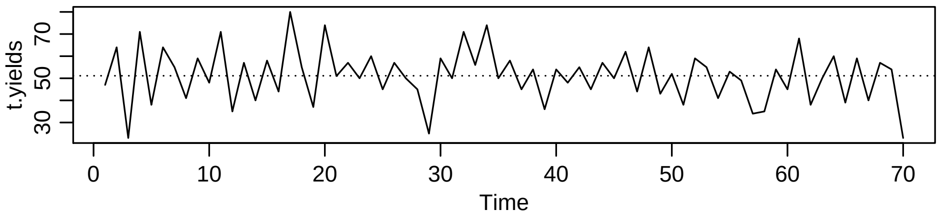
Solution to Series 3

1. **a)** Plotting and calculating the mean:

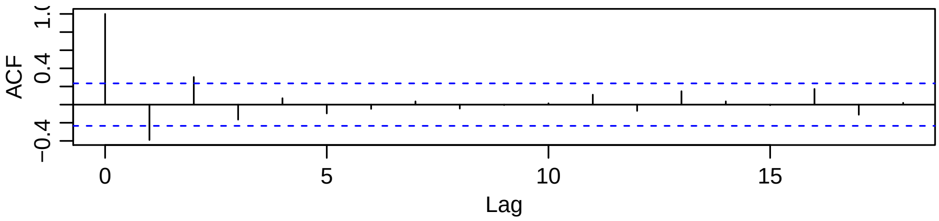
```
> mu <- mean(t.yields)
> plot(t.yields)
> abline(h = mu, lty=3)
```



We can regard this time series as being stationary.

b) Plotting the ACF:

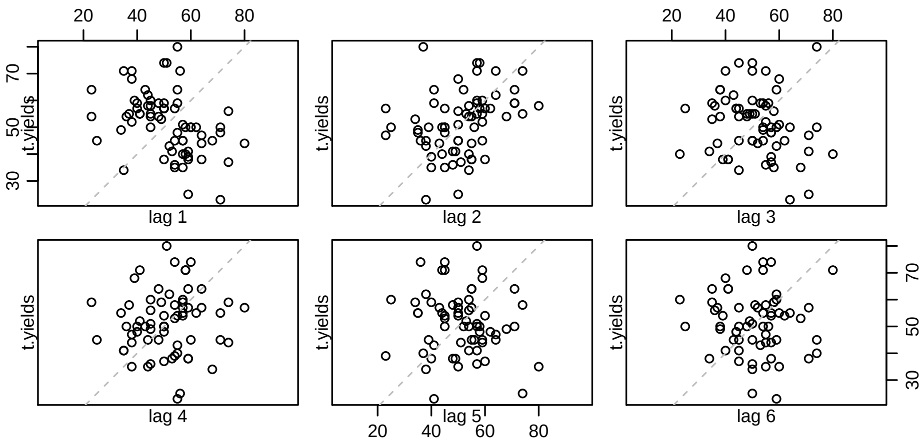
```
> acf(t.yields, plot = TRUE)
```



The correlogram shows us that for lags $k \geq 3$, the estimated autocorrelations $\widehat{\rho}\langle k \rangle$ do not differ significantly from 0. The first of these autocorrelations is negative; as the time series oscillates very noticeably, this negativity is not at all surprising.

Looking at lagged scatterplots:

```
> lag.plot(t.yields, lag = 6, layout = c(2, 3), do.lines = FALSE)
```



In the lagged scatterplot with lag 1 the pairs $[x_t, x_{t+1}]$ show the negative linear relationship we expected from the correlogram. For lag 2, however, the lagged scatterplot shows up a positive linear relationship, and for lag $k \geq 4$ we see no further correlation. The pairs $[x_t, x_{t+3}]$ (lagged scatterplot at lag 3) still have a slightly negative connection, but the correlogram tells us that we can assume $\widehat{\rho}\langle 3 \rangle = 0$.

c) The variance of the arithmetic mean $\widehat{\mu}$ is

$$\mathrm{Var}(\widehat{\mu}) = \frac{1}{n^2}\,\gamma(0)\left(n + 2\sum_{k=1}^{n-1}(n-k)\,\rho(k)\right) .$$

From the correlogram in Part b) we see that the estimated autocorrelations $\widehat{\rho}\langle k\rangle$ do not differ significantly from 0 for lags $k \geq 3$. Thus we can set all the autocorrelations $\rho(k)$ for $k \geq 3$ to 0. We obtain

$$\mathrm{Var}(\widehat{\mu}) = \frac{1}{n^2}\,\gamma(0)\Big(n + 2(n-1)\rho(1) + 2(n-2)\rho(2)\Big).$$

To estimate the variance of $\widehat{\mu}$, we replace $\gamma(0), \rho(1)$ and $\rho(2)$ by their estimates.
R code:

```
> n <- length(t.yields)
> gamma0 <- var(t.yields) * (n - 1)/n
> rho <- acf(t.yields, plot=F)$acf
> Var.mu <- n^(-2) * gamma0 * (n + 2*sum((n - 1:2)*rho[2:3]))
```

This yields an estimated variance of $\widehat{\mathrm{Var}}(\widehat{\mu}) = 1.643$.
The bounds of an approximate 95% confidence interval for the mean yield are then given by

$$\widehat{\mu} \pm 1.96 \cdot \mathrm{se}(\widehat{\mu}) = \widehat{\mu} \pm 1.96 \cdot \sqrt{\widehat{\mathrm{Var}}(\widehat{\mu})} .$$

In our case, we get a confidence interval of $[48.62, 53.64]$.
If we assume independence, the variance of $\widehat{\mu}$ is estimated as

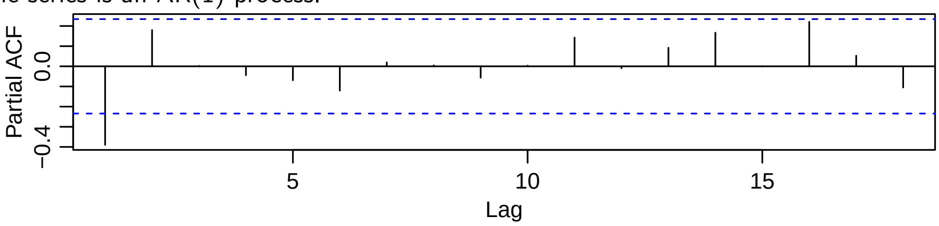
$$\widehat{\mathrm{Var}}(\widehat{\mu}) = \frac{1}{n^2}\sum_{s=1}^{n}\widehat{\mathrm{Var}}(X_s) = \frac{\widehat{\gamma}(0)}{n} = 1.997 .$$

Under this independence assumption, therefore, an approximate 95% confidence interval for the mean yield is given by

$$[\widehat{\mu} - 1.96 \cdot \mathrm{se}(\widehat{\mu}), \widehat{\mu} + 1.96 \cdot \mathrm{se}(\widehat{\mu})] = [48.36, 53.90] .$$

Thus the correct specification of the independence structure here leads to a confidence interval which is 10% narrower.

d) Only the first partial autocorrelation is significantly different from zero. Thus we can assume that our time series is an AR(1) process.



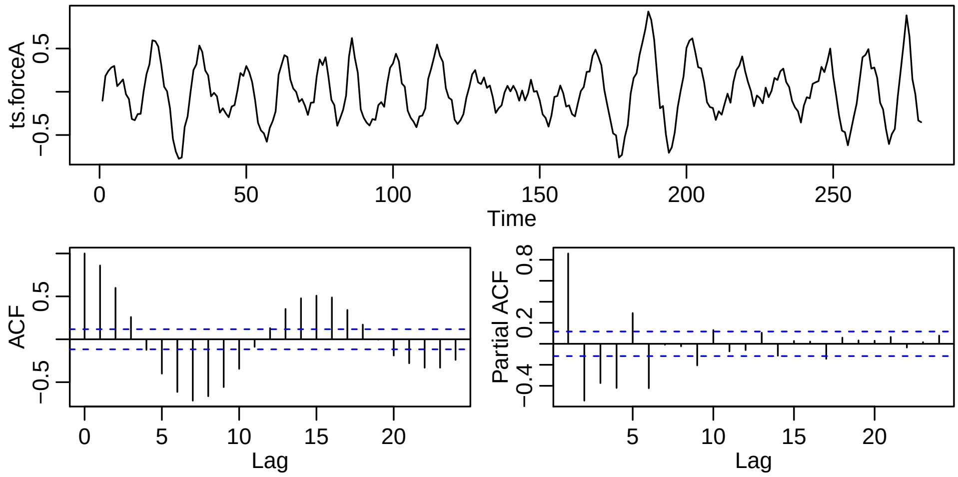
e) For an AR(1) model, the Yule-Walker equations simplify to

$$\widehat{\rho}(1) = \widehat{\alpha} \cdot \widehat{\rho}(0), \quad \widehat{\rho}(0) = 1$$

σ^2 can be estimated by $\widehat{\sigma}^2 = \widehat{\sigma}_X^2 \cdot (1 - \widehat{\alpha}^2)$. Here, we get $\widehat{\alpha} = \widehat{\rho}(1) = -0.390$ and $\widehat{\sigma}^2 = 120.266$.
Determining parameters with the help of R:

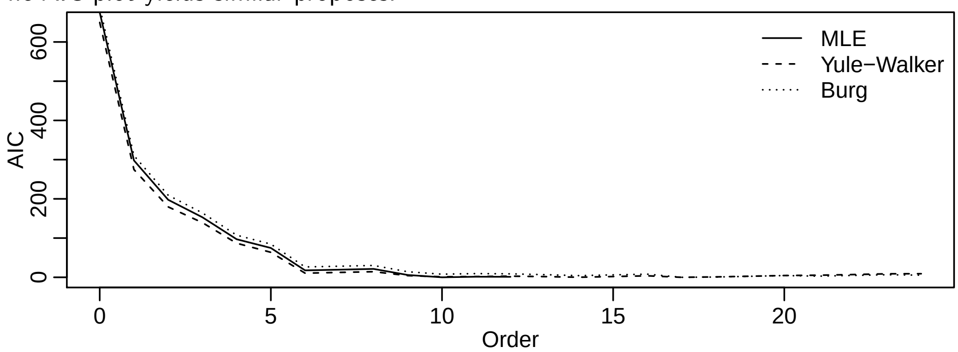
```
> r.yw <- ar(t.yields, method = "yw", order.max = 1)
> r.yw$ar
[1] -0.3898783
> r.yw$var.pred
[1] 122.0345
```

2. **a)** The experimental setup leads us to expect a period of 2 seconds. Since our measurements are spaced apart by 0.15 seconds, each 2-second period covers $2/0.15 = 13.3$ measurements. This period, however, is subject to fluctuations, which are visible in both the time series plot and the correlogram of ordinary autocorrelations:



b) The PACF (see Part a)) is clearly significant for lags up to 6; also lags 9, 10 and 17 are slightly significant. We could therefore use an AR model of order 6, 9 or 17. Order 17 seems quite high (and hence difficult to interpret), so orders 6 or 9 would be preferred. However, we cannot see by eye whether order 6 or 9 is really sufficient; this can be done with a residual analysis, see Part c).

The AIC plot yields similar proposes:



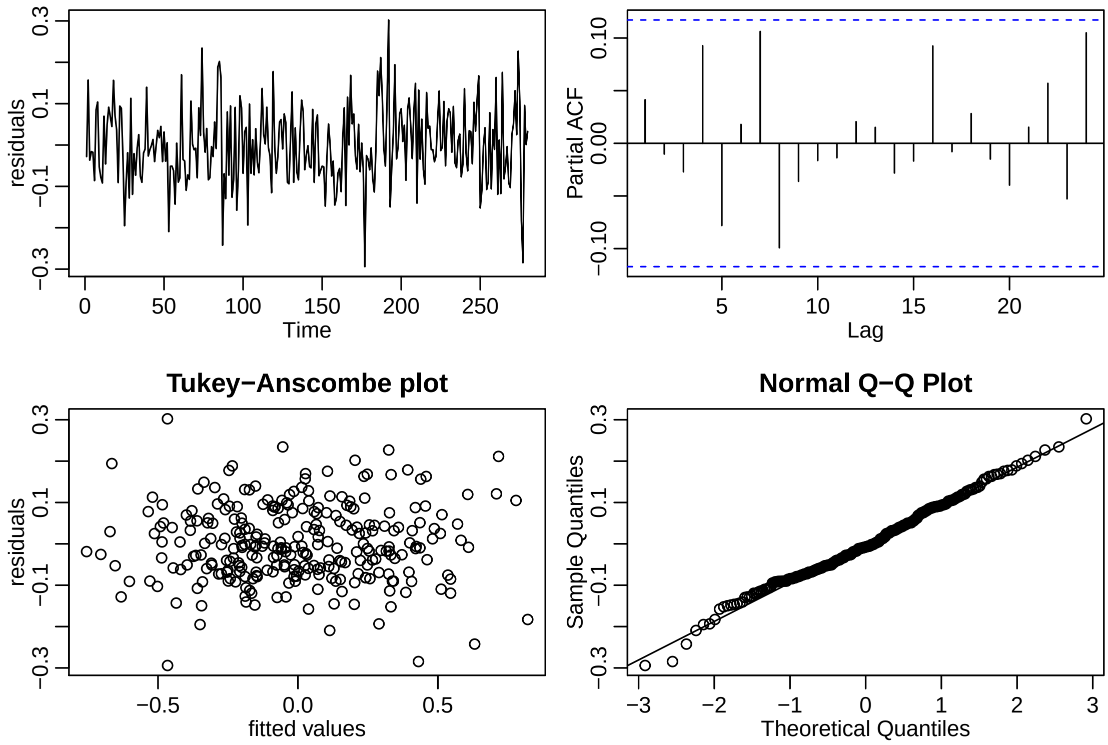
There is a big jump of the AIC at order 6, and a smaller one at order 9; hence one of these orders would be plausible to fit the given time series. However, the minimum of the AIC is attained at order $p = 10$ (MLE), $p = 17$ (Yule-Walker) or $p = 17$ (Burg), respectively.

c) Fitting the AR model:

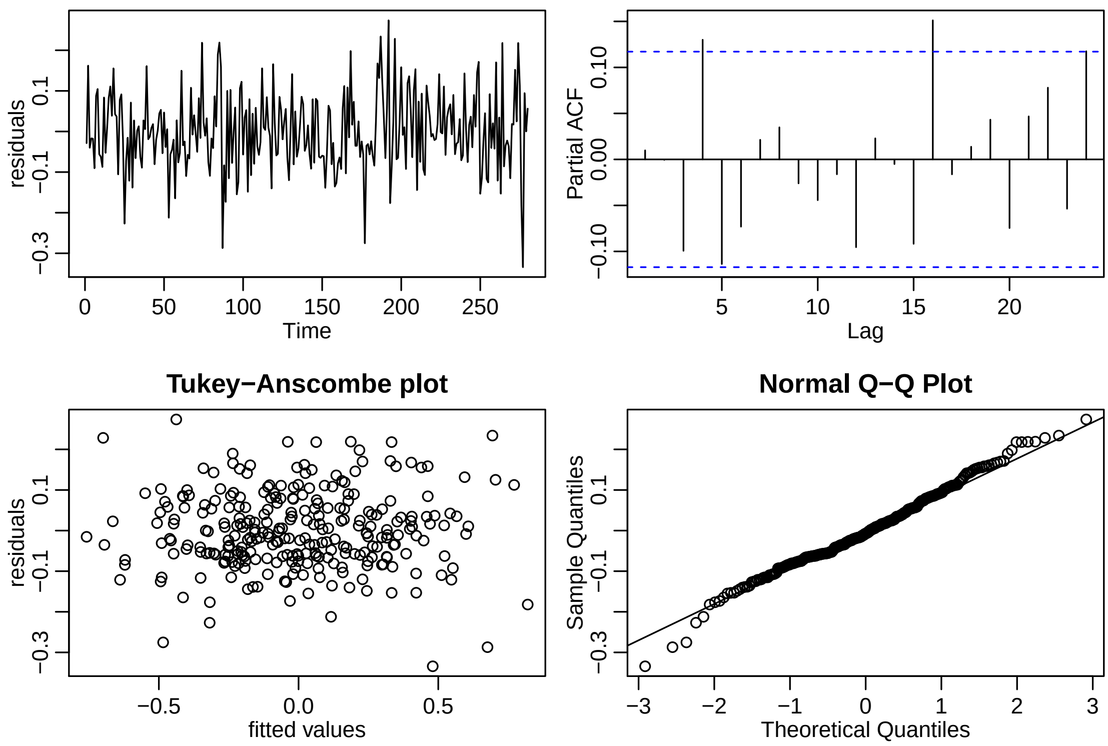
```
> p <- 9
> ar.force <- arima(ts.forceA, order = c(p, 0, 0), method = "ML")
```

The residuals of this model look as follows:

```
> par(mfrow = c(2, 2), mar = c(3, 3, 2, 0.1))
> plot(ar.force$residuals, ylab = "residuals")
> acf(ar.force$residuals, type = "partial", plot = TRUE, main = "")
> plot(ts.forceA - ar.force$residuals, ar.force$residuals, xlab = "fitted values",
   ylab = "residuals", main = "Tukey-Anscombe plot")
> qqnorm(ar.force$residuals)
> qqline(ar.force$residuals)
```



The model with order 9 is acceptable. Its residuals are normally distributed, they have constant variance, and the correlogram as well as the Tukey-Anscombe plot do not indicate any dependence. If we take order $p = 6$ instead of order $p = 9$, the residuals look different. They still show some (weak) correlation, indicating that the order is not sufficient for this time series:



d)
```
> force.pred <- predict(ar.force, n.ahead = 40)
> plot(window(ts.force, start = 250), ylab = "")
> lines(force.pred$pred, lty = 2)
```

```
> lines(force.pred$pred + 1.96*force.pred$se, lty = 3)
> lines(force.pred$pred - 1.96*force.pred$se, lty = 3)
```