

Gaussian Processes for Time Series

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

August 28, 2025

Abstract

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.

Literature Review

((Rasmussen and Williams, 2006)) We have a training set of \mathcal{D} of n observations, $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, \dots, n\}$, where \mathbf{x} denotes the input vector (covariates) of dimension \mathcal{D} and y denotes a scalar output or target (dependent variable); the column vector of inputs for all n cases are arranged in the $\mathcal{D} \times n$ design matrix X , and the targets are collected in the vector \mathbf{y} , so we can write $\mathcal{D} = (X, \mathbf{y})$.

Definition 2.1.1 ((Rasmussen and Williams, 2006)). *A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.*

A Gaussian process is completely specified by its mean function and co-variance function. We define mean mean function $m(\mathbf{x})$ and the covariance function $k(\mathbf{x}, \mathbf{x}')$ for a real process $f(\mathbf{x})$ as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})],$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))],$$

and will write the Gaussian process as

$$f(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')).$$

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

Consider a simple example shown in figure 1, where $m(\mathbf{x}) = \mathbf{x}$ and $k(\mathbf{x}, \mathbf{x}') = \mathbf{I}_p$.

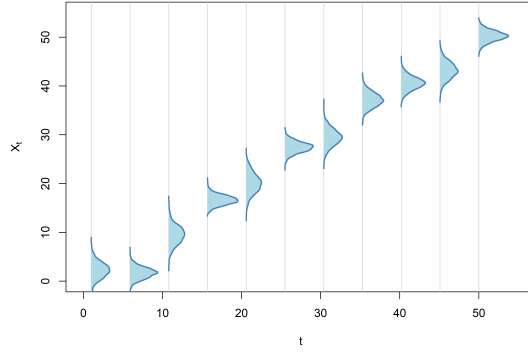


Figure 1: *An example of a Gaussian process on $t \in \{0, 10, \dots, 50\}$.*

Often people use

$$m(x_t) = 0$$

and

$$k(x_i, x_j) = \alpha^2 \exp\left(-\frac{1}{2} \left(\frac{|x_i - x_j|}{\rho}\right)^2\right) + \sigma \delta_{ij},$$

where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$, α , ρ , and σ are hyperparameters.

The latter is known as the squared-exponential kernel. Some kernel functions common in the literature are the following:

Air pollution example

Exploratory data analysis

| Variable | Name | Description | Unit |
|------------------|-----------------------|---|--------------------------|
| NO ₂ | Nitrogen dioxide | A harmful gas from vehicles and industry. | $\mu\text{g}/\text{m}^3$ |
| PM ₁₀ | Particulate matter 10 | Small inhalable dust particles. | $\mu\text{g}/\text{m}^3$ |
| SO ₂ | Sulphur dioxide | Mainly from burning fossil fuels. | $\mu\text{g}/\text{m}^3$ |
| 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.*

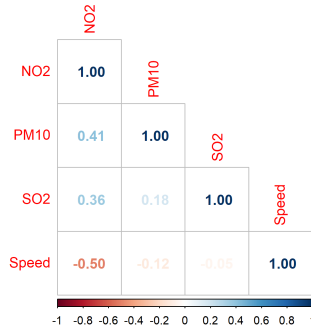
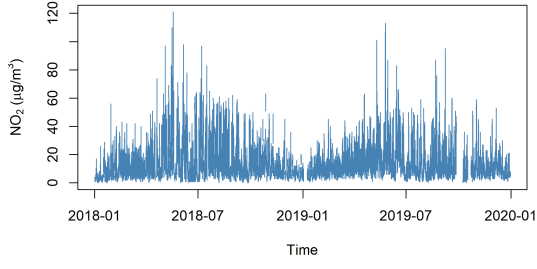


Figure 2: *Correlation plot of the air pollution dataset.*

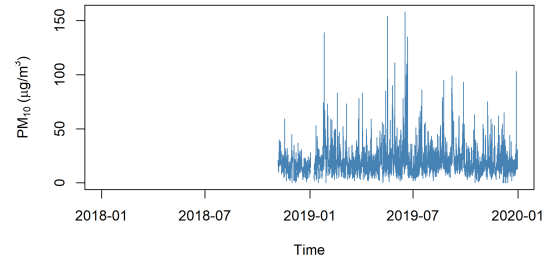
Our response variable NO₂ appears to be moderately positively correlated with PM₁₀ and SO₂, and moderately negatively correlated with Speed. These are not ideal explanatory variables since we typically would like them to be strongly correlated with the response variable. The explanatory variables are 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.

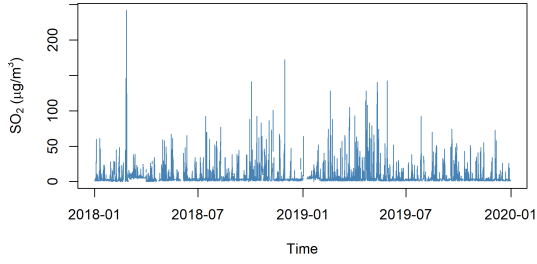
The prior distributions of the exponentiated quadratic kernel are the following:
 $\alpha \sim \text{half-normal}(0, 1)$, $\rho \sim \mathcal{IG}(20, 10)$, and $\sigma \sim \text{half-normal}(0, 1)$.



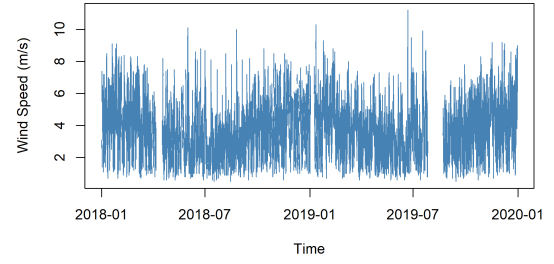
(a) Nitrogen dioxide



(b) Particulate matter 10

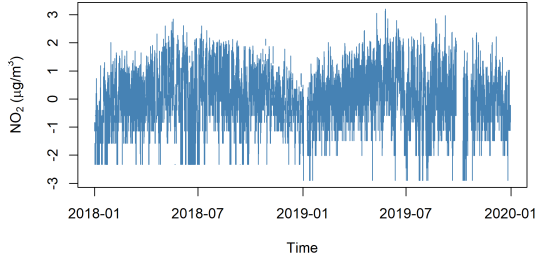


(c) Sulphur dioxide

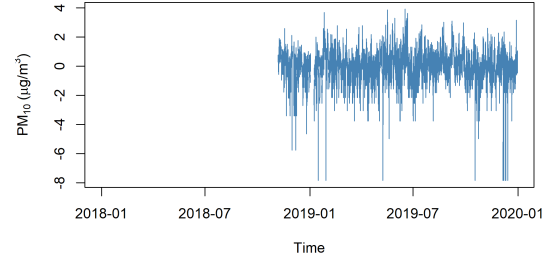


(d) Wind speed

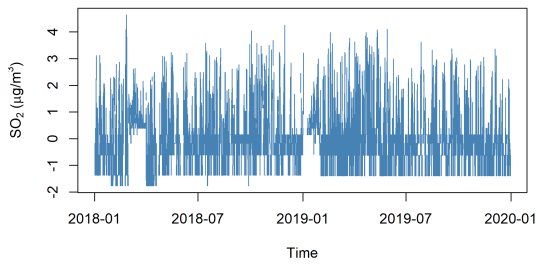
Figure 3: *Time-series plots of nitrogen dioxide and meteorological processes from 01/01/2018 to 31/12/2019 measured hourly. Similar cyclic patterns can be observed in the meteorological processes, and a weaker seasonality component can be noted in the yearly nitrogen dioxide processes.*



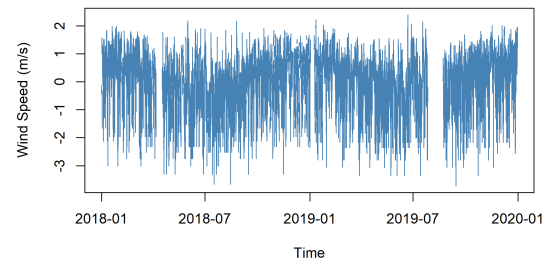
(a) Nitrogen dioxide



(b) Particulate matter 10



(c) Sulphur dioxide



(d) Wind speed

Figure 4: *Post-processed daily curves of the nitrogen dioxide and meteorological data, both standardized by their overall mean and standard deviation.*

Components of the time series

The response variable time-series x_t is not stationary because the mean and variance are changing with time. To achieve stationarity there is a need for detrending and a variance stabilizing transformation. In order to stabilize the variance we use Box-Cox transformations in the training set, $y_t = (x_t^\lambda - 1)/\lambda$. The Box-Cox transformations allow us to experiment with a wide variety of λ values. A good value of λ is one that makes the variation in the data constant through time (Neil Watson, 2024). The R package `forecast` was used to perform the Box-Cox transformation which yielded an optimal value of $\hat{\lambda} \approx 0$, suggesting a logarithmic transformation. Then, conducting first-order differencing at lag one to remove the trend $z_t = y_t - y_{t-1}$ for $t \in \{2, 3, \dots, p\}$. The time series z_t is now stationary. Thus, we only need to account for the mean and variance in our models.

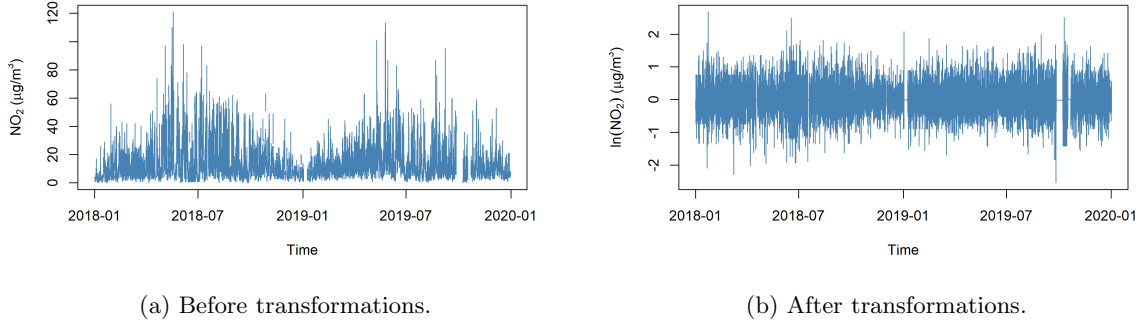


Figure 5: Comparison of NO_2 time series before and after transformations.

Results

| Model | RMSE | | | MAE | | |
|---------|--------------------------------|------|-------|--------------------------------|------|------|
| | Forecasts | | | Forecasts | | |
| | $(h \text{ day time horizon})$ | | | $(h \text{ day time horizon})$ | | |
| | 24 | 168 | 720 | 24 | 168 | 720 |
| Average | 8.61 | 7.16 | 6.62 | 8.54 | 6.48 | 5.73 |
| Naive | 1.35 | 5.58 | 7.10 | 1.00 | 3.59 | 4.80 |
| SNaive | 3.72 | 5.71 | 7.21 | 2.48 | 4.05 | 5.28 |
| Drift | 1.35 | 5.59 | 7.16 | 1.00 | 3.60 | 4.86 |
| ARIMA | 5.01 | 6.46 | 6.45 | 4.91 | 5.81 | 5.57 |
| GP | 6.07 | 8.73 | 10.72 | 5.47 | 7.29 | 9.14 |

Table 3: RMSE and MAE of forecasting models across horizons.

References

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