HEC-501 Risk Analytics

Final Report

Prof. Frédéric Aviolat

Group 5
Emily Sun Reed (370435)
Théodore Dominique Constantin Rosato-Rossi (378264)
Mathias Michel Häberli (376506)
Giulia Facchini (377474)
Maria Fernanda Cladera (23409527)

UNIL | Université de Lausanne

16.12.2024

Contents

1 Practical 1

1.1 Part 1 : Financial returns and normality

(a) The p-value of 0.3885, greater than 5% meaning we can't reject the null hypothesis that the data is non-stationary, and the weakly negative Dickey-Fuller statistic (-2.4484) suggest that Bitcoin prices are non-stationary, indicating trends or changing volatility over time rather than a constant mean and variance (as it is the case for stationary data). (See Appendix Figure 24 for the Time Series)

(b) The raw Bitcoin prices were non-stationary, with a changing mean and variance over time. After transforming the data into negative log returns, the series became stationary (p-value = 0.01, meaning we can reject the null hypothesis that the data is non-stationary), with mean and variance constant over time (see Appendix Figure 25). In finance, while raw prices are often non-stationary, negative log returns are used as they tend to be stationary, which makes them easier to model and forecast with various financial methods.

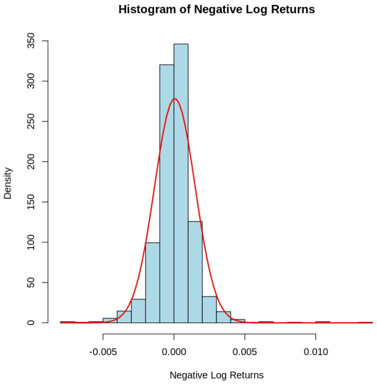


Figure 1: Histogram of Negative Log Returns

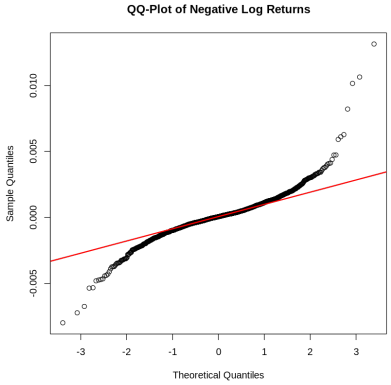


Figure 2: QQ-plot of Negative Log Returns

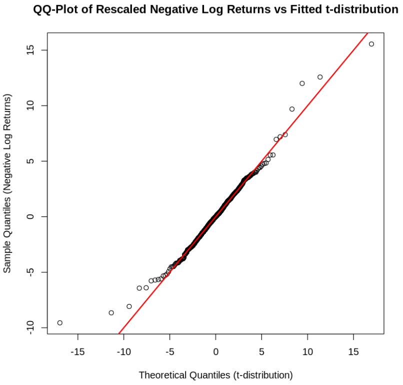


Figure 3: QQ-plot of Negative Log Returns vs fitted t-distribution

(c) Figure 1 displays the distribution of the negative log returns (blue bars) with the normal distribution (red line). The negative log returns roughly follow a bell curve. However, the tails are more spread than a typical normal distribution. Figure 2 compares the quantiles of negative log returns to those of a normal distribution. In an ideal normal distribution, points align along the red line. In the Figure 2, points deviate significantly at the extremes, indicating that the negative log returns are not perfectly normally distributed and exhibit heavy tails, with more extreme values than a normal distribution predicts. The Anderson-Darling test checks if data follows a normal distribution. The results (A = 26.277, p-value $< 2.2e{-}16$) indicate an extremely small p-value, which means we can reject the null hypothesis (H0: the negative log returns follow a normal distribution). Thus, the test strongly suggests that negative log returns are not normally distributed. Visually, Figure 1 and Figure 2 show some resemblance to a normal distribution, but Figure 2 reveals significant tail deviations. Statistically, the Anderson-Darling test confirms the negative log returns are not normally distributed (p-value < 0.01). The data exhibits fat tails and many extreme values, typical for financial returns with "heavy tails" and pronounced price movements.

(d) The t-distribution is well-suited for modeling data with heavy tails, a common characteristic of financial returns. It effectively captures extreme outliers and returns further from the mean compared to the normal distribution. The estimated parameters for the t-distribution are : Mean = 0.056, Standard deviation = 0.841, Degrees of freedom = 2.77. The df is a key parameter of the t-distribution. A lower df indicates heavier tails, meaning the distribution captures extreme values more effectively. In this case, with a df of approximately 2.77, the t-distribution effectively models the data's tails. The normal distribution assumes that the data is symmetrically distributed around the mean with tails that decay exponentially. However, financial returns, especially negative log returns, often exhibit fatter tails with more extreme values than what the normal distribution would suggest. The parameter estimates for the normal distribution are: Mean = 0.066, Standard deviation = 1.43. Figure 3 shows that the theoretical quantiles fit well the data points, indicating that the t-distribution is well-suited to model the negative log returns, capturing well the behavior of both the center and the tails of the data. In contrast, the QQ-plot for the normal distribution (Figure 2) shows deviations from the theoretical quantiles, particularly in the tails, suggesting that the normal distribution underestimates the frequency and magnitude of extreme returns, a common issue when modeling financial data with a normal distribution.

(e) By looking at Figure 4, in the normal distribution, the tails are thin and the probability of extreme values (very high or very low negative log returns) is low. In the t-distribution, especially when the df are low, the tails are thicker. This means there's a higher probability of extreme values occurring compared to the normal distribution. Because the t-distribution has heavier tails, we expect more extreme, unexpected events (eg.: large Bitcoin price swings) than in the normal distribution. In contrast, the normal distribution assumes extreme events are rare and underestimates their occurrence. Bitcoin price changes are known to be very volatile, with large and sudden price spikes or drops. The t-distribution, with its heavier tails, better predicts these extreme events, more common in financial markets than what the normal distribution would predict. The t-distribution suggests that we can expect more extreme events than the normal distribution, making it more suitable for capturing the volatility in Bitcoin prices.

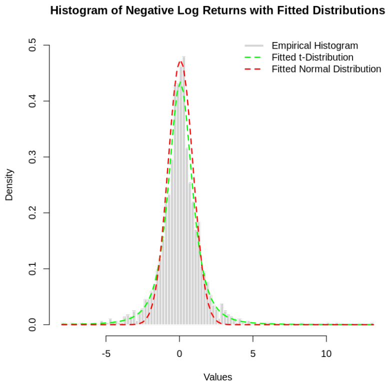


Figure 4: Histogram of Negative Log Returns with fitted distributions

1.2 Part 2: Financial time series, heteroscedasticity and the random walk hypothesis

(a) Based on the ACF plots (see Appendix Figure 26 and 27), the negative log returns are easier to model compared to the raw series as they are more stationary (mean and variance are more constant over time). Stationarity is a prerequisite for many time series models. Then, the ACF plot for the raw series (prices) shows strong autocorrelation, meaning the current value is highly influenced by previous values, making it harder to model. We can also see that as the lag increases, autocorrelation decreases, common in financial series. In contrast, the ACF for negative log returns shows much weaker autocorrelation. The few varying positive and negative peaks indicate that returns are mostly uncorrelated over time. Finally, negative log returns have more stable variance, which makes them better suited for models that assume homoscedasticity. Therefore, the negative log returns are easier to model, especially when applying models that assume stationarity and stable variance.

(b) For the raw Bitcoin series, the test statistic is very large (26873), and the p-value extremely small p-value ($< 2.2e{-}16$), which means we can reject the null hypothesis that there is no autocorrelation. This suggests that the raw price series has significant serial dependence, meaning that past values of the series influence future values, indicating non-randomness and autocorrelation over time. For the negative log returns, while the test statistic is smaller (33.356), the p-value is 0.03082, which means we can reject the null hypothesis that there is no autocorrelation at a 5% significance level but not at 1%. This implies that there is still some degree of serial dependence in the negative log returns, but it is much weaker and less significant compared to the raw Bitcoin series. (See Appendix Table 1 for Ljung-Box tests results)

(c) The ACF plot for negative log returns shows very little autocorrelation, suggesting that the series may behave like white noise. However, the PACF plot exhibits significant spikes at early lags, particularly at lags 1 and 2, indicating that an autoregressive component is likely necessary for modeling (see Appendix Figure 28 for ACF and PACF of Negative Log Returns). Based on these visual tools, several ARIMA models with autoregressive and moving average components were fitted (ARIMA(1,0,0), ARIMA(0,0,1), and ARIMA(1,0,1)). Among these, ARIMA(1,0,1) performed best, achieving the lowest AIC (-14762.32) and producing residuals with minimal autocorrelation, though the Ljung-Box test p-value (0.013) indicates some remaining serial correlation in the residuals (see Appendix Table 2 and Figure 29 for ARIMA(1,0,1) residuals). Using the `auto.arima()` function, an ARIMA(2,0,2) model with a non-zero mean is selected. This model includes two autoregressive (AR) terms and two moving average (MA) terms. Its coefficients suggest that AR(2) and MA(2) are the most significant terms in capturing the dynamics of the data. This model demonstrates a higher log-likelihood (7391.82) and lower AIC (-14771.65) and BIC (-14740.02) compared to the manually fitted models, indicating better overall fit. Additionally, the residual variance is very small, and the Ljung-Box test returns a p-value of 0.5727, far above the 5% significance level (see Appendix Table 4 and Figure 30 for ARIMA(2,0,2) residuals). This confirms the absence of significant residual autocorrelation, meaning the ARIMA(2,0,2) model effectively captures the temporal dependencies in the series (see Appendix Table 3 for ARIMA(2,0,2) results). In conclusion, while initial models based on ACF and PACF provided a good starting point, the `auto.arima()` selected ARIMA(2,0,2) model achieves superior performance, demonstrating the value of leveraging automated model selection for refining ARIMA parameters.

(d) By evaluating the residuals for GARCH(1,1) models with normal and t-distribution, we can see that both models capture well the dynamics of the negative log returns (see Appendix Table 5 for GARCH(1,1) results). In fact, there is almost no autocorrelation in the residuals. The residuals from both models are centered around zero with no visible patterns or clustering, suggesting that the models have adequately explained the volatility in the data. The difference between the two models lies in the distribution of the residuals (see Appendix Table 6 for GARCH(1,1) residuals). The GARCH(1,1) model with t-distribution better handles extreme values and heavier tails than with the normal distribution, and therefore provides a better fit for the negative log returns as heavy tails are common in financial series. In both cases, the Ljung-Box test indicates no significant autocorrelation of the negative log returns (p-value for GARCH(1,1) with a t-distribution = 0.3507, p-value for GARCH(1,1) with normal distribution = 0.3419).

(e) After fitting the ARIMA(2,0,2) model with non-zero mean on the negative log returns, and the GARCH(1,1) model on the residuals of the ARIMA(2,0,2) fit, the residual analysis indicates that the ARIMA-GARCH model performs well (see Appendix Table 8 for GARCH(1,1) coefficients). The residual time series plot shows that the residuals are centered around zero with no discernible patterns or clustering, suggesting that the ARIMA-GARCH model has effectively captured both the mean and volatility structure of the data. There are no persistent deviations from zero, indicating that the model has adequately removed serial dependence and volatility. The ACF plot of the residuals further supports this conclusion, as it reveals no significant autocorrelation across lags. All values fall within the confidence bounds, indicating that the residuals resemble white noise, meaning the model has successfully accounted for any autocorrelation and volatility clustering in the series. The histogram of the residuals shows a symmetric distribution centered around zero, which is characteristic of a well-fitted model. The residuals appear to be approximately normally distributed, with no visible skewness or excessive kurtosis, confirming that the model has effectively captured the data's volatility dynamics. (See Appendix Figure 31 for the ARIMA-GARCH(1,1) plots)

(f) The ARIMA(2,0,2) model effectively removes serial correlation and provides residuals that resemble white noise, as confirmed by the Ljung-Box test (p-value = 0.5727) (see Appendix Table 4). However, ARIMA assumes homoscedasticity and cannot capture volatility clustering, a key feature of financial data, such as instability in Bitcoin prices. The GARCH(1,1) models capture heteroscedasticity effectively, and therefore handle well Bictoin prices fluctuations. The t-distribution variant performs better than the normal distribution, as it accounts for heavy tails often observed in financial returns. Nonetheless, applying GARCH directly to the negative log returns can leave residual serial correlation, suggesting it does not fully capture temporal dependence between past and future returns. The ARIMA-GARCH model combines the advantages of both approaches. ARIMA removes linear dependencies, while GARCH models removes the remaining volatility clustering. When we look at the residuals, it confirms that this combined model captures both the mean and volatility dynamics effectively, with no significant autocorrelation and approximately normally distributed residuals. In summary, the ARIMA-GARCH model is the most suitable as it addresses both serial correlation and volatility clustering. Homoscedasticity is only assumed in the ARIMA(2,0,2) model, while the GARCH and ARIMA-GARCH models handle the heteroscedastic nature of financial data, making them better suited for this application.

1.3 Part 3: Dependence between time series

(a) The correlation coefficient of -0.0031 indicates almost no linear dependence between Bitcoin's and Ethereum's negative log returns. With a p-value of 0.905 (above the 5% significance level), we cannot reject the null hypothesis of independence, but this does not confirm independence. Correlation only measures linear relationships. A near-zero correlation and high p-value suggest a lack of linear dependence but do not rule out non-linear relationships. The test simply indicates no strong evidence of linear dependence without confirming overall independence. Non-linear dependencies between the series may still exist.

(b) At lag zero, the CCF is nearly zero, indicating no significant linear relationship between Bitcoin's and Ethereum's negative log returns at the same time. Across all lags, CCF values remain close to zero and mostly within the confidence interval, showing no significant cross-correlation or lead-lag relationship overall. At lag -5, however, a significant positive cross-correlation suggests that Ethereum's negative log returns lead Bitcoin's by 5 time steps. This indicates Ethereum's performance may predict Bitcoin's, with increases or decreases in Ethereum's returns followed by similar movements in Bitcoin's. Despite this, the lack of significant correlations at other lags supports the conclusion that the two series are largely independent. (See Appendix Figure 32 for the CCF plot)

(c) A Granger causality test was conducted to assess whether one cryptocurrency's past performance predicts the other's future movements. The results show a one-way relationship: Bitcoin's past negative log returns Granger-cause Ethereum's future returns, with a highly significant p-value ($< 2.2e-16$). This indicates that Bitcoin strongly influences Ethereum, suggesting Bitcoin's dominant role in the cryptocurrency market. In contrast, Ethereum's past returns do not Granger-cause Bitcoin's future returns (non significant p-value of 0.7132), implying that Ethereum lacks predictive power over Bitcoin. In summary, Bitcoin serves as a leading indicator for Ethereum, reinforcing its influence on market dynamics, while Ethereum's movements have little impact on Bitcoin.

(d) The Granger causality results suggest distinct outcomes during significant market events for Bitcoin or Ethereum. 1) A sharp drop in Bitcoin is likely to cause a similar decline in Ethereum, reflecting Bitcoin's strong influence on Ethereum. Major shifts in Bitcoin often ripple through other cryptocurrencies, with Ethereum being particularly affected. 2) Conversely, a sudden drop in Ethereum is unlikely to impact Bitcoin, as Ethereum's performance does not predict Bitcoin's movements, highlighting Bitcoin's independence. In summary, Bitcoin drives market dynamics, influencing Ethereum, while Ethereum lacks the same impact on Bitcoin.

2 Practical 2

2.1 Part 1: Block maxima approach

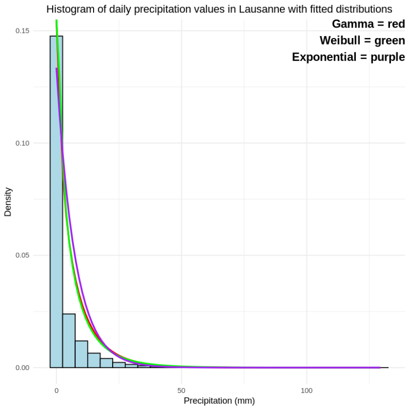


Figure 5: Histogram of daily precipitation values in Lausanne

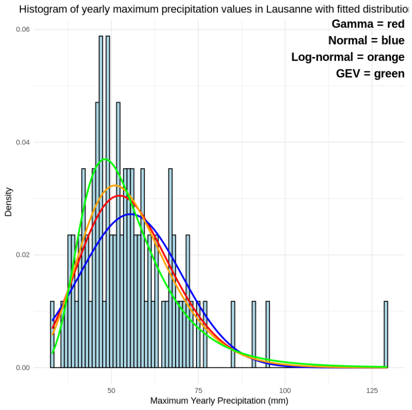


Figure 6: Histogram of yearly maximum precipitation values in Lausanne

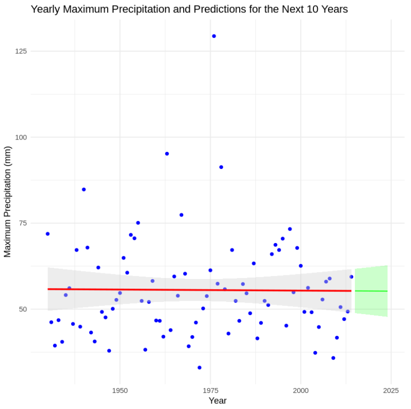


Figure 7: Yearly maximum precipitation and predictions in the next 10 years

(a) The histogram (Figure 5) shows a right-skewed distribution, with the majority of values concentrated around 0 mm. This means that most days record little to no precipitation, and only a few days record high precipitation. To model this data, the Exponential, Gamma, and Weibull distributions all fit the data well. In order to assess which distribution fits the data the best, we used different tests (see Appendix Table 9). In all cases, the Weibull distribution performs the best. In fact, the Weibull distribution is often used to model meteorological data.

(b) Based on the histogram (Figure 6), the GEV distribution seems to best fit the maximum yearly precipitations. This is also the case when looking at the AIC values obtained. Lower AIC values indicate a better fit for the model, in this case the GEV distribution has the lowest AIC value (-660.9433).

(c) This is not a reasonable approach since there is a lot of variability in the extreme values which cannot be captured by a linear model. The adjusted R-squared (-0.01194) and F-statistic (0.008718) with a very high p-value (0.9258), indicate that the model doesn't fit the data well. The projections (Figure 7) are linearly predicted values, which fail to account for the peaks present in historical data and that are very likely to be present in the next 10 years.

(d) The model with constant parameters gets lower AIC and BIC values (AIC = 672.94 and BIC = 680.27) compared to the model with time-varying location parameter (AIC = 674.89 and BIC = 684.66). Lower AIC and BIC values indicate a better fit to the data. Therefore, the model with constant parameters is recommended.

(e) The diagnostic plots (see Appendix Figure 33) for the GEV model suggest a generally good fit to the data. The Q-Q plot indicates a strong alignment between observed and theoretical quantiles across the central portion of the data, with some deviation in the upper tail, meaning the most extreme precipitation events are underestimated. The P-P plot shows strong similarity between observed and theoretical cumulative probabilities, confirming that the GEV model captures the overall data distribution effectively. The sharp decline in return levels as return periods increase in the return level plot might indicate that the model's parameters do not fully capture the behavior of extreme values. Finally, the residual histogram reveals that residuals are approximately centered around zero, confirming that the GEV model captures the central tendency of the data well, less true for extreme values. Overall, the GEV model is a strong fit for most of the data, but has its limitations in capturing extreme values.

(f) The predicted 10-year return level is 73.60 mm, meaning that a precipitation exceeding this threshold is expected to occur once every 10 years. Figure 8 illustrates the historical yearly maximum precipitation alongside the predicted 10-year return level.

(g) The predicted return levels for the 10, 20, 50, and 85-year periods are 73.60 mm, 82.52 mm, 94.89 mm, and 102.44 mm, respectively. Historical exceedances of these levels were observed 6, 4, 2, and 1 times, aligning well with theoretical expectations. For instance, approximately 10% of years are expected to exceed the 10-year return level, and the observed 6 exceedances out of 85 years confirm the GEV model's reliability. In contrast, the linear model from part (c) fails to capture the probabilistic nature of extreme events, rendering it unsuitable for return level predictions.

The GEV model provides a robust framework for analyzing extremes, with its return levels and exceedance frequencies validating its ability to model rare precipitation events effectively.

(h) The return period for a precipitation of 100 mm is 71.78 years. This means that a precipitation event of 100 mm or greater is expected to occur, on average, once every 71.78 years. The value aligns with the tail behavior modeled by the GEV distribution, where higher precipitation levels correspond to longer return periods.

(i) The probability of exceeding 150 mm of precipitation on at least one day in the next year is approximately 20.90%, meaning such extreme precipitation is expected to occur on roughly 1 in 5 days over the year. Given that 150 mm significantly exceeds the predicted return levels for even longer return periods (e.g., the 85-year return level is 102.44 mm), it is consistent that such events would remain uncommon within a single year.

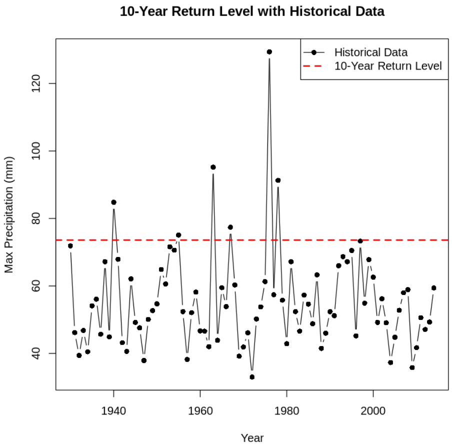


Figure 8: Historical yearly maximum precipitation and Predicted 10-year return level

2.2 Part 2: Peaks-over-threshold approach

(a) The time series plot (Figure 9) shows daily precipitation levels, with most days having minimal or no rainfall (Median: 0 mm, Mean: 3.2 mm), and occasional extremes values up to 129.4 mm.

(b) Using the POT approach, we set the 95th percentile (17.7 mm) as the threshold, considering the top 5% of daily precipitation as extreme. The Mean Residual Life Plot (Figure 10) shows the mean excess above the threshold exhibits a stable linear trend, which aligns with the assumptions of the Generalized Pareto Distribution (GPD). Beyond 20 mm, the mean excess becomes irregular, indicating data sparsity and suggesting a higher threshold may not be optimal. Figure 11 shows daily precipitations with values above the 20 mm threshold.

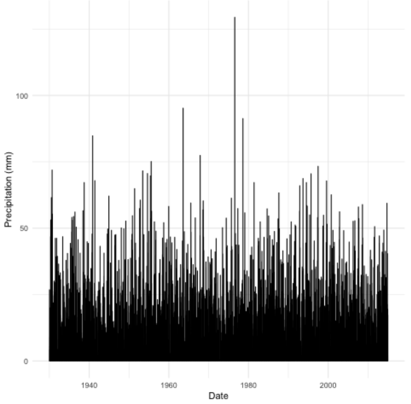


Figure 9: Daily precipitations over time

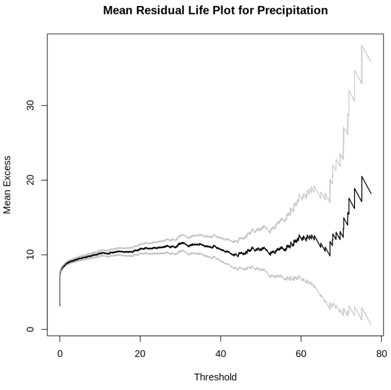


Figure 10: Mean Residual Life Plot for daily precipitations in Lausanne

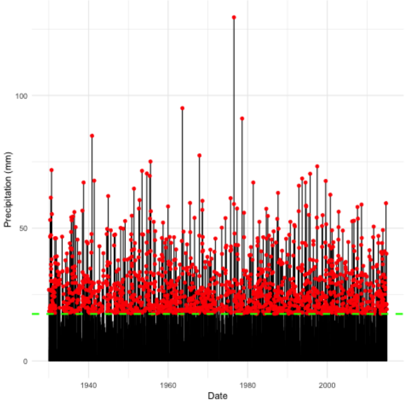


Figure 11: Daily precipitations with values above threshold

(c) The GPD model, fitted to precipitation data exceeding the 17.7 mm threshold, shows a strong alignment with observed extremes. The Return Level Plot (see Appendix Figure 34) confirms that the model accurately captures return levels across most periods, with fewer deviations for very high return periods due to data limitations. Diagnostic plots (see Appendix Figure 35) support the validity of the threshold choice, showing stable residuals and consistent parameter estimates. Overall, the model is effective for analyzing extreme precipitation events.

(d) The return levels for 10, 20, 50, and 85-year return periods are 10.82, 17.64, 26.99, and 32.59 mm respectively. These return levels represent the expected precipitation levels associated with extreme events over the specified timescales.

(e) Using the same fitted Generalized Pareto Distribution (GPD) model, we find a return period of 21'966 for a precipitation level of 100 mm, meaning that such an extreme event is expected once every 21'966 days (60.18 years).

(f) The probability of a day exceeding 150 mm of precipitation in the next year is 0.06% (return period of 639'872.6 days or 1753 years).

(g) After our analysis, we consider that the Peaks-Over-Threshold (POT) approach offers significant advantages for our extreme precipitation events compared to the Block Maxima method. By selecting a threshold of 17.7 mm (95th percentile) and focusing on all exceedances, the POT method captured 152 extreme events, providing a richer dataset for modeling extremes. In contrast, the Block Maxima approach analyzes only yearly maxima, reducing the dataset to a single value per year (approx. 25 data points) and discarding intermediate extremes. The analysis shows that the GPD fit in the POT approach provided greater parameter stability and a well-fitting return level plot. For example, it estimated a 100 mm precipitation return period of 60.18 years and a 0.06% probability of exceeding 150 mm annually. In contrast, the GEV model in the Block Maxima method fit well overall but shows deviations in the tails, limiting its ability to capture extreme events. It estimated a 10-year return level of 73.60 mm, compared to 26.99 mm from the POT model, reflecting methodological differences. While GEV summarizes annual extremes effectively, it lacks POT's granularity for capturing multiple rare events in a year.

2.3 Part 3: Clustering and Seasonal Variations

(a) Figure 12 reveals clear seasonal patterns, with peaks in summer and troughs in winter. During summer months (June to September), temperatures are consistently higher, with July and August as warmest months. Figure 13 highlights extreme temperatures in July and more variability in June and September.

(b) We chose the 95th percentile as the threshold (we consider the top 5% values as extreme) for extreme temperatures, which corresponds to 24.94°C. The extremal index is $\theta = 0.26$. As it is closer to 0, it means that extreme events are not independent but occur in clusters. There is a 73.87% probability of clustering, meaning that when one extreme event occurs, this is the probability that an extreme event happens the day after. Additionally, the probability that an extreme event today is followed by another extreme event tomorrow is 26.13%.

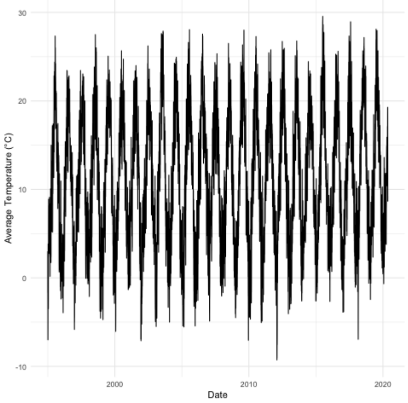


Figure 12: Time series of daily temperatures in Geneva

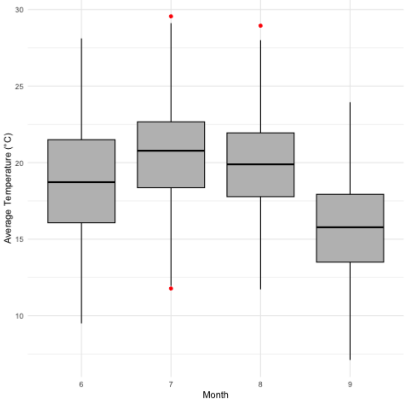


Figure 13: Boxplot of Geneva daily temperatures for summer months

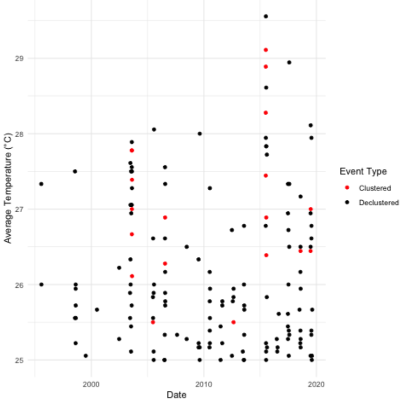


Figure 14: Declustered Geneva summer daily temperatures

(c) We declustered the summer temperatures (June–September) using a threshold of 24.94°C (95th percentile) (Figure 14). Out of 152 exceedances above the threshold, 67 independent extreme events were identified.

(d) For the model with raw data we obtain an AIC of 370.96. For the model with declustered data, the AIC is 303.01, indicating that this model provides a better fit. The 10-year return level for the raw data model is 25.78°C, while for the declustered data model it is 25.45°C. While the 10-year return levels are similar, the lower AIC value for the declustered model demonstrates its superior fit. This shows the importance of declustering, as it accounts for the dependence between extreme events and provides a more theoretically robust framework for analyzing extremes.

We plotted diagnostic plots to check the fit of the GEV model with constant parameters. The Q-Q plot indicates that the GEV model fits well across most quantiles, though minor deviations in the tails suggest slight discrepancies in capturing extreme values. The P-P plot further supports the fit, showing that the GEV model effectively captures the cumulative distribution of the data. On the return level plot, we can see that the observed points closely follow the theoretical predictions, indicating that the model provides reliable estimates for rare and extreme events. Finally, the residual histogram shows residuals centered around zero, with no significant skewness or irregularities. This symmetry suggests that the model's predictions align well with the observed data, supporting the overall validity of the GEV fit.

3 Practical 3

3.1 New Dehli air pollution and PM2.5 concentrations

New Delhi is one of the world's most polluted cities, with PM2.5 levels often exceeding the safe limits of 35 µg/m^3 set by the World Health Organization (World Health Organization, 2021). PM2.5 refers to tiny particles in the air, with a diameter of less than 2.5 micrometers, that can penetrate deep into the lungs and bloodstream, causing serious health issues such as respiratory and cardiovascular diseases. For context, PM2.5 concentrations below 12 µg/m^3, typically observed in cities like Zurich or Vancouver, are considered "good" and pose minimal health risks. In contrast, levels in New Delhi frequently surpass 200 µg/m^3, particularly during winter months, and can even reach "hazardous" levels above 300 µg/m^3, comparable to conditions during severe smog episodes in Beijing or Lahore. Such high concentrations pose a severe health risk to the entire population. The primary sources of PM2.5 in New Delhi include vehicular emissions, industrial activities, construction dust, and crop stubble burning in surrounding regions (Guttikunda and Gurjar, 2012). Seasonal factors, such as reduced wind speed and temperature inversions during winter, exacerbate pollution by trapping pollutants closer to the ground (Cichowicz et al., 2017). This practical focuses on analyzing PM2.5 levels in New Delhi to better understand these sources and their contributions to severe air pollution.

3.2 Block maxima analysis

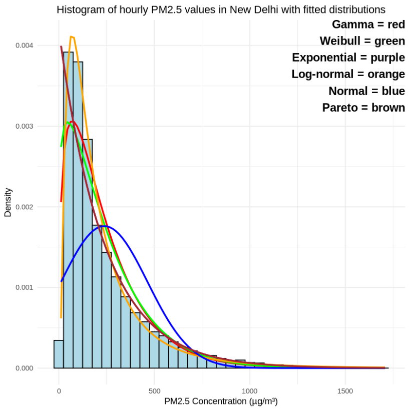


Figure 15: Hourly PM2.5 concentrations in New Dehli

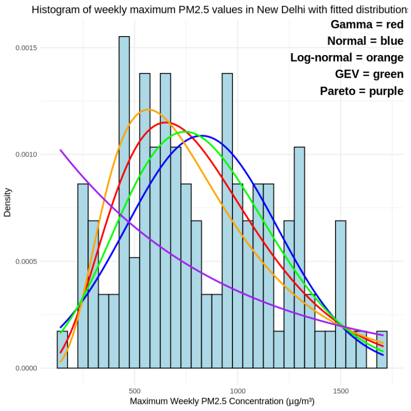


Figure 16: Weekly maximum PM2.5 concentrations in New Dehli

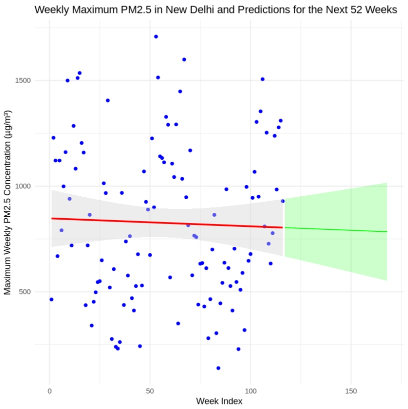


Figure 17: Weekly maximum PM2.5 concentrations and predictions for 52 weeks

Figure 15 shows a right-skewed distribution of PM2.5 concentrations in New Delhi, with most values under 200 $\mu g/\text{m}^3$, indicating high pollution for most observations and rare extremes. Visual analysis and goodness-of-fit tests (see Appendix 10) identify the Log-normal distribution as the best fit.

Figure 16 suggests the Log-normal and GEV distributions fit the weekly maximum PM2.5 concentrations in New Dehli the best. AIC values confirm this, with the Log-normal at 240403.5 and the GEV at -1688.5, indicating the GEV as the best model, also due to its ability to better capture extremes (see Appendix 11).

Figure 17 shows that a linear model poorly predicts future 52 weeks PM2.5 concentrations, as reflected by the adjusted R-squared (-0.007578), high p-value (0.714) of the F-statistic (0.1351), and a large residual standard error (369.8). The linear model fails to capture historical peaks and troughs, resulting in overly simplified predictions that do not represent future extreme fluctuations. There is a need for a model addressing non-linear trends and extremes.

Therefore, we fit a GEV model with constant and time-varying location parameters. The constant model yields a lower AIC (1700.549) and BIC (1708.809) compared to the time-varying location model (AIC = 1702.511, BIC = 1713.525). Since the time-varying location parameter does not improve the model's fit, the simpler GEV model with constant parameters is recommended for analyzing historical weekly maximum PM2.5 values.

We plotted diagnostic plots (see Figure Appendix 36) to check the fit of the GEV model with constant parameters. The Q-Q plot indicates that the GEV model fits well across most quantiles, though minor deviations in the tails suggest slight discrepancies in capturing extreme values. The P-P plot further supports the fit, showing that the GEV model effectively captures the cumulative distribution of the data. On the return level plot, we can see that the observed points closely follow the theoretical predictions, indicating that the model provides reliable estimates for rare and extreme events. Finally, the residual histogram shows residuals centered around zero, with no significant skewness or irregularities. This symmetry suggests that the model's predictions align well with the observed data, supporting the overall validity of the GEV fit.

The predicted 52-week return level is 1627.26 µg/m^3, indicating that PM2.5 concentrations exceeding this threshold are expected to occur less than once per year. Figure 18 illustrates the historical weekly maximum PM2.5 concentrations alongside the predicted 52-week return level, with most observed data points falling below this threshold.

The predicted return levels for the 52, 104 (2 years), 260 (5 years), and 442-week (8.5 years) periods are 1627.53 µg/m^3, 1735.15 µg/m^3, 1857.40 µg/m^3, and 1919.22 µg/m^3, respectively. Historical exceedances (1, 0, 0, and 0 times) align with theoretical expectations, confirming the GEV model's reliability in capturing extreme pollution events (eg.: 1 exceedance is expected for the 52-week return level, and there is 1 observed exceedance). Unlike simpler linear models, the GEV model effectively accounts for the variability of extremes, providing robust predictions that align with observed data, such as return levels and exceedance frequencies, demonstrating its ability to effectively model and predict rare air pollution events.

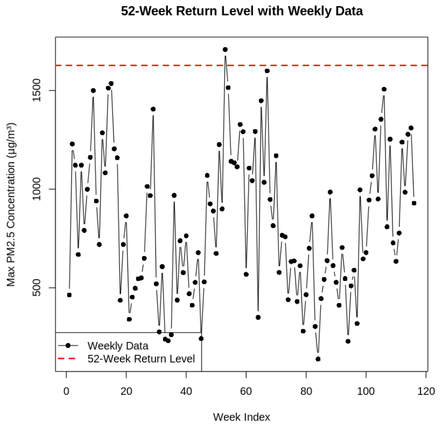


Figure 18: Weekly maximum PM2.5 concentrations & 52-week return level

The computed return period for a PM2.5 concentration of 2000 µg/m^3 is approximately 966.47 weeks. This indicates that an air pollution event with PM2.5 levels reaching or exceeding 2000 µg/m^3 is expected to occur, on average, once every 966.47 weeks, or roughly 18.6 years. This result highlights the rarity of such extreme pollution levels and aligns with the tail behavior captured by the GEV distribution, where extreme PM2.5 concentrations correspond to longer return periods.

The probability of exceeding 2500 µg/m^3 of PM2.5 on at least one day in the next year is about 0.00000022%, reflecting the rarity of such extreme events modeled by the GEV distribution. This indicates the near impossibility of such air pollution levels.

3.3 Analysis of Seasonalities with Peak-over-threshold approach

This analysis examines the seasonal patterns of PM2.5 This analysis investigates the seasonal patterns of PM2.5 concentrations using the Peak-over-Threshold (POT) approach. As observed in Figure 19, the time series plot of the daily maximum PM2.5 concentrations shows substantial fluctuations, with occasional extreme pollution spikes. These variations are further emphasized by Figure 20, the histogram, which reveals a right-skewed distribution, where most observations fall within moderate ranges below 500 µg/m^3, with only a small portion exceeding these thresholds. The seasonal distribution of the daily maximum PM2.5 concentrations, as shown in Figure 21, highlights the differences in extremes across the Monsoon, Summer, and Winter seasons. This overview suggests that while most daily concentrations remain at moderate to elevated levels, extreme pollution events surpassing 1000 µg/m^3 are rare but significant, particularly during certain seasons. Such extremes, though infrequent, pose serious health risks and warrant close attention. Based on these observations, we decided to segment the dataset by season and assign specific thresholds for each season, reflecting their distinct pollution behaviors.

The analysis identifies extreme days based on season-specific thresholds, determined through the Mean Residual Life (MRL) plot technique (see Appendix Figure 37). Based on this, we set the thresholds to 500 µg/m^3 for Winter, 200 µg/m^3 for Summer, and 150 µg/m^3 for Monsoon. The statistics present that Winter has 185 extreme days, Summer has 168, and Monsoon has 159. These results highlight Winter as the season with the highest number of extreme events. To further understand the nature of these extremes, the extremal index was calculated, revealing that Winter (0.51) has more clustered extremes, suggesting longer periods of high pollution, while Summer (0.35) and Monsoon (0.33) exhibit more sporadic extremes.

We applied the declustered Peak-Over-Threshold (POT) analysis to get an understanding of independent extreme PM2.5 events by removing clustered data points. The AIC values for the Generalized Pareto Distribution (GPD) fits reveal notable differences between raw and declustered analyses. In the raw analysis, significantly higher AIC values—Winter (2048.54), Monsoon (1780.06), and Summer (1814.10)—indicate poorer model fits, reflecting the confounding effect of clustered events. By contrast, the declustered analysis yields much lower AIC values—Winter (18.03), Monsoon (18.03), and Summer (18.12)—demonstrating a more accurate representation of independent extreme events. The marginally higher AIC for Summer in both analyses suggests greater variability and unpredictability of its extremes, while Winter and Monsoon exhibit more stable fits, consistent with their stronger clustering tendencies.

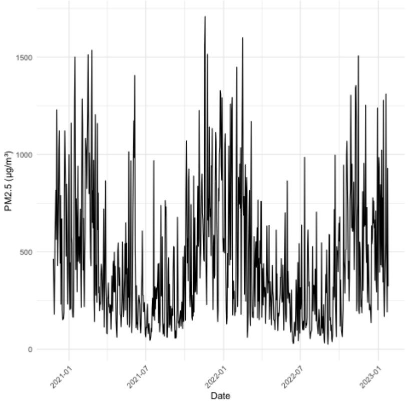


Figure 19: Time series of daily maximum PM2.5 concentrations in New Delhi

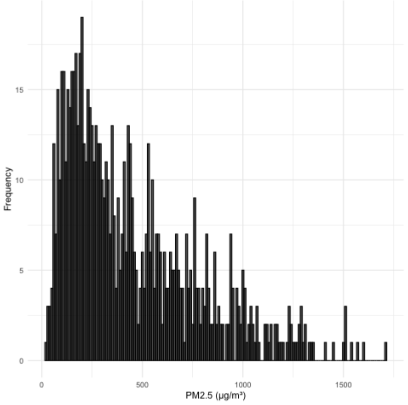


Figure 20: Histogram of daily maximum PM2.5 concentrations in New Delhi

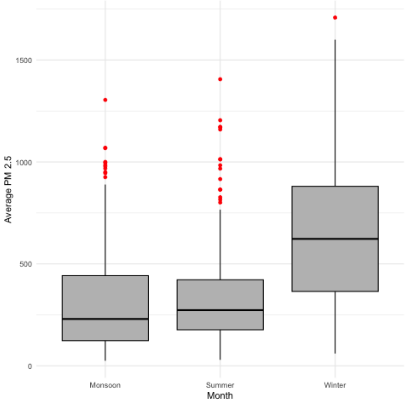


Figure 21: Seasonal box plot of daily maximum PM2.5 concentrations in New Delhi (Monsoon, Summer, and Winter)

The return level analysis quantifies the expected PM2.5 concentrations corresponding to specific return periods, offering insights into the severity of extreme pollution events in Winter. For the raw model, the predicted return levels are 1669.97 µg/m^3, 1722.44 µg/m^3, and 1776.92 µg/m^3 for return periods of 1, 2, and 5 years, respectively. The declustered model produces slightly higher return levels for shorter return periods—1706.14 µg/m^3 for 1 year and 1707.66 µg/m^3 for 2 years—while stabilizing for longer periods at 1708.58 µg/m^3 for 5 years. These results indicate that declustering mitigates the influence of event clustering on extreme predictions, presenting more conservative estimates for shorter return periods and greater consistency.

Overall, the Block Maxima and Peak-Over-Threshold (POT) approaches provide complementary insights into extreme PM2.5 events. The Block Maxima method, using GEV, gives a return level of 1627.26 µg/m^3 (1 year) but may underestimate extremes due to data aggregation. POT, focusing on threshold exceedances, estimates higher return levels (1706.14 µg/m^3 for 1 year), capturing rare, high-impact events more effectively. POT's declustering ensures independence, offering greater precision for rare events, while Block Maxima provides a broader overview.

3.4 Dependencies and Factors of PM2.5

In this section, we explore the causes of the significant PM2.5 emission peaks in New Delhi, focusing first on emission sources and then on environmental factors contributing to these peaks. Studies have shown that Nitric Oxide (NO) and Carbon Monoxide (CO) emissions are directly linked to combustion engines in the numerous vehicles in New Delhi (Nagpure et al., 2015). Additionally, Pant et al., 2015 found that traffic contributes to at least 20% of PM2.5 emissions. Using NO and CO as traffic proxies, we examine their effects on PM2.5 levels and vice versa to identify any dependencies. As previously discussed, PM2.5 emissions in our data exhibit seasonal patterns, with higher peaks during winter compared to summer in New Delhi. Cichowicz et al., 2017 not only confirmed these seasonal trends but also identified wind speed differences between winter and summer as a key factor influencing PM2.5 levels. This section also investigates these dependencies.

3.4.1 Sources of PM2.5 Variations: Nitric Oxide (NO) & Carbon Monoxide (CO) as Proxies

Correlation between PM2.5 and NO, CO: We first conducted a correlation test using hourly emission data for PM2.5, NO, and CO, available in our primary dataset for the period 2021-01 to 2023-01. For NO and PM2.5, we found a significant positive correlation of 0.816 (95% confidence interval: 0.812 to 0.821, t = 193.73, df = 18,774, $p < 2.2e-16$). This correlation is reasonable, as NO undergoes atmospheric chemical reactions that contribute to PM2.5 formation. (See Appendix Figures 40 and 41) Similarly, the correlation between CO and PM2.5 is significantly positive and even stronger, at 0.937 (95% confidence interval: 0.935 to 0.939, t = 366.98, df = 18,774, $p < 2.2e-16$). This stronger effect may be due to CO being emitted alongside various types of PM2.5 during combustion processes such as vehicle emissions, heating, or wood burning.

Cross-Correlation Function (CCF) Analysis: The Cross-Correlation Function (CCF) shows a strong relationship between NO, CO, and PM2.5 at lag 0 (NO: 0.816, CO: 0.937), indicating significant simultaneous correlations and supporting the hypothesis that these gases directly contribute to PM2.5 formation. (See Appendix Figures 42 and 43) Negative lags (e.g., -5: NO: 0.445, CO: 0.550) reveal that NO and CO emissions from prior hours strongly influence PM2.5 levels, aligning with their role as precursors. Positive lags also show significant but weaker correlations (e.g., +5: NO: 0.355, CO: 0.459), suggesting co-emission during combustion. Overall, the results confirm the dominant causal role of NO and CO in PM2.5 variations.

NO and CO as predictors of PM2.5 - Granger Causality Analysis: The Granger Causality tests reveal clear relationships between NO, CO, and PM2.5. For NO predicting PM2.5, the test yields a highly significant result (F = 470.95, $p < 2.2e - 16$), confirming that NO emissions strongly explain PM2.5 variations. Conversely, PM2.5 predicting NO shows a much weaker, though still significant, relationship (F = 86.32, $p < 2.2e - 16$). Similarly, CO significantly predicts PM2.5 (F = 226.24, $p < 2.2e - 16$), while PM2.5 predicting CO is comparatively weaker (F = 86.57, $p < 2.2e - 16$). These results underscore that both NO and CO are stronger predictors of PM2.5 than vice versa, reinforcing their role as primary contributors to PM2.5 formation. In conclusion, the dependencies observed through these tests strongly suggest that traffic-related emissions, a major source of NO and CO in New Delhi, play a crucial role in driving PM2.5 levels. This highlights the significant impact of vehicular emissions on air pollution in the region.

3.4.2 Wind Intensity as a predictor of Seasonal Trends

Correlation between PM2.5 and Average Wind Speed (mph): In this section, we examined the effects of daily average wind speed on PM2.5 levels. To do so, we conducted a correlation test using daily average PM2.5 data, as external wind speed data was only available in daily averages (Weather Underground, n.d.). The correlation test results showed a significant negative relationship between daily average wind speed and PM2.5 levels (correlation: -0.634, t = -22.989, df = 786, $p < 2.2e - 16$, 95% confidence interval: [-0.674, -0.590]). This suggests that higher wind speeds are associated with lower PM2.5 concentrations, likely due to the dispersal of particulate matter. These findings support the hypothesis that stronger winds in winter help disperse PM2.5, reducing its concentration. (See Figure 22 and Appendix Figure 44)

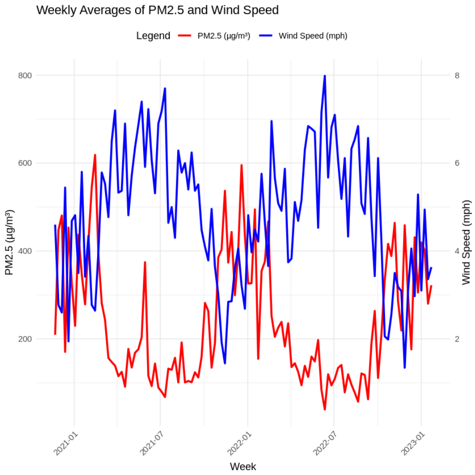


Figure 22: Weekly averages of PM2.5 and Wind Speed

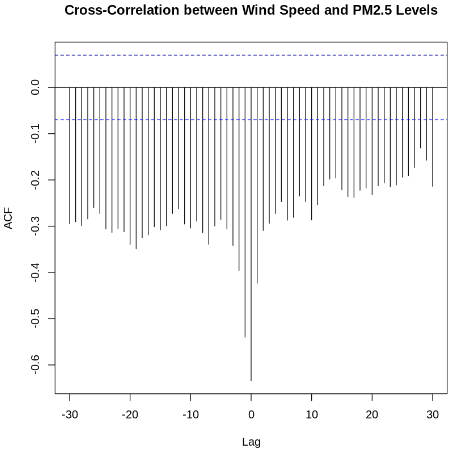


Figure 23: Cross-Correlation between Wind Speed and PM2.5 Levels

Cross-Correlation Function (CCF) Analysis: The cross-correlation between wind speed and PM2.5 at lag 0 revealed a significant negative relationship (-0.634), indicating that higher wind speeds within the same hour strongly correlate with lower PM2.5 concentrations (see Figure 23). Negative lags, such as -1 (-0.540) and -5 (-0.286), remain significant, suggesting that higher wind speeds earlier in the day effectively reduce PM2.5 levels. This supports the theory that wind disperses particulate matter. In contrast, positive lags show weaker correlations, such as +1 (-0.423) and +5 (-0.247), which decrease over time. These findings reinforce that wind speed influences PM2.5 levels more significantly than the reverse, supporting the hypothesis of wind as a key factor in PM2.5 dispersion.

Daily Average Wind Speed as a Predictor of PM2.5 Seasonality - Granger Causality Analysis: The Granger Causality test provided clear results: wind speed significantly predicts PM2.5 levels (p = 0.0065), confirming its dispersive effect on particulate matter. Conversely, PM2.5 does not predict wind speed (p = 0.1435), indicating no reverse causality. These findings align with the understanding that higher wind speeds help disperse PM2.5 concentrations, reducing pollution levels (Cichowicz et al., 2017), while PM2.5 emissions have no clear direct influence on wind behavior. This emphasizes wind's critical role in mitigating air pollution.

4 Appendix

Part 1.1 a)

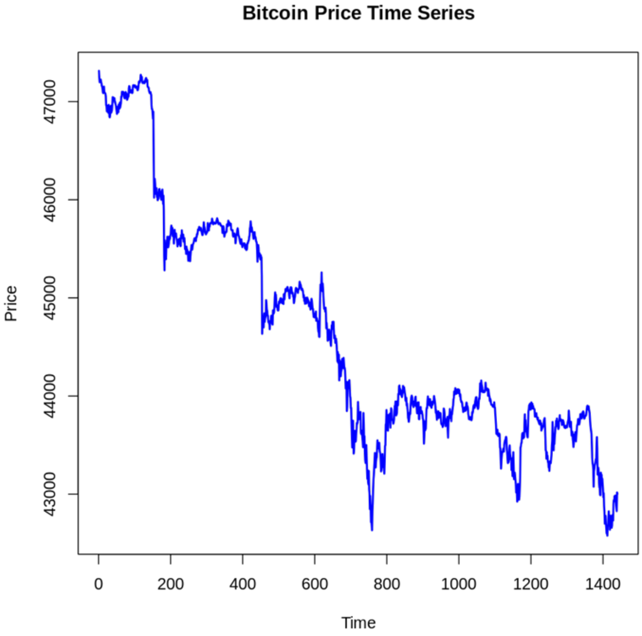


Figure 24: Bitcoin Price Time Series

Part 1.1 b)

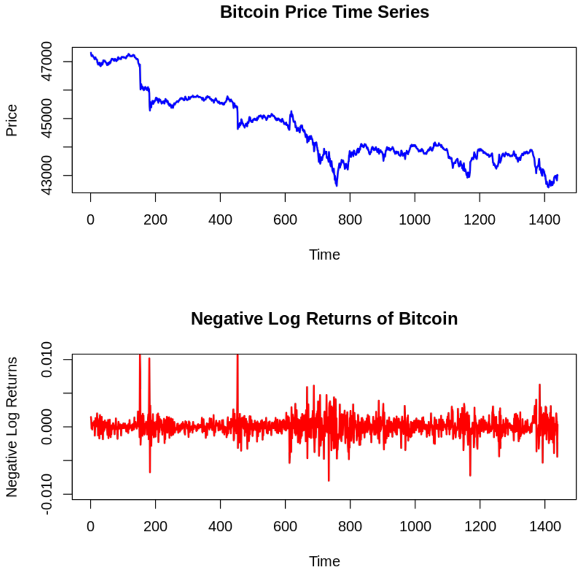


Figure 25: Time series of Bitcoin Prices and Negative Log Returns

Part 1.2 a)

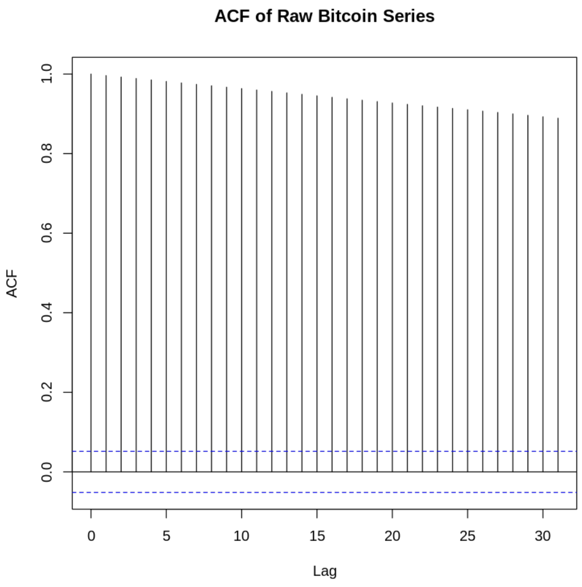


Figure 26: ACF of Raw Bitcoin Series

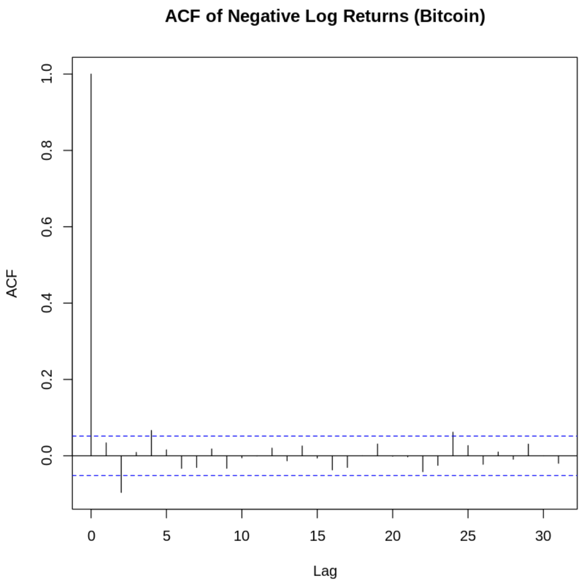


Figure 27: ACF of Negative Log Returns

Part 1.2 b)

	Test Statistic (X-squared)	Degrees of Freedom (df)	p-value
Raw Bitcoin Series	26873	20	< 2.2e-16
Negative Log Returns	33.356	20	0.03082

Table 1: Ljung-Box Test Results for Raw Bitcoin Price Series and Negative Log Returns

Part 1.2 c)

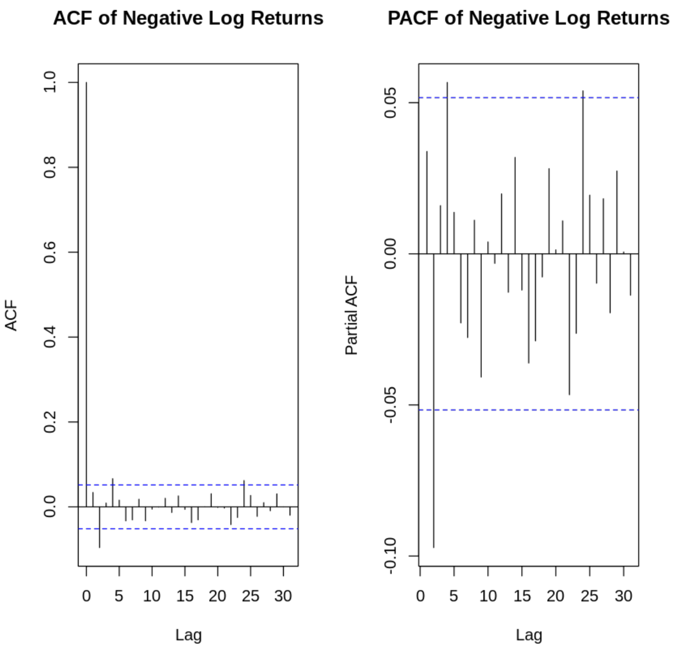


Figure 28: ACF and PACF of Negative Log Returns

	Q*	Degrees of Freedom (df)	p-value
Residuals from ARIMA(1,0,1)	19.33	8	0.01319
Model df	2		
Total lags used	10		

Table 2: Ljung-Box Test Results for Residuals from ARIMA(1,0,1) with Non-Zero Mean

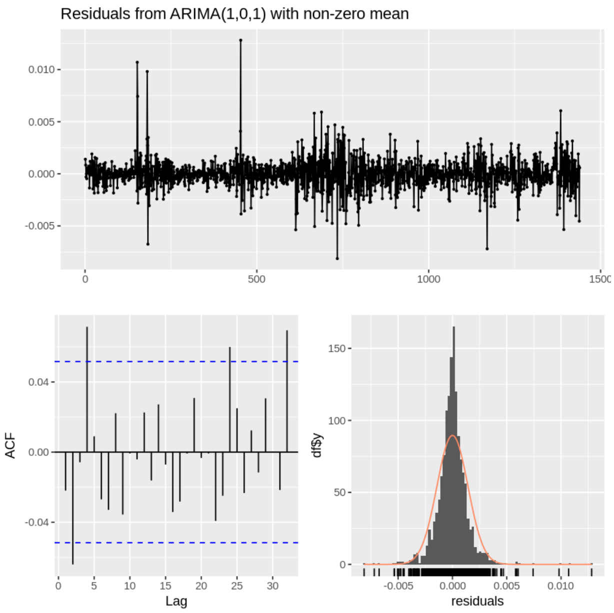


Figure 29: Residuals from ARIMA(1,0,1) model

Parameter	**Value**
Series	log_returns
Model	ARIMA(2,0,2) with non-zero mean
Coefficients	
ar1	-0.0520
ar2	-0.5415
ma1	0.0853
ma2	0.4479
mean	1e-04
Standard Errors	
ar1	0.1717
ar2	0.1664
ma1	0.1824
ma2	0.1773
mean	0
Sigma^2	2.029e-06
Log Likelihood	7391.82
AIC	-14771.65
AICc	-14771.59
BIC	-14740.02

Table 3: ARIMA Model Summary for Negative Log Returns

	Q*	Degrees of Freedom (df)	p-value
Residuals from ARIMA(2,0,2)	4.7774	6	0.5727
Model df	4		
Total lags used	10		

Table 4: Ljung-Box Test Results for Residuals from ARIMA(2,0,2) with Non-Zero Mean

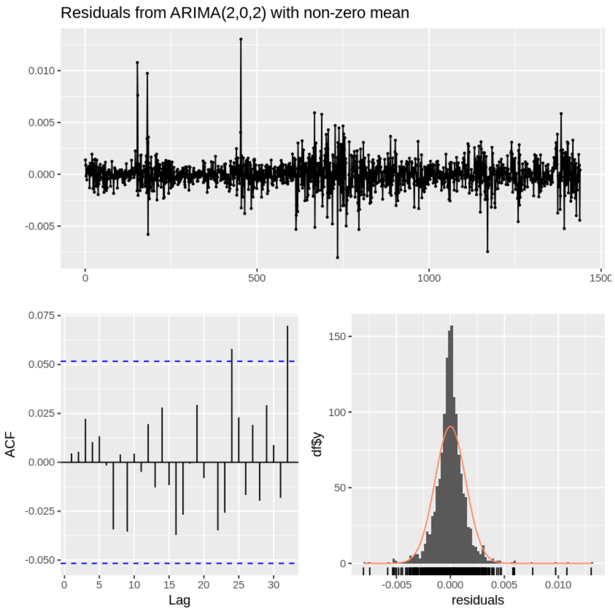


Figure 30: Residuals from ARIMA(2,0,2) model

Part 1.2 d)

	Normal Distribution	**t-Distribution**
Mean (μ)	1.5195×10^{-5}	3.5590×10^{-5}
Omega (ω)	5.5039×10^{-8}	4.0913×10^{-8}
Alpha1 (α_1)	0.25355	0.19082
Beta1 (β_1)	0.76629	0.81752
Shape (t-distribution only)	-	4.2798
Log Likelihood	7632.108	7736.355
AIC	-10.60196	-10.74545

Table 5: GARCH(1,1) Model Summary for Negative Log Returns

	Statistic	**p-Value**
Jarque-Bera Test (Normal)	2244.82	0.0000
Shapiro-Wilk Test (Normal)	0.9477	0.0000
Ljung-Box Test $Q(10)$ **(Normal)**	11.20	0.3419
Ljung-Box Test $Q(15)$ **(Normal)**	12.30	0.6559
Ljung-Box Test $Q(20)$ **(Normal)**	13.76	0.8425
Ljung-Box Test $Q(10)$ **(t-Distribution)**	11.09	0.3507
Ljung-Box Test $Q(15)$ **(t-Distribution)**	12.24	0.6604
Ljung-Box Test $Q(20)$ **(t-Distribution)**	13.47	0.8563

Table 6: Standardized Residuals Tests for GARCH (1,1) Models

Part 1.2 e)

Coefficient	Estimate
ar1	-0.0520
ar2	-0.5415
ma1	0.0853
ma2	0.4479
mean	1×10^{-4}

Table 7: ARIMA(2,0,2) Coefficients for Negative Log Returns

Coefficient	Estimate
mu	-1.966×10^{-6}
omega	6.188×10^{-8}
alpha1	0.25996
beta1	0.75793

Table 8: GARCH (1,1) Coefficients for ARIMA Residuals

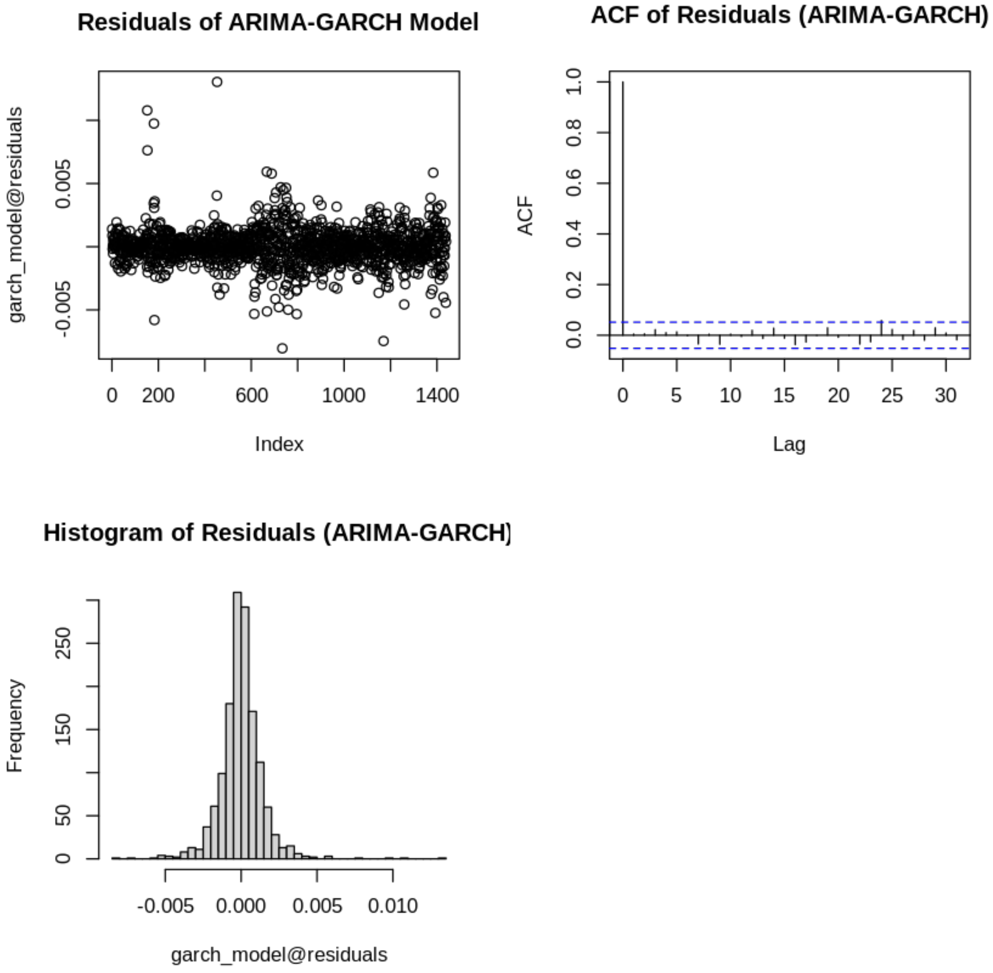


Figure 31: Residual Time Series and ACF Plot from ARIMA-GARCH(1,1) Model

Part 1.3 b)

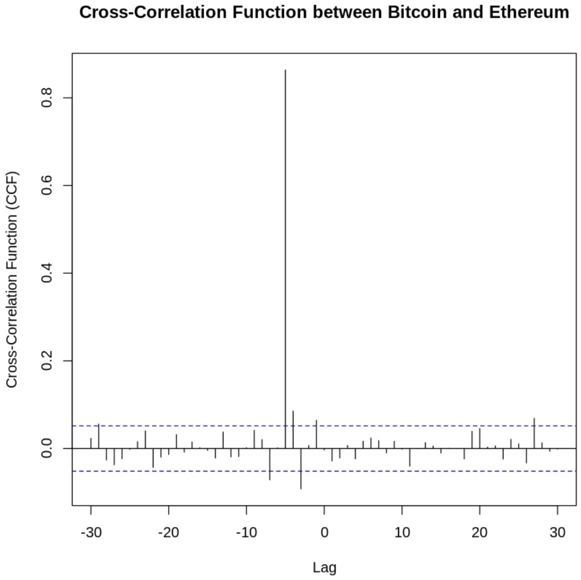


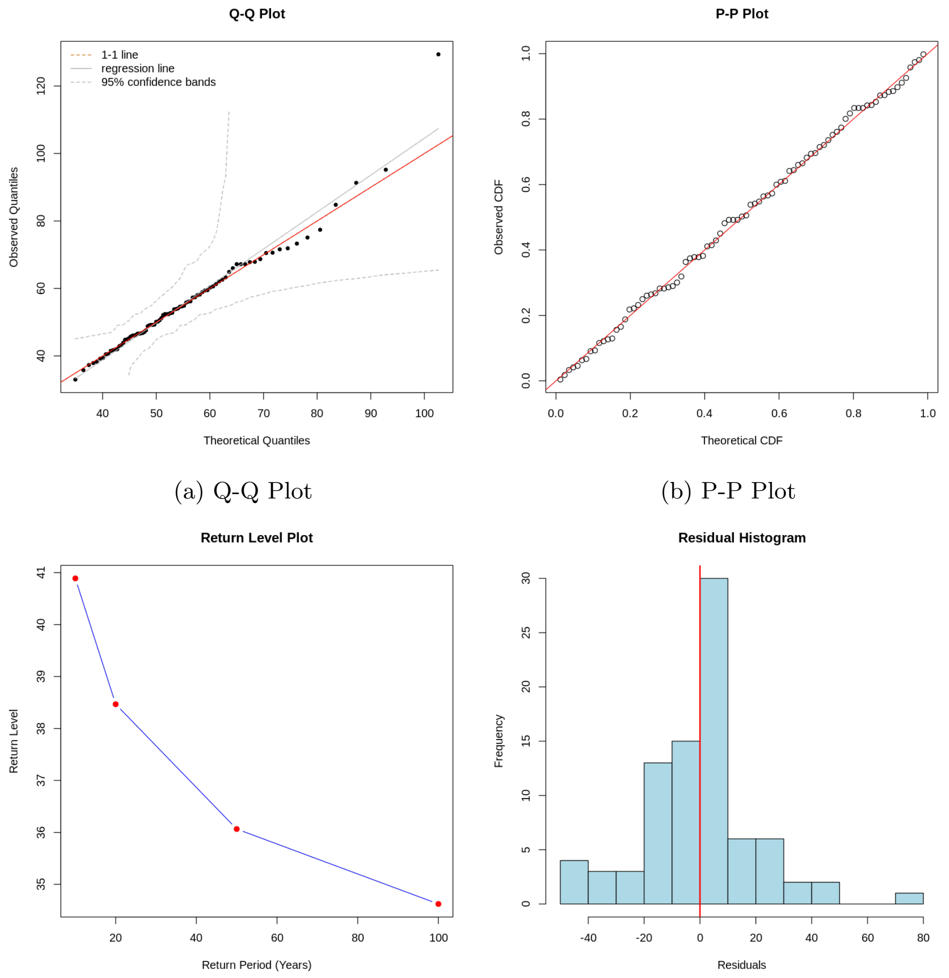
Figure 32: Cross-Correlation Function between Bitcoin and Ethereum

Part 2.1 a)

Statistic / Criterion	**Gamma**	**Weibull**	**Exponential**
Kolmogorov-Smirnov Statistic	0.0404	0.0358	0.1141
Cramer-von Mises Statistic	3.5981	1.9311	56.2976
Anderson-Darling Statistic	30.2724	20.9436	396.2175
Akaike's Information Criterion (AIC)	78870.89	78739.12	80072.31
Bayesian Information Criterion (BIC)	78885.88	78754.11	80079.80

Table 9: Goodness-of-Fit Statistics for Precipitation Data

Part 2.1 e)



(a) Q-Q Plot

(b) P-P Plot

(c) Return Level Plot

(d) Residual Histogram

Figure 33: Diagnostic plots for the GEV model fitted to the yearly maximum precipitation data

Part 2.2 c)

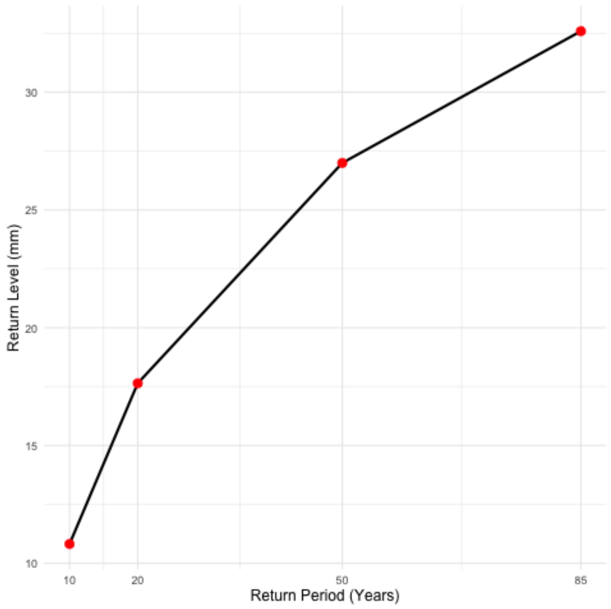
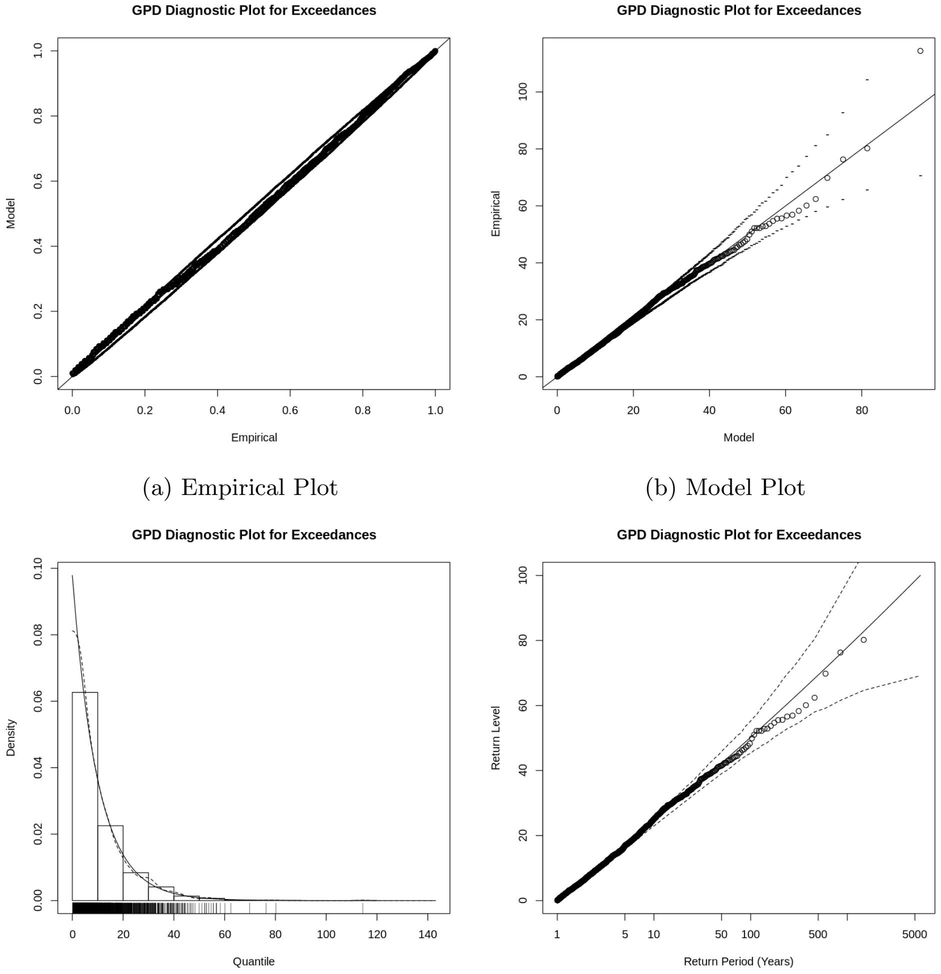


Figure 34: Return Level for different timescales



(a) Empirical Plot

(b) Model Plot

(c) Caption for Quantile Plot

(d) Return Period Plot

Figure 35: Diagnostic plots of the GPD model

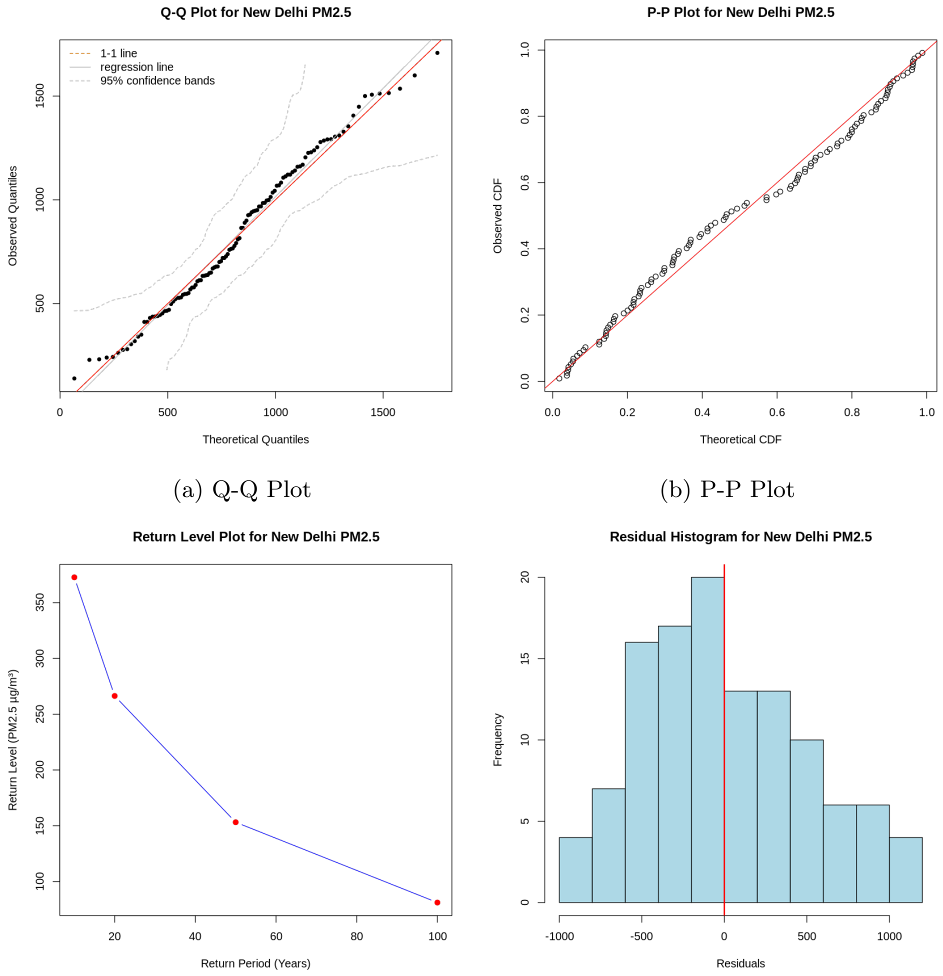
Part 3.2

Statistic / Criterion	**Gamma**	**Weibull**	**Exponential**	**Log-normal**	**Normal**	**Pareto**
Kolmogorov-Smirnov Statistic	0.0700	0.0651	0.0906	0.0200	0.1624	0.0906
Cramer-von Mises Statistic	29.1726	26.0202	31.3425	2.4447	199.2459	31.3413
Anderson-Darling Statistic	161.9768	170.2775	263.8264	17.4764	1125.6599	263.8158
Akaike's Information Criterion (AIC)	241866.9	242382.7	243069.3	240403.5	256927.5	243071.3
Bayesian Information Criterion (BIC)	241882.6	242398.4	243077.1	240419.2	256943.2	243086.9

Table 10: Goodness-of-Fit statistics for PM2.5 hourly concentrations

Distribution	**AIC**	**BIC**
Normal	256927.472	256943.2
Gamma	241866.883	241882.6
Log-normal	240403.487	240419.2
Pareto	243071.265	243086.9
GEV	-1688.549	NA

Table 11: AIC and BIC values for fitted distributions of maximum weekly PM2.5 concentrations



(a) Q-Q Plot

(b) P-P Plot

(c) Return Level Plot

(d) Residual Histogram

Figure 36: Diagnostic plots for the GEV model on weekly max. PM2.5 concentrations

Part 3.3

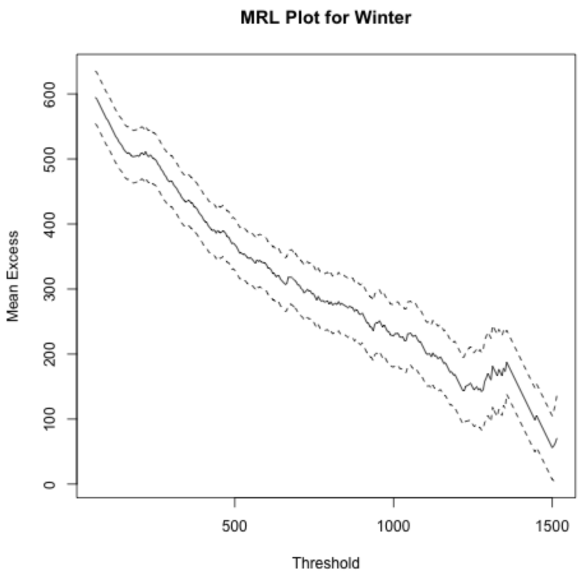


Figure 37: Winter Season PM2.5 Concentration Distribution

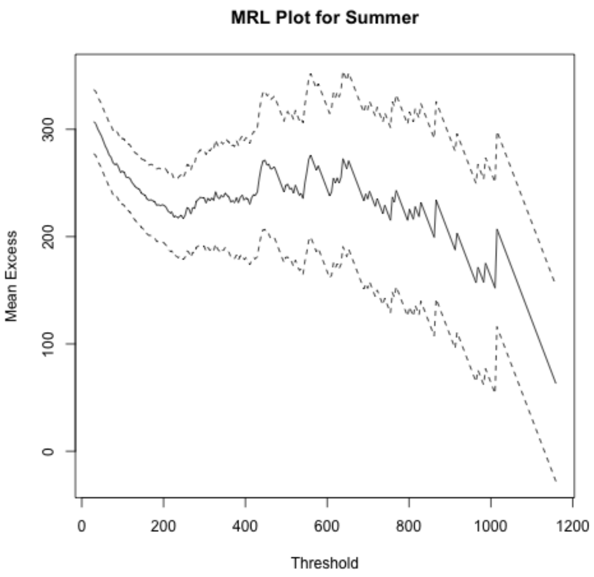


Figure 38: Summer Season PM2.5 Concentration Distribution

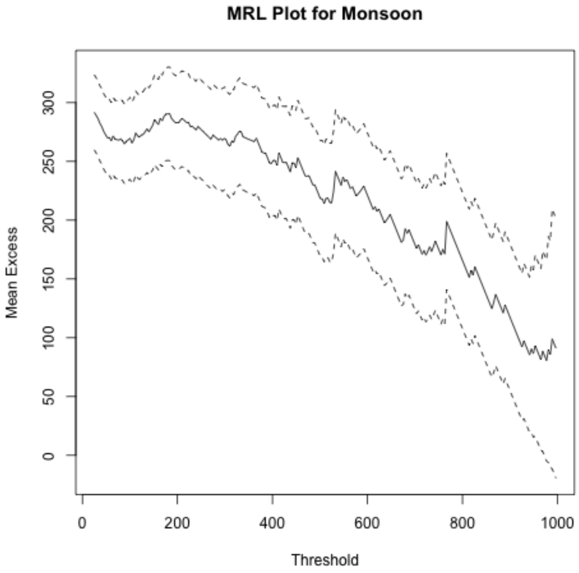


Figure 39: Monsoon Season PM2.5 Concentration Distribution

Part 3.4

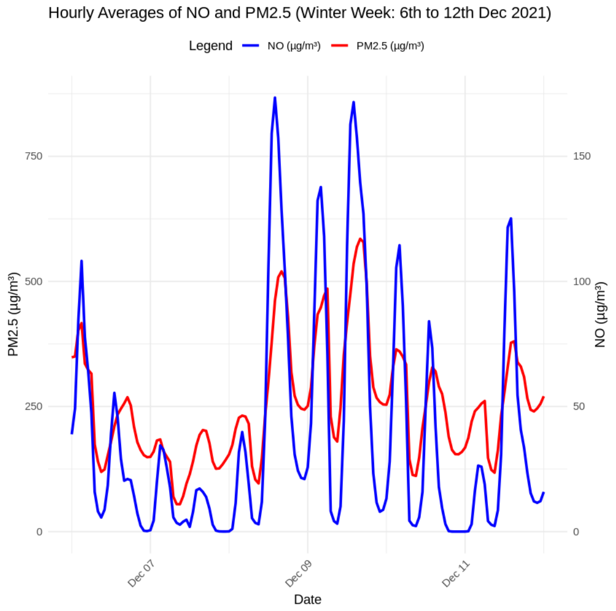


Figure 40: Comparison of hourly averages of NO and PM2.5 (Winter week : 6-12 Dec 2021)

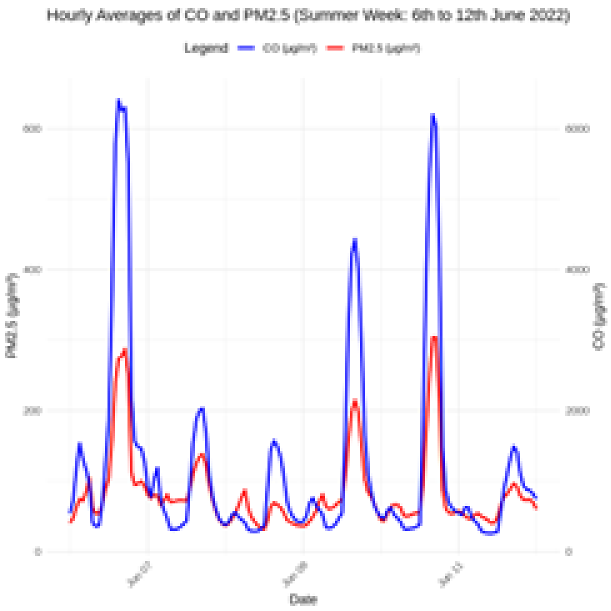


Figure 41: Comparison of hourly averages of CO and PM2.5 (Summer week : 6-12 June 2021)

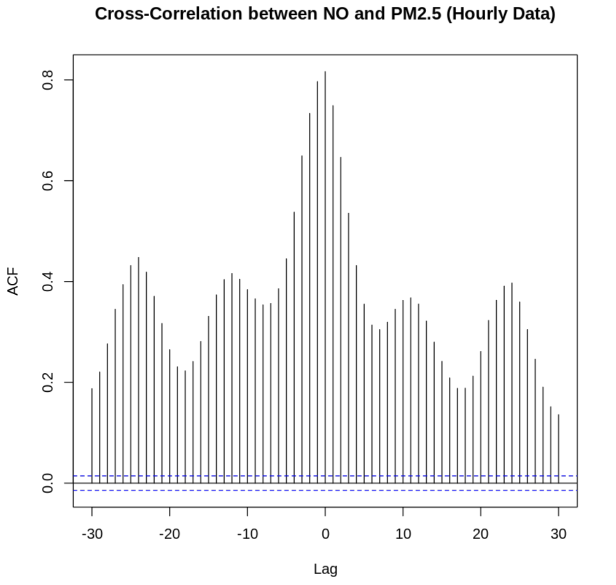


Figure 42: Cross-Correlation between NO and PM2.5 (hourly data)

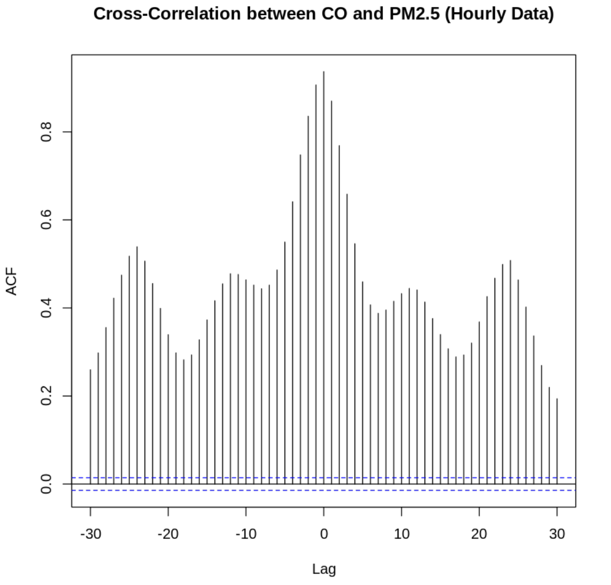


Figure 43: Cross-Correlation between CO and PM2.5 (hourly data)

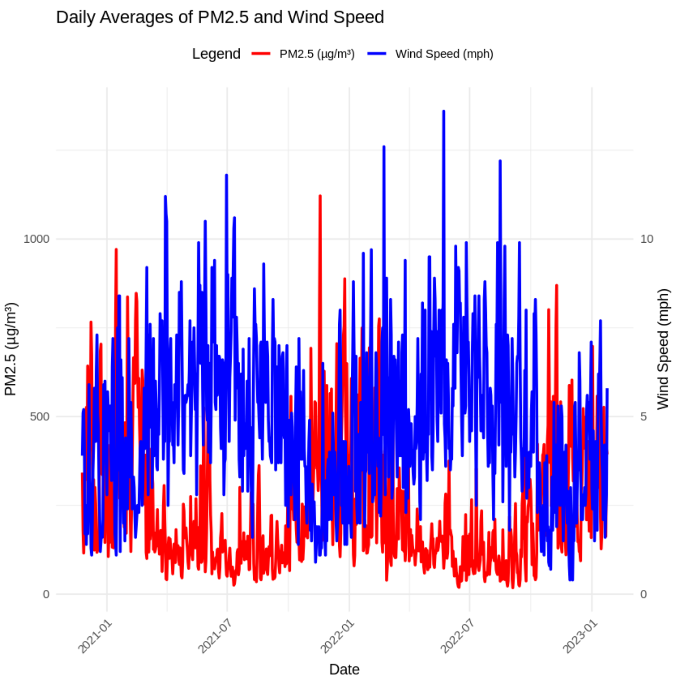


Figure 44: Daily Averages of PM2.5 and Wind Speed

References

Cichowicz, R., Wielgosiński, G., & Fetter, W. (2017). Dispersion of atmospheric air pollution in summer and winter season. *Environmental Monitoring and Assessment*, *189*(12). https://doi.org/10.1007/s10661-017-6319-2

Guttikunda, S. K., & Gurjar, B. R. (2012). Role of meteorology in seasonality of air pollution in megacity delhi, india. *Environmental Monitoring and Assessment*, *184*(5), 3199–3211. https://doi.org/10.1007/s10661-011-2182-8

Nagpure, A. S., Gurjar, B., Kumar, V., & Kumar, P. (2015). Estimation of exhaust and non-exhaust gaseous, particulate matter and air toxics emissions from on-road vehicles in delhi. *Atmospheric Environment*, *127*, 118–124. https://doi.org/10.1016/j.atmosenv.2015.12.026

Pant, P., Shukla, A., Kohl, S. D., Chow, J. C., Watson, J. G., & Harrison, R. M. (2015). Characterization of ambient pm2.5 at a pollution hotspot in new delhi, india and inference of sources. *Atmospheric Environment*, *109*, 178–189. https://doi.org/10.1016/j.atmosenv.2015.02.074

Weather Underground. (n.d.). New delhi, india weather conditions — weather underground [n.d.]. https://www.wunderground.com/weather/in/new-delhi

World Health Organization. (2021). Air quality guidelines: Global update 2021 [Accessed December 2024]. https://www.who.int/publications/i/item/9789240034228