ARMA(1,1) Model Analysis for Nile River Flow

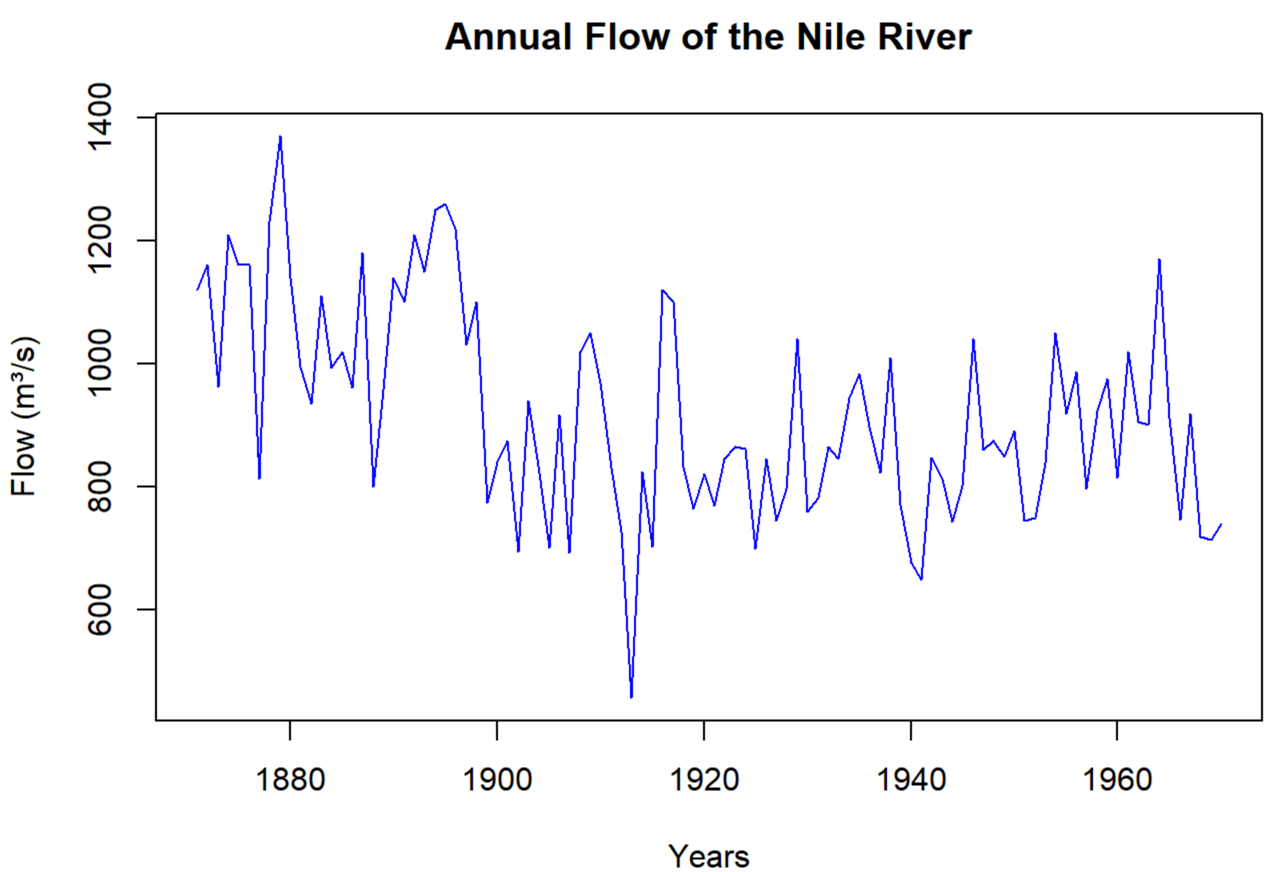
Victor Cornejo

2025-02-14

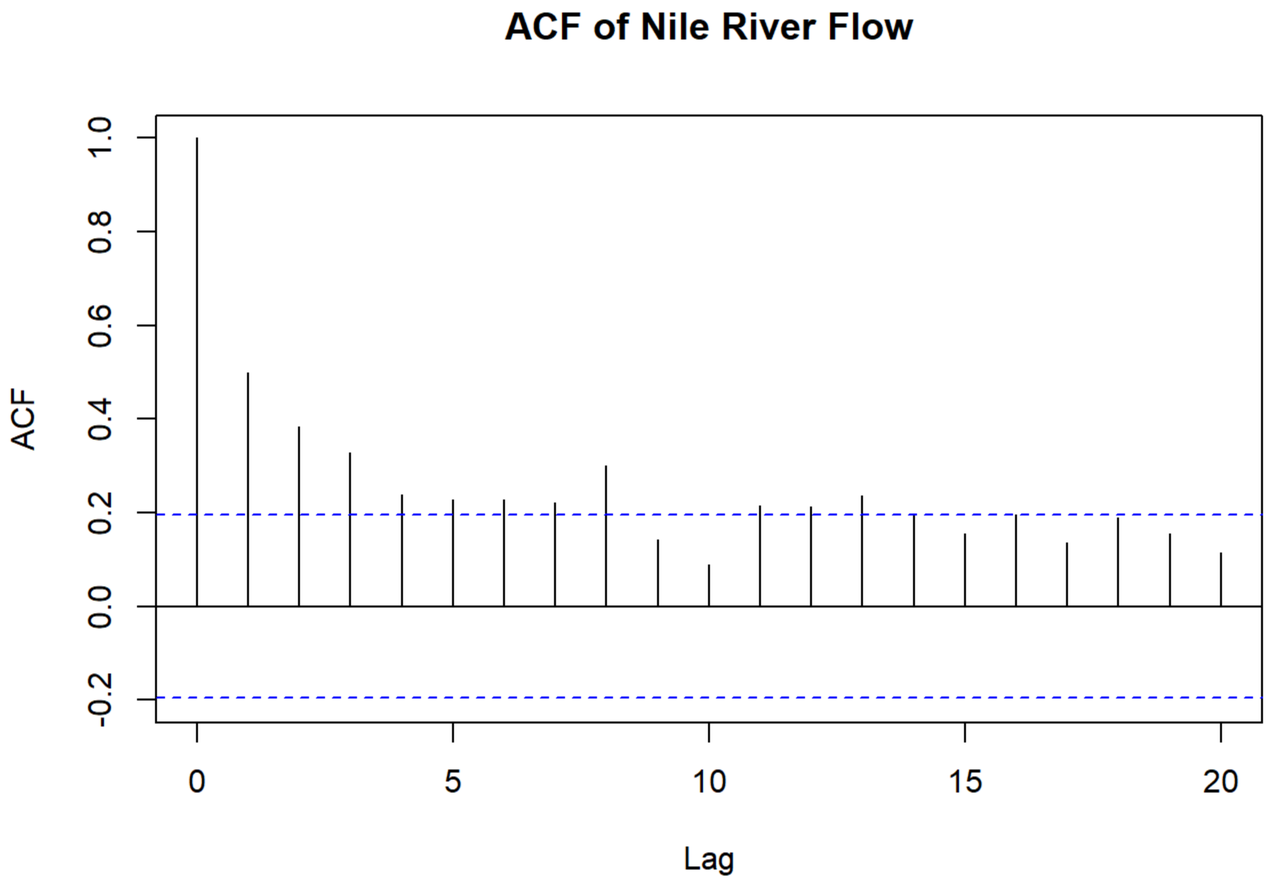
This document analyzes the annual flow of the Nile River using an ARMA(1,1) model. The analysis includes:

- **Filtering the time series using an ARMA(1,1)** model to compute conditional means and innovations.
- **Estimating the model parameters** and assessing residuals.
- **Forecasting future flow values** using the estimated ARMA model.
- **Computing the log-likelihood function** to optimize model parameters.

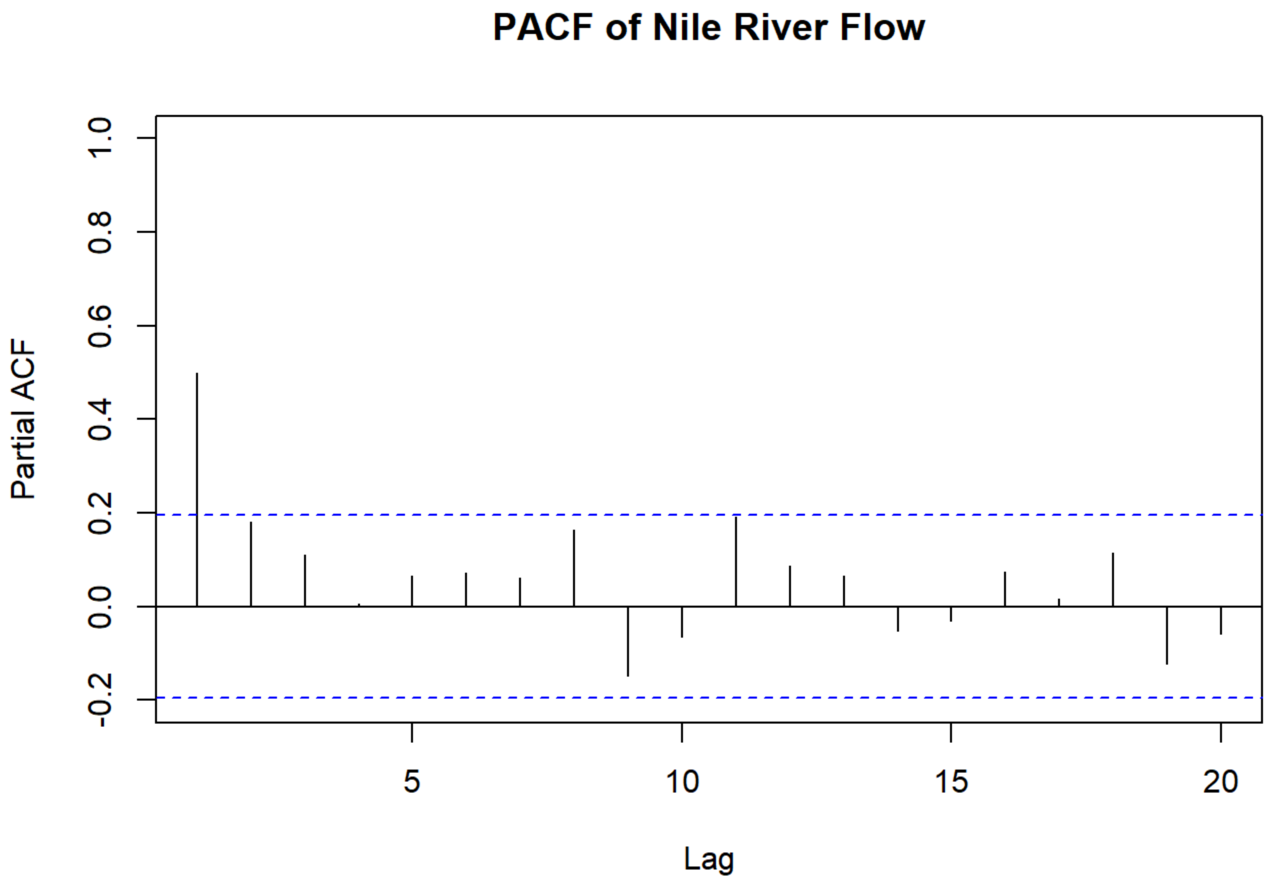
```
# Load necessary library
library(datasets)
# This library contains dataset about diverse topics, including the Annual Flow of the Nile Rive
r at Aswan from 1871 to 1970.
# Load the Nile dataset
plot(Nile, main="Annual Flow of the Nile River", col="blue", xlab="Years", ylab="Flow (m³/s)")
```



```
acf(Nile, ylim=c(-0.2,1), main="ACF of Nile River Flow")
```




```
pacf(Nile, ylim=c(-0.2,1), main="PACF of Nile River Flow")
```




```r
# Convert to numeric format
y <- as.numeric(Nile)
```

Interpretation of the ACF Plot

- The first lag has a strong positive autocorrelation (~0.9), indicating that the current flow value is highly dependent on the previous year's value. Which is characteristic of an autoregressive (AR) process.
- Since the decay is not strictly exponential but rather gradual, this suggests that an AR(1) or AR(2) process might be a good model. The slow decay suggests the presence of a trend component or a long-term dependency structure.

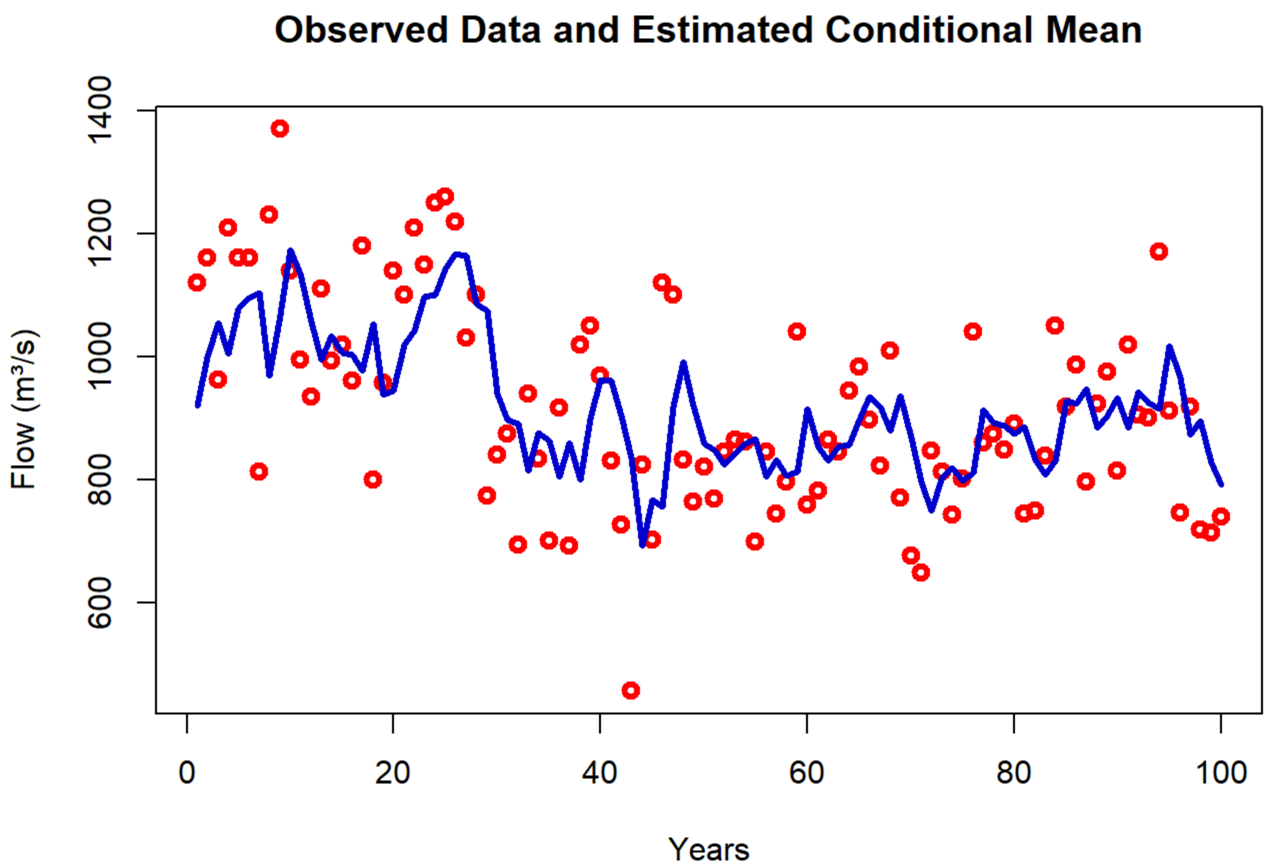
Interpretation of the PACF Plot

- The PACF shows a significant spike at lag 1 (~0.9) and then quickly drops near zero for subsequent lags. This suggests that most of the predictive power is captured by the first lag only, supporting an AR(1) process.
- The small spikes at other lags are within the confidence intervals, meaning they are likely due to random noise.

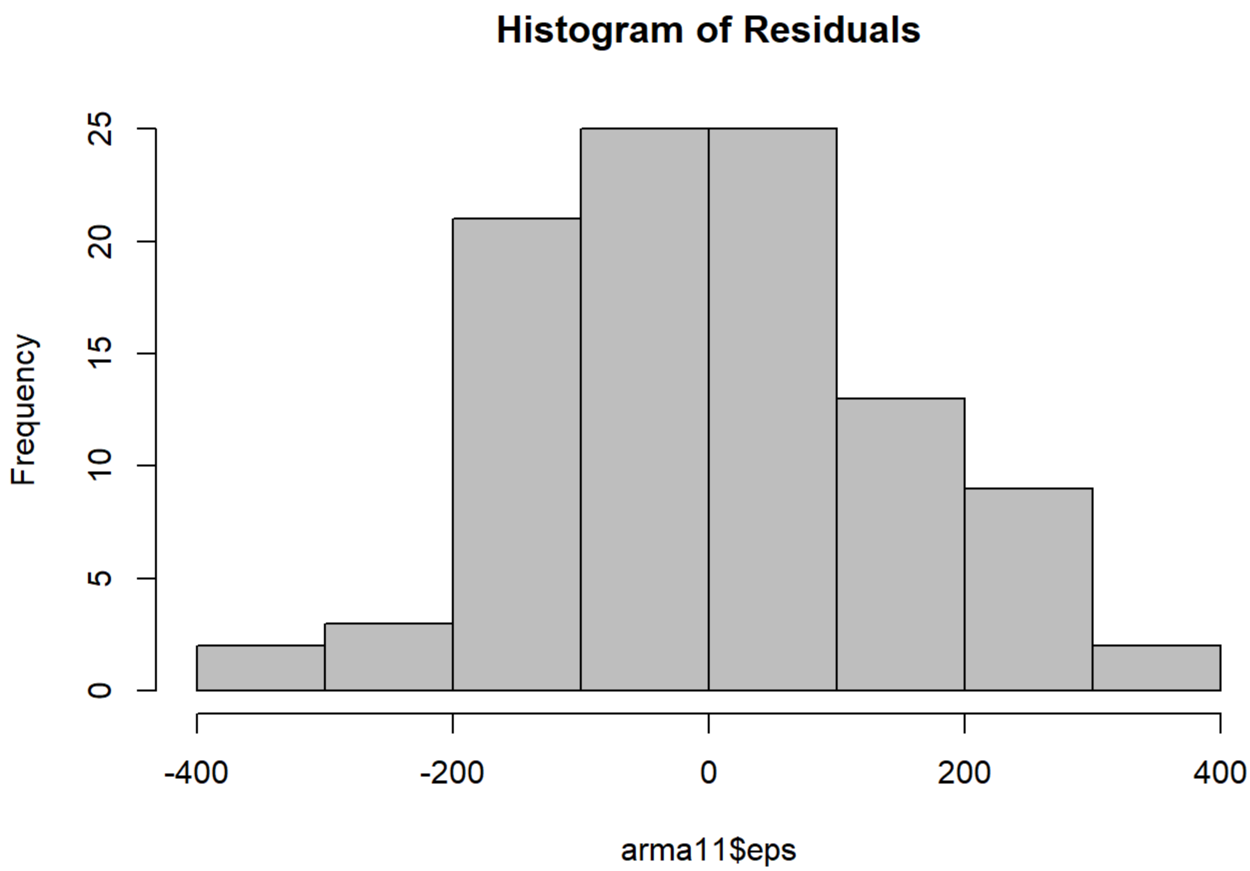
Now, we will analyze the inclusion of an MA section in the model.

ARMA(1,1) Filter Function

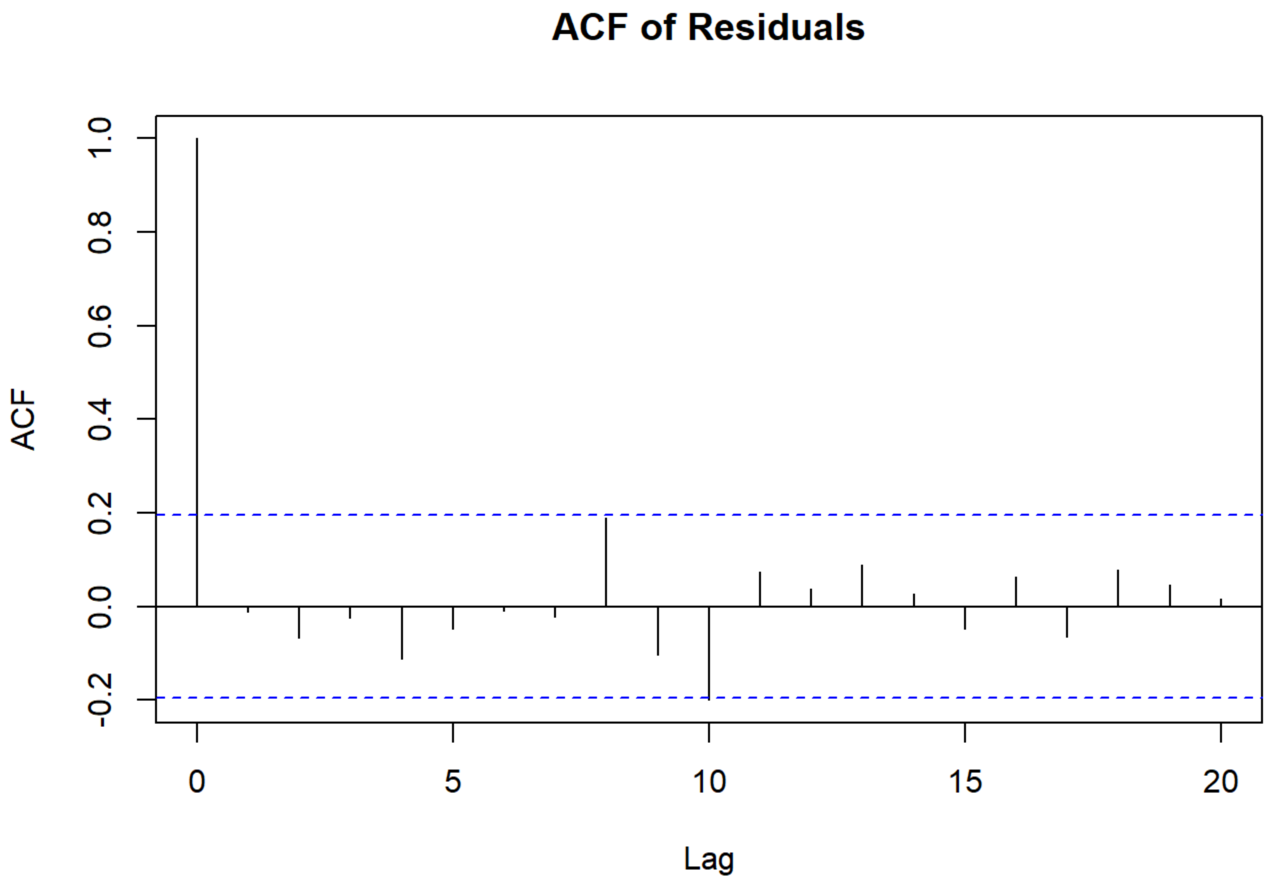
```r
# Function to compute conditional means and innovations using ARMA(1,1)
my.arma11.filter <- function(y, param) {
  c <- param[1]
  phi1 <- param[2]
  theta1 <- param[3]

  T <- length(y)
  mu <- numeric(T)
  eps <- numeric(T)

  mu[1] <- c / (1 - phi1)
  eps[1] <- y[1] - mu[1]

  for (t in 2:T) {
    mu[t] <- c + phi1 * y[t - 1] + theta1 * eps[t - 1]
    eps[t] <- y[t] - mu[t]
  }

  return(list(mu = mu, eps = eps))
}

# Define initial parameters
param <- c(mean(y) * (1 - 0.9), 0.9, -0.5)

# Apply the ARMA filter
arma11 <- my.arma11.filter(y, param)

# Plot the results
plot(y, lwd=3, col="red", main="Observed Data and Estimated Conditional Mean", xlab="Years", yla
b="Flow (m³/s)")
lines(arma11$mu, lwd=3, col="blue3")
```



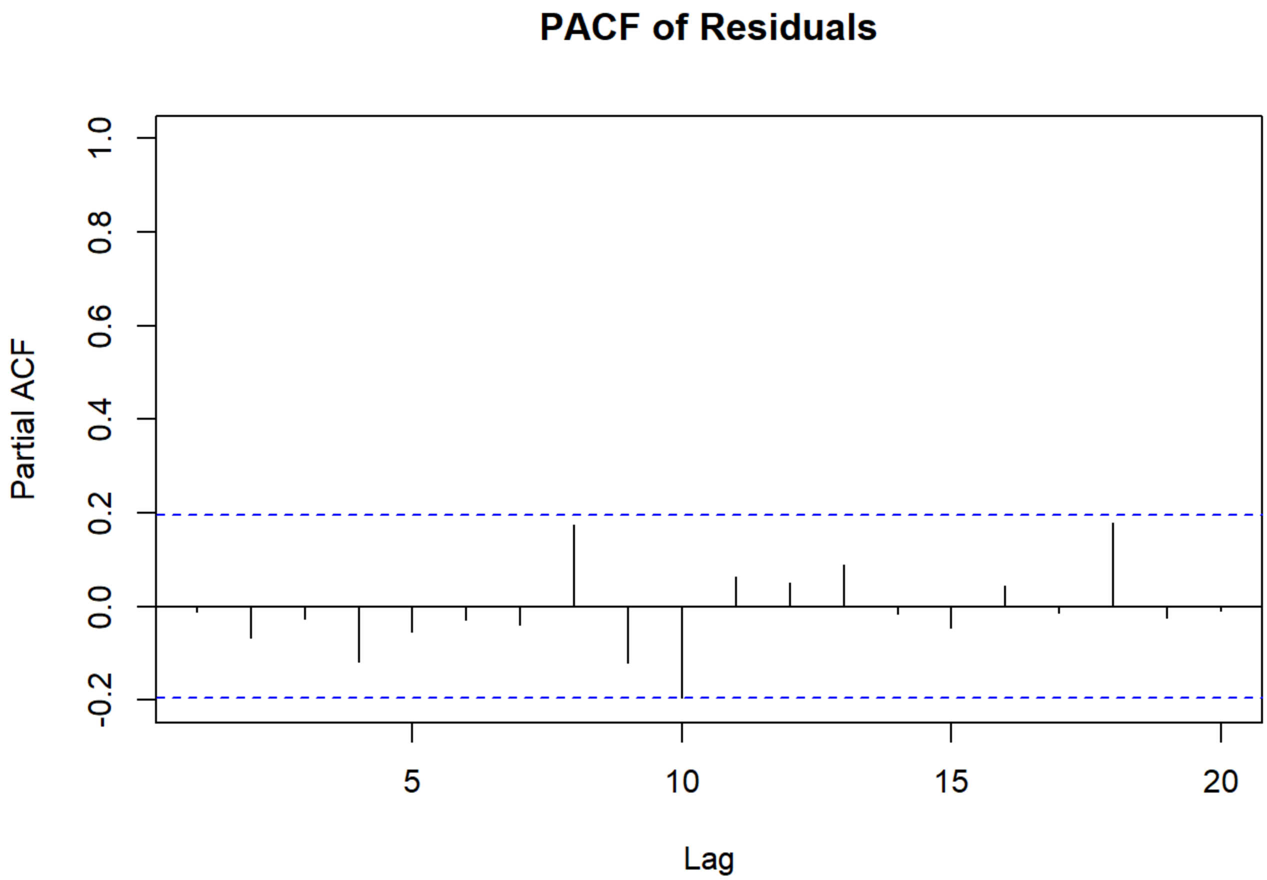

```
# Analyze residuals
hist(arma11$eps, main="Histogram of Residuals", col="gray")
```




```
acf(arma11$eps, ylim=c(-0.2,1), main="ACF of Residuals")
```




```
pacf(arma11$eps, ylim=c(-0.2,1), main="PACF of Residuals")
```



Interpretation of the Observed Data and Estimated Conditional Mean Plot

- The red points represent the observed Nile River flow (m^3/s) over time. The blue line represents the estimated conditional mean from the ARMA(1,1) model.
- The model captures some of the underlying trends and variations in the data but still shows deviations from the observed values. The discrepancies suggest that there might be unexplained variance, potentially indicating a need for further model refinement

Interpretation of the Histogram of Residuals

- The residuals appear to be approximately normally distributed, with a slight right skew. The majority of residuals are centered around zero, which suggests that the model does not have a strong bias.

Interpretation of the ACF Plot of Residuals

- The first lag shows a significant autocorrelation, but all subsequent lags fall within the confidence interval. The fact that later lags are insignificant implies that most dependencies have been captured by the ARMA(1,1) model.

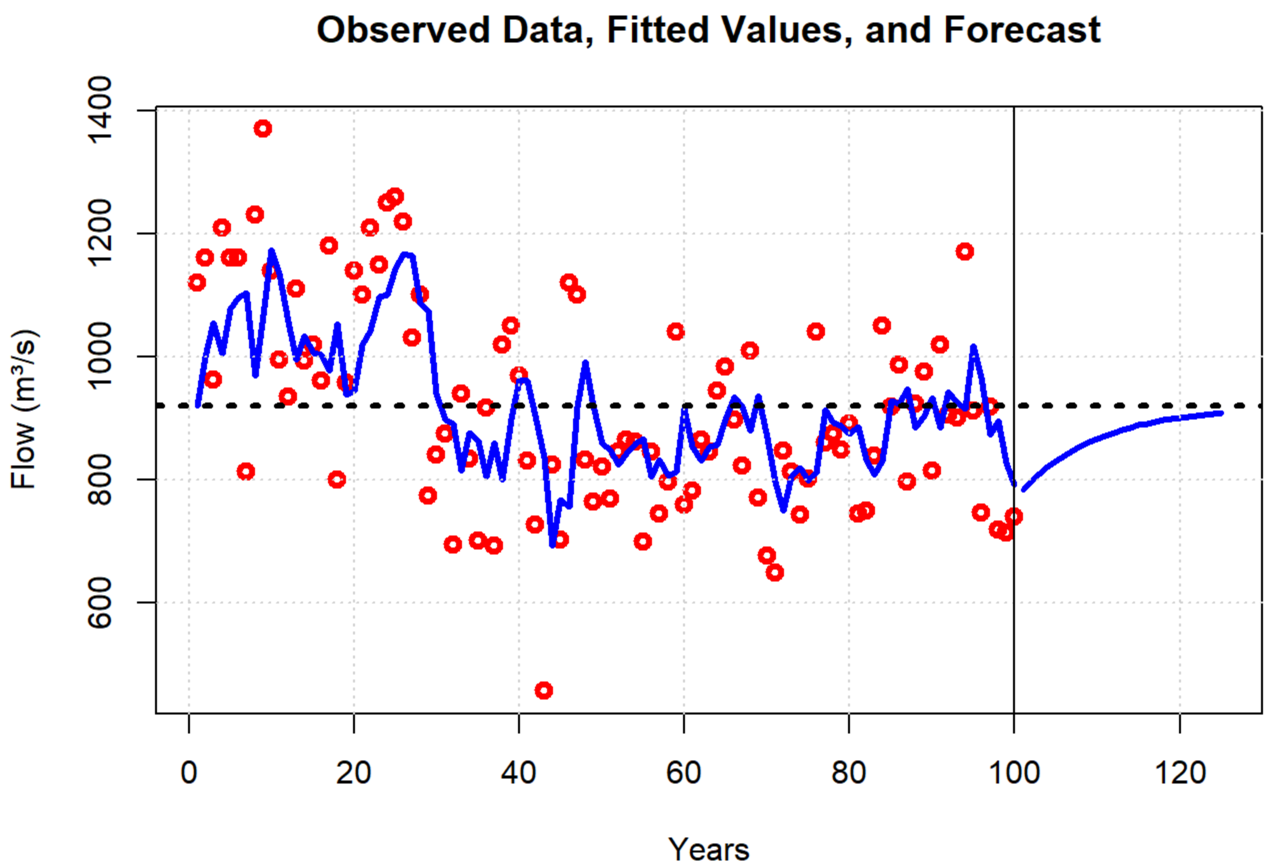
Interpretation of the PACF Plot of Residuals

- The PACF does not show any significant spikes, meaning there is no clear remaining autoregressive structure in the residuals. If we see significant spikes, it means that the errors are correlated at those lags and would indicate that our model is missing an important autoregressive (AR) component and might need an additional AR term.

- The first few lags are small and stay within the blue confidence bands. This means that there is no significant pattern left in the residuals. Since the bars do not extend beyond the blue confidence intervals, this suggests that no additional AR terms are needed.
- Remember that residuals primarily diagnose the MA component, but their PACF can also help check if more AR terms are needed.

Forecasting Future Values

```
# Function to generate forecasts using the ARMA(1,1) model
my.arma11.forecast <- function(y, param, H) {
  c <- param[1]
  phi1 <- param[2]
  theta1 <- param[3]

  T <- length(y)
  filter_result <- my.arma11.filter(y, param)
  mu <- filter_result$mu
  eps <- filter_result$eps

  y_forecast <- numeric(H)

  for (h in 1:H) {
    if (h == 1) {
      y_forecast[h] <- c + phi1 * y[T] + theta1 * eps[T]
    } else {
      y_forecast[h] <- c + phi1 * y_forecast[h - 1]
    }
  }

  return(y_forecast)
}

# Forecast 25 future years
H <- 25
y.hat <- my.arma11.forecast(y, param, H)

T <- length(y)

# Plot observed data, fitted values, and forecasts
plot(y, lwd=3, col="red", xlim=c(1, T + H), main="Observed Data, Fitted Values, and Forecast",xl
ab="Years", ylab="Flow (m³/s)")
lines(arma11$mu, lwd=3, col="blue1")
lines((T+1):(T+H), y.hat, lwd=3, col="blue1")
grid()
abline(h=mean(y), lty=3, lwd=3)
abline(v=T, lwd=1)
```



Interpretation of the Observed Data, Fitted Values, and Forecast Plot

- The black dotted line represents the long-term mean of the flow. The forecasted values (beyond the vertical black line) indicate how the model predicts future river flow.
- The fitted values (blue line) follow the observed data reasonably well, suggesting that the ARMA(1,1) model captures much of the time series dynamics. The forecasted values converge toward the long-term mean, which is expected in an ARMA model because it assumes the series will revert to a stable mean over time.

Log-Likelihood Function

```
# Function to compute the log-likelihood of an ARMA(1,1) model
my.arma11.loglik <- function(y, param) {
  c <- param[1]
  phi1 <- param[2]
  theta1 <- param[3]
  sigma2 <- param[4]

  filter_result <- my.arma11.filter(y, param[1:3])
  mu <- filter_result$mu
  eps <- filter_result$eps

  T <- length(y)
  loglik <- - (T / 2) * log(2 * pi * sigma2) - (1 / (2 * sigma2)) * sum(eps^2)

  return(loglik)
}

# Compute log-likelihood for initial parameter estimates
sigma2 <- 19892
arma11.loglik <- my.arma11.loglik(y, c(param, sigma2))
```

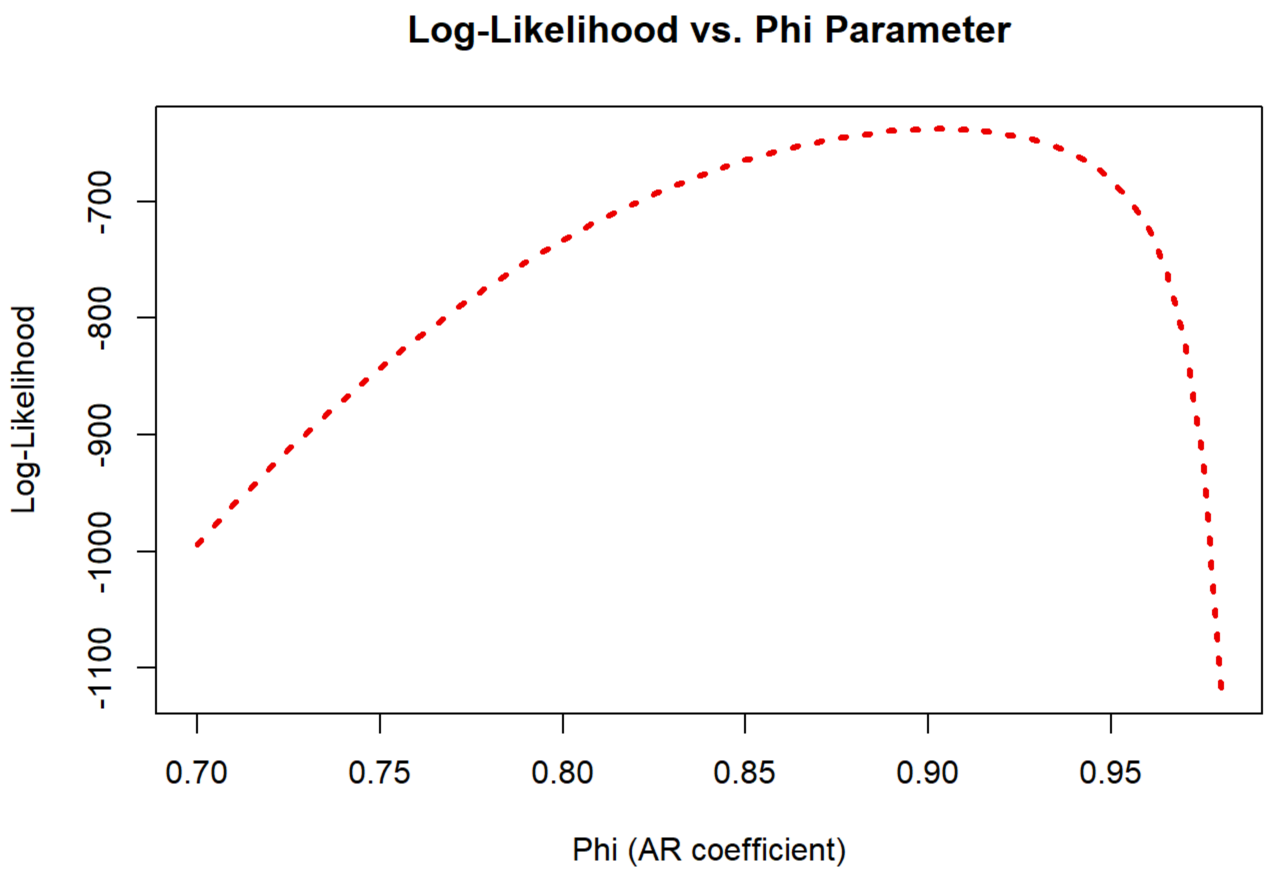
Analyzing Log-Likelihood Across Phi Values

```
# Define a range of phi values
phi.range <- seq(0.7, 0.98, 0.005)
loglik.range <- numeric(length(phi.range))

# Compute log-likelihood for each phi value
for (i in 1:length(phi.range)) {
  cat(".")
  loglik.range[i] <- my.arma11.loglik(y, c(param[1], phi.range[i], param[3], sigma2))
}
```

```
## .........................................................
```

```
cat("\n")
```

```
# Plot log-likelihood vs. phi parameter
plot(phi.range, loglik.range, lwd=3, col="red2", lty=3, type="l",
     main="Log-Likelihood vs. Phi Parameter",
     xlab="Phi (AR coefficient)", ylab="Log-Likelihood")
```



Interpretation of the Log-Likelihood vs. Phi Parameter Plot

- The **x-axis** represents the AR(1) coefficient (Phi), which controls the level of persistence in the model. The **y-axis** represents the log-likelihood, which indicates how well the model fits the data for each Phi value. The **red dashed line** shows the variation in log-likelihood across different Phi values.
- The log-likelihood **increases until around Phi ≈ 0.92**, suggesting this is the optimal value for the AR(1) parameter. Beyond **Phi > 0.92**, the log-likelihood **declines**, meaning models with larger Phi values provide a worse fit. This confirms that an **AR(1) process with Phi ≈ 0.92 is the most likely optimal choice**.

Conclusion

This analysis applied an **ARMA(1,1) model** to the Nile River flow data set. Key insights include:

- The ARMA(1,1) filter provides smoothed conditional means and innovations.
- The forecast function suggests future trends based on ARMA dynamics.
- The log-likelihood function allows for parameter estimation by maximizing likelihood.