

Gaussian Processes for Time Series

Raphaela Azar and Sbonelo Gumede

July 25, 2025

Introduction

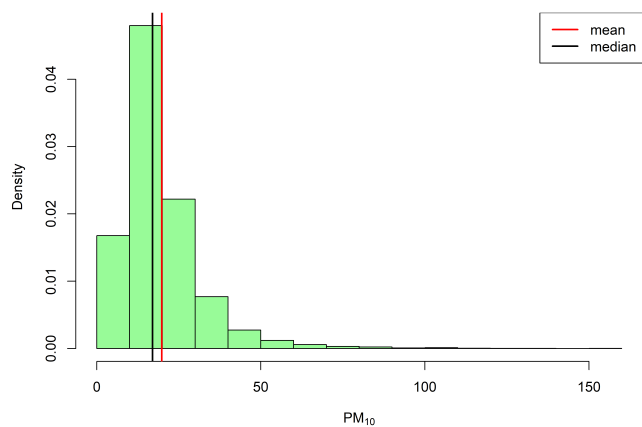
Variable	Feature Name	Feature Description	Unit
NO ₂	Nitrogen dioxide	A harmful gas from vehicles and industry.	$\mu\text{g}/\text{m}^3$
PM ₁₀	Particulate matter 10	Small inhalable dust particles.	$\mu\text{g}/\text{m}^3$
SO ₂	Sulphur dioxide	Mainly from burning fossil fuels.	$\mu\text{g}/\text{m}^3$
Direction	Wind direction	Indicates where the wind is coming from.	Degrees (0–360°)
Speed	Wind speed	How fast the wind is moving.	m/s

Table 1: Description of variables used in the analysis.

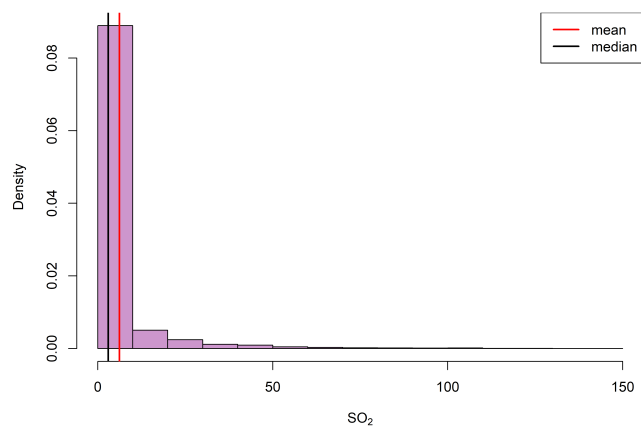
Exploratory Data Analysis



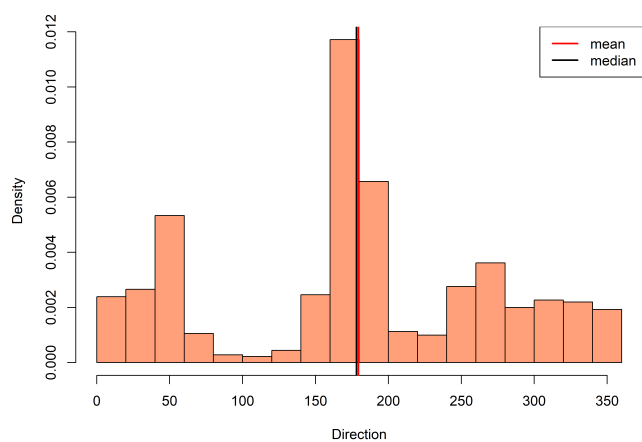
Figure 1: Correlation plot of the variables.



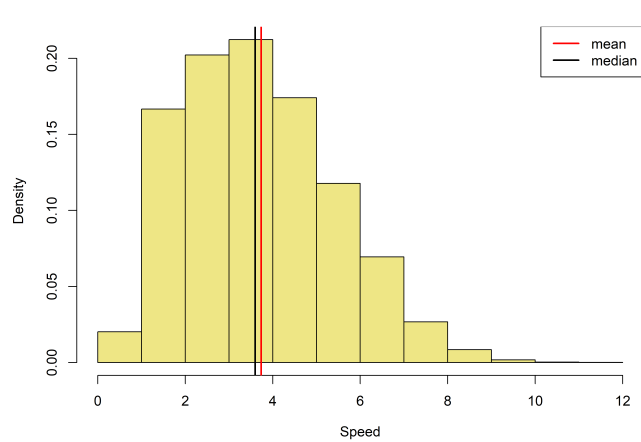
(a) Histogram of PM₁₀.



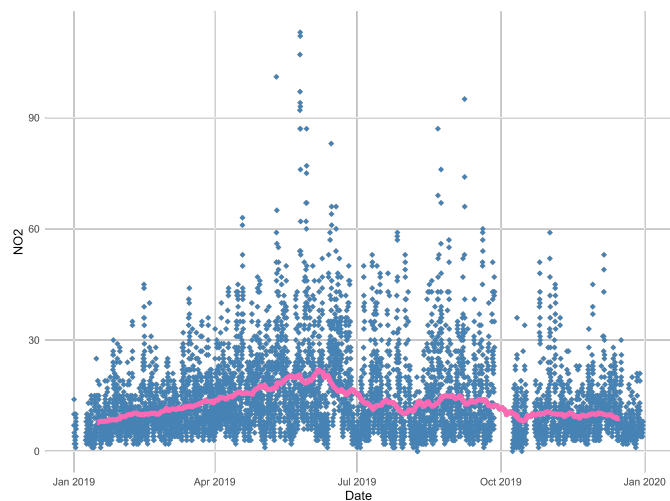
(b) Histogram of SO₂.



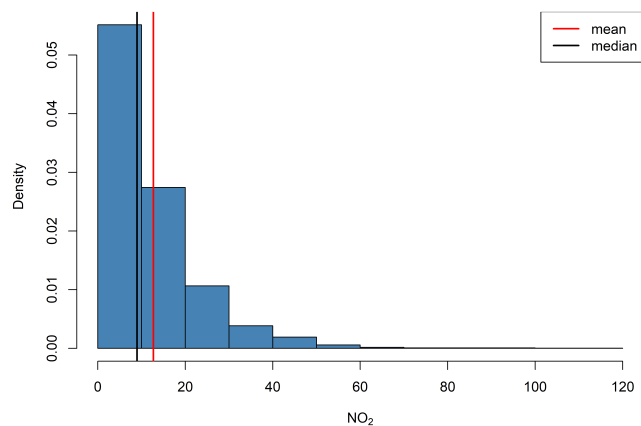
(c) Histogram of Direction.



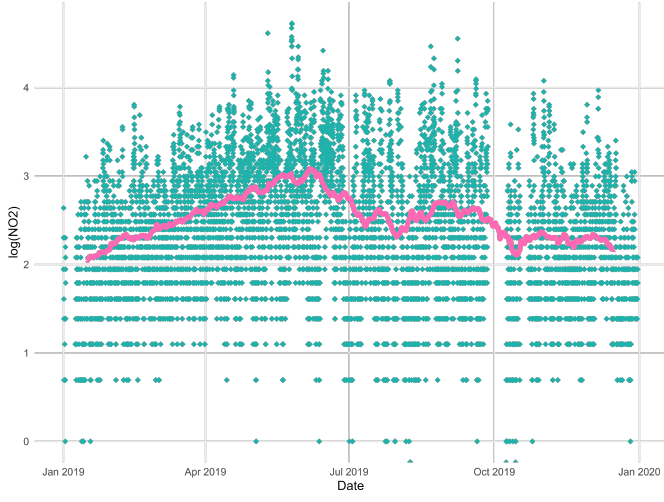
(d) Histogram of Speed.



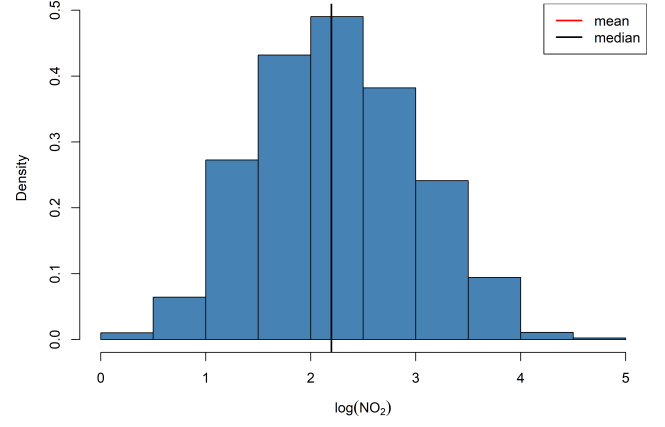
(a) Scatter plot of NO₂.



(b) Histogram of NO₂.



(a) Scatter plot of $\log(\text{NO}_2)$.



(b) Histogram of $\log(\text{NO}_2)$.

Let $Y := \text{NO}_2$ levels in the atmosphere. Notice that $Y \not\sim \mathcal{N}(\mu, \sigma)$ but $\log(Y) \sim \mathcal{N}(\mu, \sigma)$.

Our Gaussian process is of the form $h(x) \sim \mathcal{N}(f(x), g(x))$, $x \in \mathbb{R}$. The resultant covariance matrix of $g(x)$ must be positive semi-definite. Thus $g(x)$ must be chosen appropriately. The proposed model for the mean function is $f(x) = \alpha + x^T \beta$.

References

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