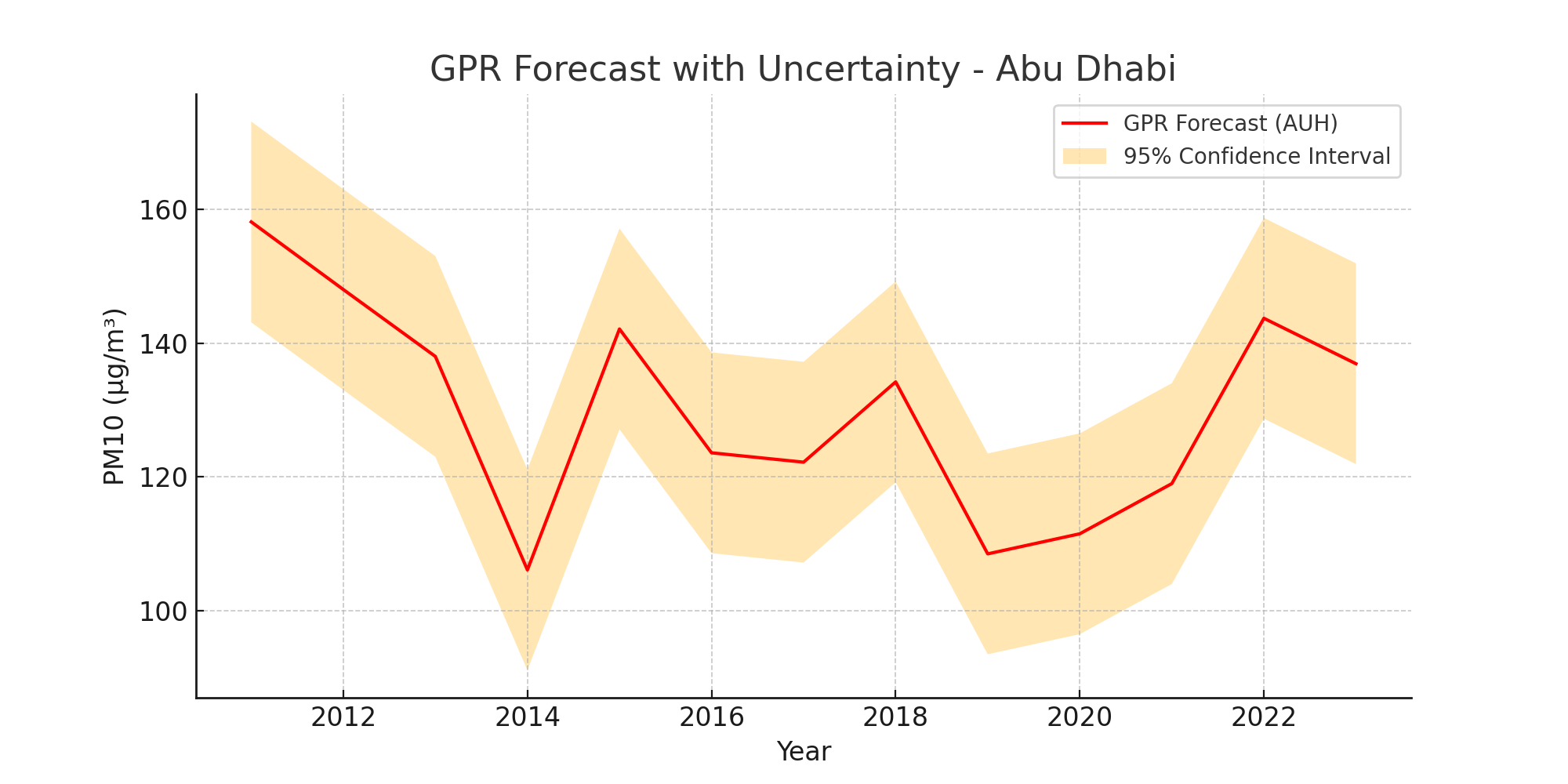
Figure 1. Annual PM10 concentration in Abu Dhabi and Dubai (2011–2022).



# 3. Modeling Approaches

Two models were implemented:  
- ARIMA(1,1,1): A classical time series forecasting method without uncertainty quantification.  
- Gaussian Process Regression (GPR): A Bayesian approach using an RBF kernel, allowing estimation of credible intervals.

Figure 2. GPR forecast with 95% uncertainty interval for Abu Dhabi PM10.



# 4. Results and Discussion

The models yielded the following 2023 forecasts (in µg/m³):  
- ARIMA: Abu Dhabi = 135.62, Dubai = 129.75  
- GPR: Abu Dhabi = 136.91 [95% CI: 121.9–151.9], Dubai = 127.58 [95% CI: 110.2–144.7]  
  
While ARIMA provided reasonable point forecasts, GPR offered added value through its uncertainty quantification. This is crucial for policy planning and environmental risk assessment. Despite the small dataset, the Bayesian approach produced smooth forecasts.

# 5. Conclusion

This project successfully applied ARIMA and Bayesian GPR to forecast PM10 levels in two UAE cities. The results demonstrate the advantages of probabilistic models in environmental forecasting. Future work could expand this analysis with more granular data (e.g., hourly/daily) and exogenous variables such as weather or traffic patterns.

# 6. References

- Rasmussen, C.E., & Williams, C.K.I. (2006). Gaussian Processes for Machine Learning. MIT Press.  
- Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice.  
- UAE Federal Competitiveness and Statistics Center (2022).