

## Series 5

1. During their yearly spring melt, glaciers deposit layers of sand and mud. These annual sediments, known as *varves*, can be reconstructed in New England for the whole time between the beginning (about 12'600 years ago) till the end (6'000 years ago) of glacial retreat. From these varves, approximations of paleoclimatic parameters can in turn be computed, such as temperature (a warmer year yields more sediment).

In the dataset `varve.dat`, you will find 350 annual sediment diameters (contained in lines 201 through 550) starting at 11'660 years ago. After loading these data, first construct a time series object from them:

```
> t.url <- "http://stat.ethz.ch/Teaching/Datasets/WBL/varve.dat"
> d.varve <- ts(scan(t.url))[201:550]
```

**Comment:** The procedure `scan()` is a more general data loading function than `read.table()`. We use it here to avoid putting the data into a data frame. Do not worry about the exact choice of procedure for reading data here: simply believe us when we say that `scan()` does what we need, or read the help file.

- a) Draw a time series plot. If need be, transform the original data. Is this a stationary time series? Give reasons for your conclusion.

**R hint:** `diff()`

- b) Choose a suitable model that fits the data. Which model have you chosen? Check the convergence of the algorithm and write down the model with its estimated coefficients.

**R hints:**

```
> r.varve <- arima(log(d.varve), order=c(..., ..., ...))
> str(r.varve)
> f.acf(resid(r.varve), ...)
```

- c) Does your model fit? Analyze the residuals and comment on the results.

**R hints:**

```
> t.resid <- resid(...)
> qqnorm(...)
> qqline(...)
```

**Source:** R. Shumway, D. Stoffer, Time Series Analysis and its Applications, 2000.

2. There is a study on the development of beluga whales that focusses on the nursing behaviour of mother and calf. During a total of 160 time periods (each lasting 6 hours) subsequent to birth, the following variables were observed for "Hudson", a beluga calf. Zoologists use this data to ascertain the health of this young whale.

PERIOD	Index of the time period
BOUTS	Square root of the number of nursing bouts
LOCKONS	Square root of the number of lock-ons ("docking attempts")
DAYNIGHT	Day (= 1, 8am – 8pm) or night (= 0, 8pm – 8am) indicator
NURSING	Square root of the number of seconds spent successfully nursing during the period.

A nursing bout is defined as a successful nursing episode where milk was obtained. We would like to model the nursing time by means of the other variables. Count variables have already undergone a square root transformation to stabilize their variance ("first-aid-transformation"). You will find the data in the file `beluga.dat`.

Load the data in the usual way and create a time series matrix:

```
> d.beluga <- read.table("ftp://.../WBL/beluga.dat", header=T)
> d.beluga <- ts(d.beluga)
```

- a) Fit the model

$$\text{NURSING} = \beta_0 + \beta_1 \text{PERIOD} + \beta_2 \text{BOUITS} + \beta_3 \text{LOCKONS} + \beta_4 \text{DAYNIGHT}$$

using ordinary linear regression. Check the independence of the residuals. What conclusions can zoologists draw from this analysis?

**R hint:** `f.acf(...)`

- b) Due to the correlations involved, an AR model should be assumed for the residuals. Determine the order  $p$  of this model, and estimate the parameters  $\alpha_1, \dots, \alpha_p$ .
- c) What transformation should you apply to obtain a linear model with independent errors? State it as a formula.
- \* How would you perform this transformation (or these transformations) in R? Use the transformed time series to carry out another regression, and look at the correlation structure for the errors!

**R hint:** `lag()`

- d) Estimate the regression coefficients and the AR parameters using Maximum Likelihood.

**R hints:**

```
> library(nlme) # Load the package containing the procedure gls()
> r.bel.gls <- gls(NURSING ~ BOUITS + LOCKONS + DAYNIGHT + PERIOD,
+               data = d.beluga, correlation = corARMA(form= ~ PERIOD,
+               p = r.burg$order, q = 0, fixed = FALSE), method = "ML")
> summary(r.bel.gls)
> d.resid <- ts(resid(r.bel.gls))
> f.acf(d.resid)
```

To ensure convergence of the algorithm, known estimates of the AR parameters can be passed to `corARMA()` as starting values using the optional argument `values`. In this particular case, this does not change the outcome.

```
..., correlation = corARMA(..., value = r.burg$ar, ...), ...
```