Gaussian Processes for Time Series

Raphaela Azar and Sbonelo Gumede
Supervised by Prof. Birgit Erni
The University of Cape Town

August 13, 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 peformance 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 $g(x_t) = 1 + \sin^2(x_t)$.
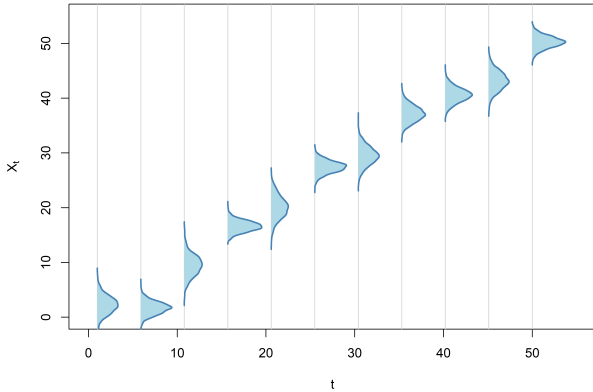

Figure 1: An example of a Gaussian process.

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

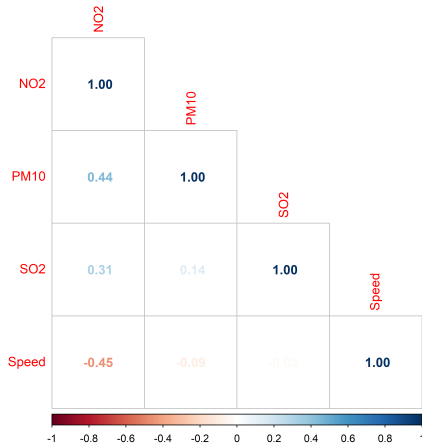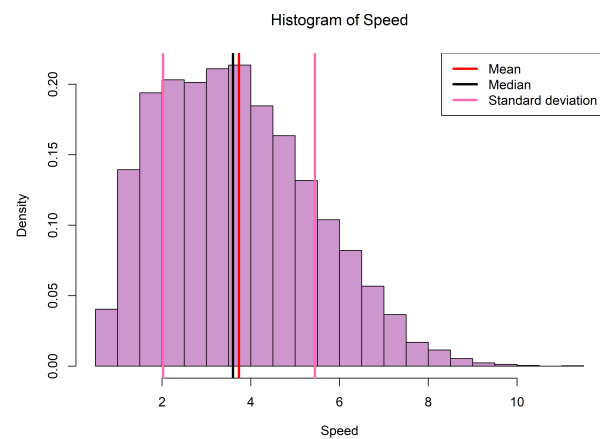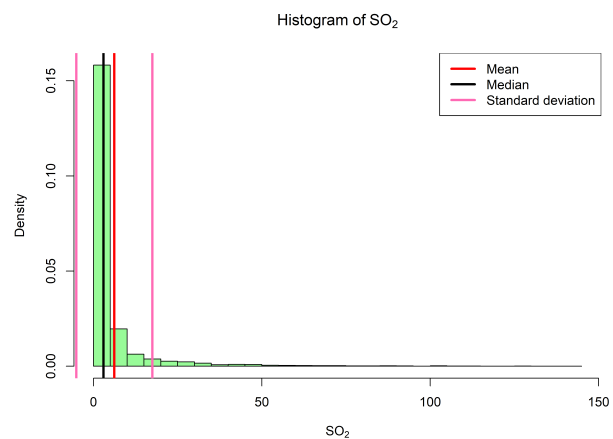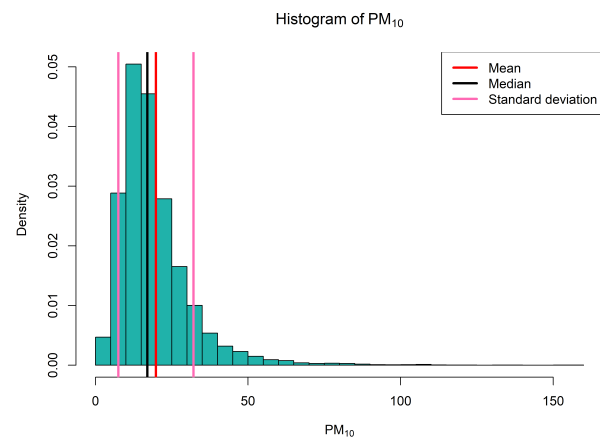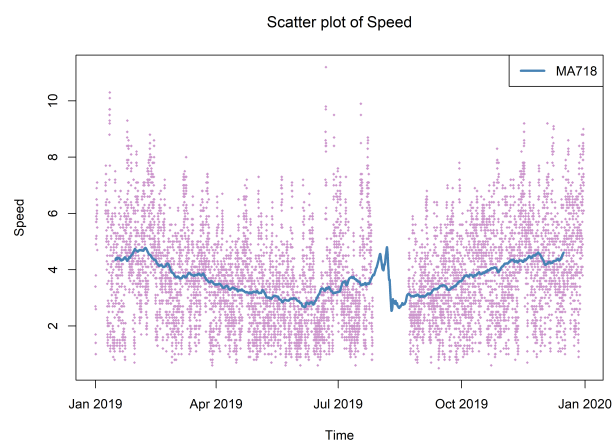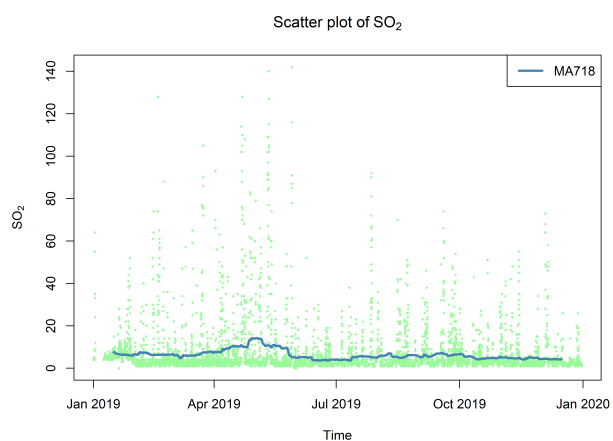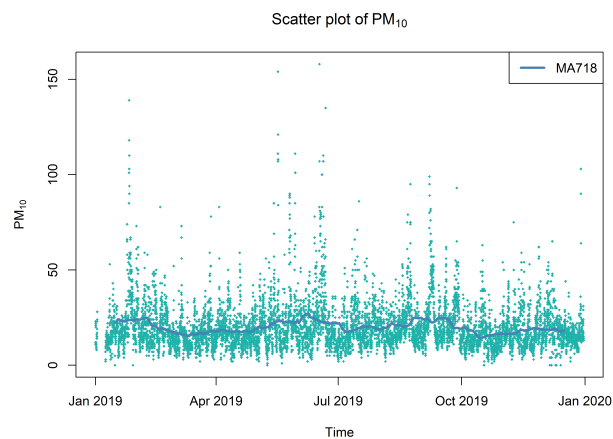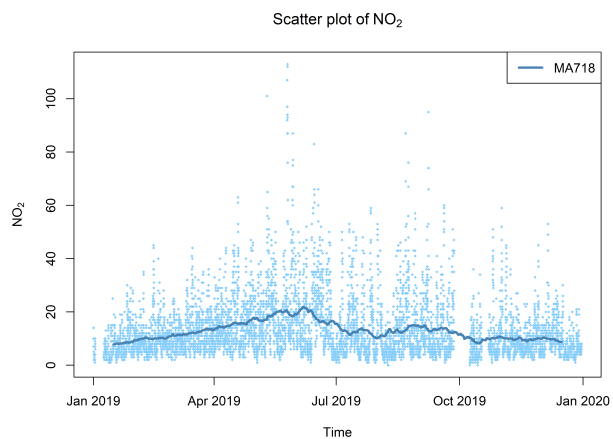Table 1: Description of variables used in the analysis.

Figure 2: Correlation plot of the variables.

Our response variable NO$_2$ appears to be moderately positively correlated with PM$_{10}$ and SO$_2$, 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 weekly 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 proposed model for the mean function is $f(x_t) = x_t^T \beta + e$ and $g(x_i, x_j) = \alpha^2 \exp(-\frac{1}{2}(\frac{|x_i - x_j|}{\rho})^2) + \sigma \delta_{ij}$ for $i, j \in \mathcal{I}$ for the covariance function. This allows us to capture the trend (if it exists) in the time series. This also allows us to model the correlation structure of the data. The squared exponential kernel is chosen because it is a natural way of describing the correlation decay.

Scatter plot of NO₂

Scatter plot of PM₁₀

Scatter plot of SO₂

Scatter plot of Speed

Histogram of NO₂

Histogram of PM₁₀

Histogram of SO₂

Histogram of Speed
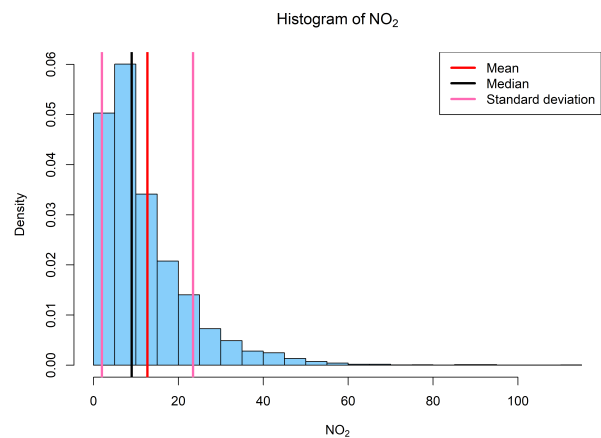
The mean appears to be constant, but the variance seems to be changing with time. Thus it is better if $f(x) = 0$ and $g(x) = \alpha^2 \exp(-\frac{1}{2}(\frac{|x_i - x_j|}{\rho})^2)$. The prior distributions of the exponantiated quadratic kernel are the following: $\alpha \sim$ half-normal$(0, 1)$, $\rho \sim \mathcal{IG}(20, 10)$, and $\sigma \sim$ half-normal$(0, 1)$. There is seasonality since regular patterns repeat for the seasons of the year. Random variation is present. However, there is no cyclical component present.

References

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