

Gaussian Processes for Time Series

Raphaela Azar and Sbonelo Gumede

Supervised by Prof. Birgit Erni

The University of Cape Town

August 15, 2025

Introduction

In this paper we hope to define what is a Gaussian process. How can you use a Gaussian process for time series analysis? Compare the performance of a Gaussian process to other time series analysis models such as state-space models. An area of application included is predicting the level of air pollution in the Table View station in Cape Town for the year of 2019. Another area of application is predicting the gold price for South Africa for the year of 2020. This is done in a Bayesian framework in order to make probabilistic statements and obtain a predictive distribution for future forecasts.

A Gaussian process is a stochastic process denoted by $\{X_t\}_{t \in \mathcal{I}}$, where the parameter space \mathcal{I} and the state space \mathcal{U} are real numbers. At each point $t \in \mathcal{I}$, the state $X_t \sim \mathcal{N}(f(x_t), g(x_t))$, where $f(x_t)$ is a mean function and $g(x_t) \geq 0$ is the covariance function. Consider a simple example shown in figure 1, where $f(x_t) = x_t + e$, where $e \sim \mathcal{N}(0, 1)$ and $g(x_t) = 1 + \sin^2(x_t)$.

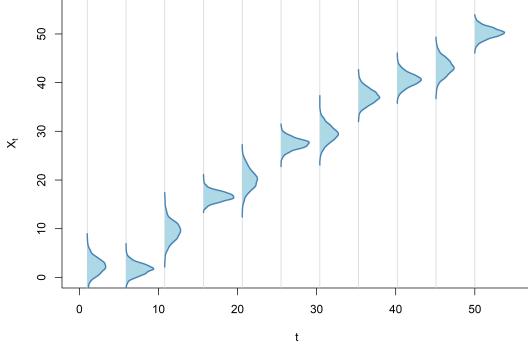


Figure 1: An example of a Gaussian process.

A special property of Gaussian processes is that if you take a countable finite set of observations $\{x_1, \dots, x_p\}$ their joint distribution is

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ where } \boldsymbol{\mu} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_p) \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} g(x_1, x_1) & g(x_1, x_2) & \cdots & g(x_1, x_p) \\ g(x_2, x_1) & g(x_2, x_2) & \cdots & g(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ g(x_p, x_1) & g(x_p, x_2) & \cdots & g(x_p, x_p) \end{bmatrix}.$$

Often in the literature people use $f(x_t) = 0$ and $g(x_i, x_j) = \alpha^2 \exp(-\frac{1}{2}(\frac{|x_i - x_j|}{\rho})^2) + \sigma \delta_{ij}$, where α and ρ are hyperparameters and δ_{ij} is an indicator function. The latter is known as the squared-exponential kernel. Some kernel functions common in the literature are the following:

There are two ways of modeling time series data. The first one is functional mapping $y = f(x) + e$, where y is the response variable, x is the explanatory variable, and e is typically white additive

noise. The noise is assumed to be uncorrelated with one another. Although the simplicity of functional mapping is attractive, functional mapping neglects the time component rendering it useless for our purposes. The second one is curve fitting, where x is often time and y is the response variable. This is ideal for our purposes since we are accounting for the time dependence between successive observations.

Air pollution example

Exploratory data analysis

Variable	Name	Description	Unit
NO ₂	Nitrogen dioxide	A harmful gas from vehicles and industry.	µg/m ³
PM ₁₀	Particulate matter 10	Small inhalable dust particles.	µg/m ³
SO ₂	Sulphur dioxide	Mainly from burning fossil fuels.	µg/m ³
Speed	Wind speed	How fast the wind is moving.	m/s

Table 1: Description of variables of the air pollution dataset.

Variable	Min.	1st Qu.	Median	Mean	Std.	3rd Qu.	Max.	NA's
NO ₂	0.0	5.0	9.0	12.73	10.72	17.0	113.0	734
PM ₁₀	0.0	12.0	17.0	19.82	12.30	24.0	158.0	298
SO ₂	0.0	2.0	3.0	6.20	11.30	5.0	142.0	638
Speed	0.5	2.3	3.6	3.74	1.71	4.9	11.2	918

Table 2: Summary statistics of the air pollution dataset.

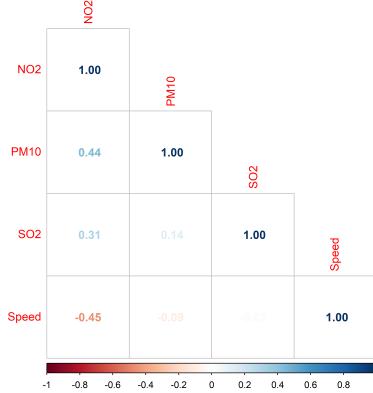
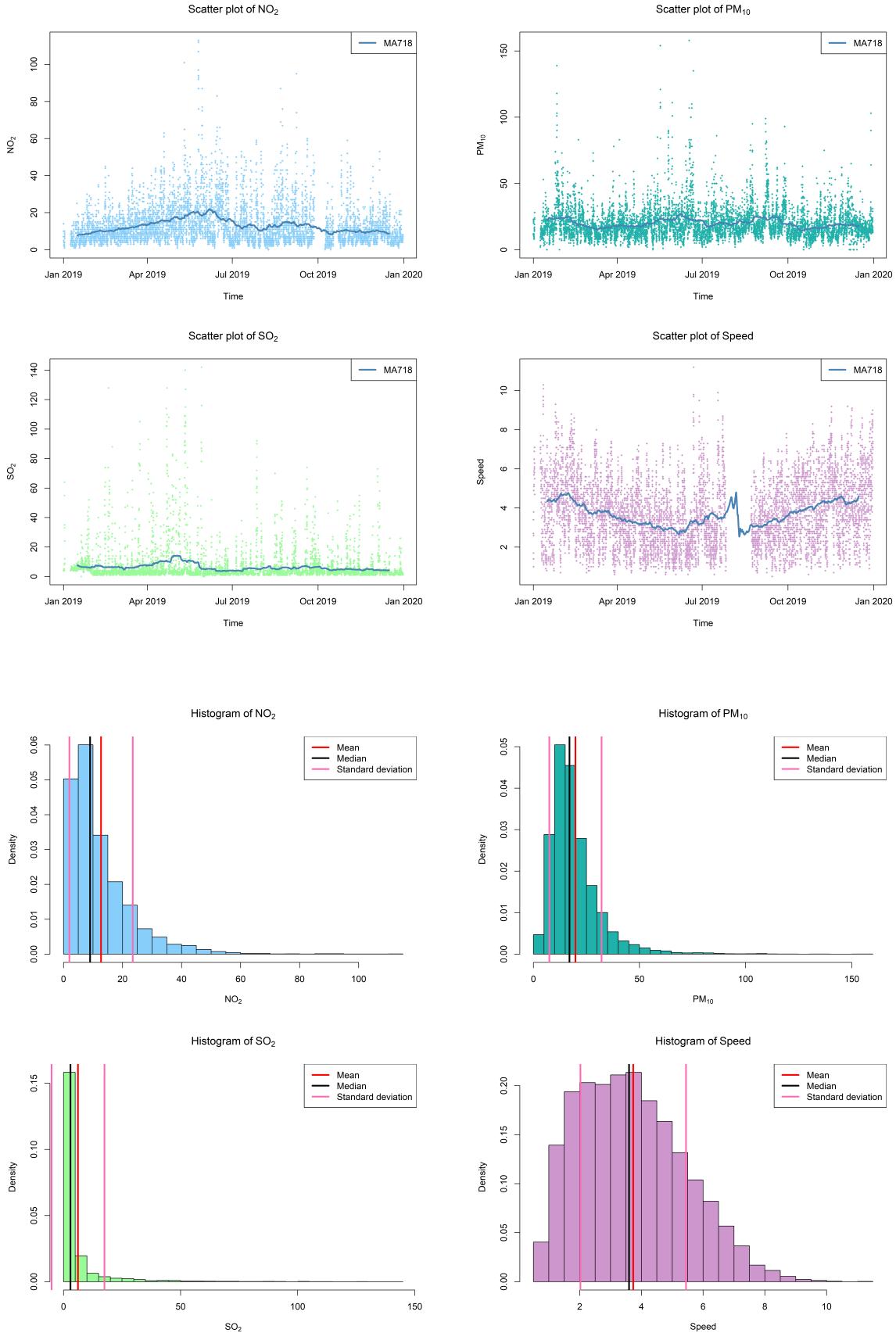


Figure 2: Correlation plot of the air pollution dataset.

Our response variable NO₂ appears to be moderately positively correlated with PM₁₀ and SO₂, and moderately negatively correlated with Speed. These are not ideal explanatory variables since we typically would like them to be strongly correlated with the response variable. The explanatory variables are weakly correlated with one another, whether it be positive or negative correlation. This is ideal since some models do not work well with correlated explanatory variables, often leading to unstable point estimates and inflated standard errors.

Before looking at the data. The proposal mean function is $f(x_t) = x_t^T \beta + e$ and the proposal covariance function is $g(x_i, x_j) = \alpha^2 \exp(-\frac{1}{2}(\frac{|x_i - x_j|}{\rho})^2) + \sigma \delta_{ij}$ for $i, j \in \mathcal{I}$. This allows us to capture the trend (if it exists) in the time series. This also allows us to capture the correlation structure of the data. The squared-exponential kernel is chosen because it is a natural way of describing the correlation decay.



The prior distributions of the exponentiated quadratic kernel are the following:
 $\alpha \sim \text{half-normal}(0, 1)$, $\rho \sim \mathcal{IG}(20, 10)$, and $\sigma \sim \text{half-normal}(0, 1)$.

Components of the time series

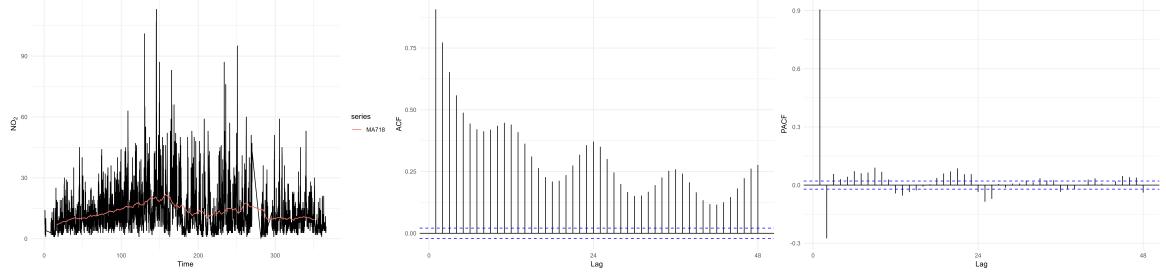


Figure 5: Time series plot of NO_2 .

This time series is not stationary. The mean and variance are changing with time. There is a need for a variance stabilizing transformation. Also possibly first-order differencing. The ACF slowly decays to 0, but the PACF shuts off abruptly after lag 2. Hence, an AR(2) model may be the appropriate order of AR model to fit the data. So we would fit an ARIMA($p=2, d, q$) and then fit some other ARIMA(p, d, q) models e.g., ARIMA($p=1, d=1, q=1$) and then compare the models with AIC.

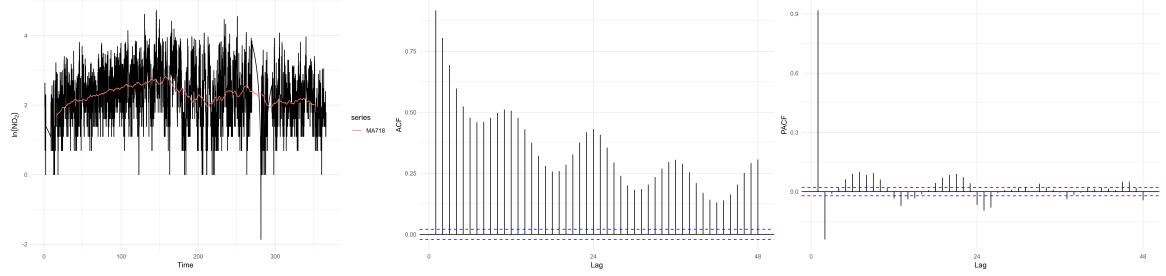


Figure 6: Time series plot of $\ln(\text{NO}_2)$.

The time series is not yet stationary since the mean is changing with time. First-order differencing is needed in order to remove the trend.

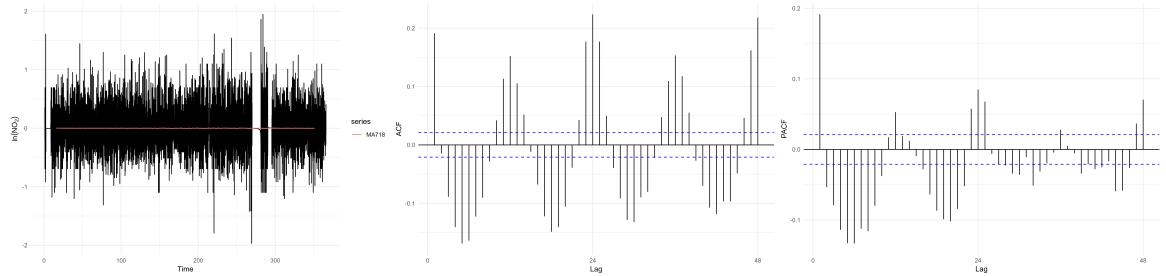


Figure 7: Time series plot of first-order differenced $\ln(\text{NO}_2)$.

Notice that the time series is now stationary. The mean and variance are not changing with time. Thus we only need to account for the mean and variance in our models.

Results

Horizon (hours)	Average	Naive	SNaive	Drift	ARIMA
1	8.79	1.00	4.00	1.00	0.44
24	8.61	1.35	3.72	1.35	3.76
168	7.16	5.58	5.71	5.59	15.57
720	6.62	7.10	7.21	7.16	48.61

Table 3: RMSE values on the test set for different forecasting methods and horizons.

References

1. <https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/nitrogen-dioxide>