Varve Data Analysis

Linda Frossard and Aurélie Glerum

24th January 2009

Summary

This report presents the analysis of the evolution of the thickness of glacial varves during 634 years, starting 11,834 years ago. Glacial varves are of interest for paleoclimatological research because their thickness varies with the climate at the time of formation. Due to the variation of the thicknesses in the original data, we applied a logarithmic transformation to the latter and differenced them. A significant correlation at lag 1 in the correlogram suggested fitting at least an MA(1) model to the differenced log data. Finally, an ARMA(1,1) appeared to give the best fit. This was confirmed by one-step-ahead predictions.

1 Initial data analysis

The data we study in this report are the measurements of the thickness of glacial varves, which are sedimentary deposits in a lake due to glacier melting. Normally, a varve consists of two layers: a thick brighter layer of silt and sand which is deposited in spring and summer; and a thin dark layer of clay which is deposited during autumn and winter. Similar to tree rings, varves provide information about the climate and its changes at the time of their formation. The importance of glacial varves for paleoclimatological studies has been discoverd by the Swedish scientist Gerard De Geer in the late 19th century. He and later his student Ernst Antevs studied varves formed during the last deglaciation in Scandinavia and the east coast of North America. Since the method of counting varves at the time was not very accurate because quite subjective and some of Antevs' arguments did not seem plausible to his colleagues, the study of varves has for a long time been neglected by paleoclimatologists. But with the new computer and photographic technologies, glacial varves have again become a research topic of interest (e.g. Shumway and Verosub, 1992).

The data we use are from the last deglaciation at one location in Massachusetts and were recorded by Antevs. Unfortunately we did not find any unit of measurement. The most plausible unit to us would be millimeters but this is just a guess. Today varves are determined by counting the occurence of certain chemical elements, such that the thickness might just be proportional to these counts. The starting point is 11,834 years ago and the data are collected for a period of 634 years with one observation per year. We took these data from Shumway and Stoffer (2006) and its website (`http://www.stat.pitt.edu/stoffer/tsa2/index.html`), but they are actually studied in Shumway and Verosub (1992), where more details can be found.

The aim of this data analysis project is to see if there are some important changes in the varve thickness during the considered time period. Experts could then use this information to learn about the climate variation at that time.

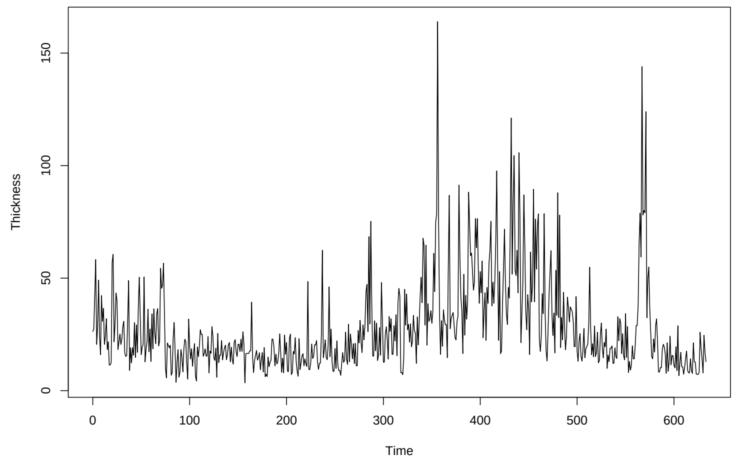

Figure 1: Varve data for a period of 634 years, starting 11,834 years ago. The unit of thickness is unknown.

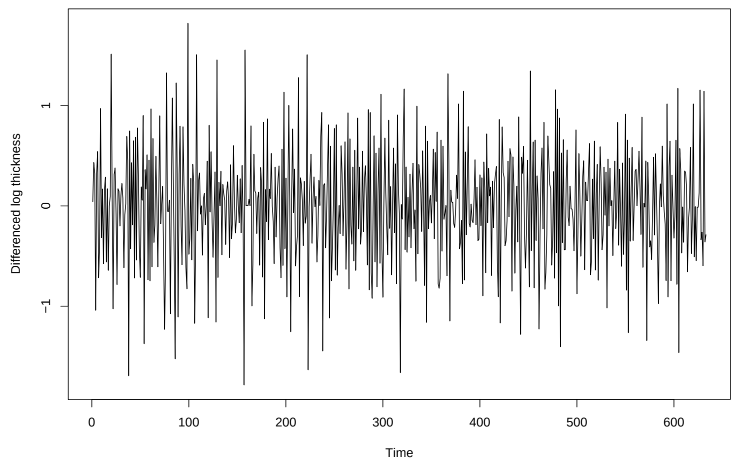

Figure 2: Differenced series of the log data.

Figure 1 shows the raw varve data. The time series is obviously not stationary. The varves take higher values of thickness between times 300 and 500 and we notice some peaks as well. Moreover, the variability seems to increase with the thickness. To remove the trend effect, we first tried to difference the time series but the result was not convincing, as there was still heteroskedasticity in the differenced data. Therefore, we decided to apply a logarithmic transformation to the varve data to balance the high and low values of thicknesses and to stabilize the variance. If we difference the time series obtained with a logarithmic transformation, we obtain the pattern shown in figure 2. The new time series now looks stationary, which enables us to apply ARMA models to it.

There does not seem to be any seasonality effect in the data. The periodogram of the raw data plotted in figure 3 on the left confirms this as we do not observe any peak. We notice that taking the difference of the log data has removed the trend effect in the original data (see figure 3 in the middle and on the right) as the spectrum is low for small frequencies. The smoothed periodogram of the differenced log data does not show any important peaks either, which lets us conclude that there are no cyclic variations in the differenced log data.

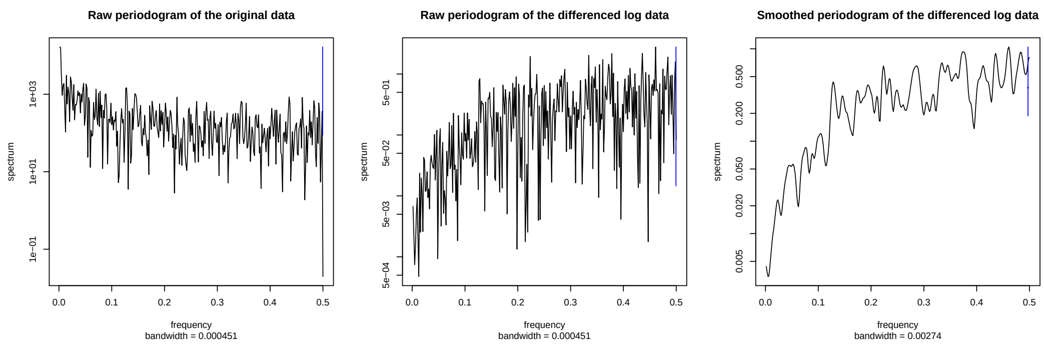

Figure 3: Periodograms of the original data (left), the differenced log data (middle) and a smoothed version of the periodogram of the differenced log data (right).

The differenced log data follow a distribution which is not too far from a normal one. Figure 4 shows a histogram of the differenced log data and a Q-Q plot of them. In the histogram we observe two high peaks at about 0 and 0.4 but otherwise the differenced log data seem to be close to a normal distribution of mean 0 and a variance 0.55^2 (red curve). The blue curve (normal density with mean 0 and variance 0.4^2) accounts for the two peaks but the tails are too light. The peak around 0 causes the horizontal segment at 0 in the Q-Q plot.

Figure 5 shows the correlogram and partial correlogram of the differenced log data. These graphs suggest us to fit an MA(1) model to the differenced log data, as the correlogram shows a nonnegligible correlation at lag 1. Moreover, the partial correlogram tails off, which supports fitting an MA model. Nevertheless, we will try to fit other models, such as MA models of higher orders or ARMA models.

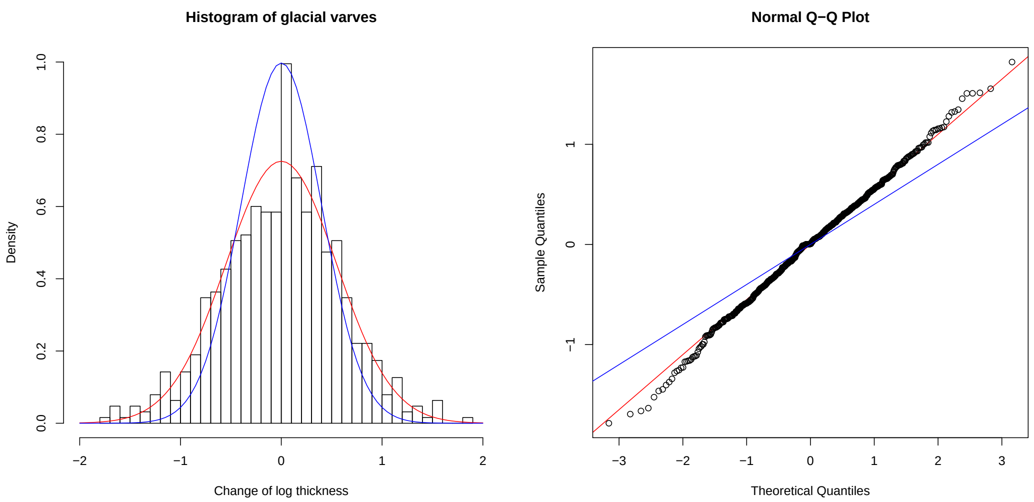

Figure 4: Histogram (left) and Q-Q plot of the differenced log data (right). The red lines correspond to a normal distribution function with mean 0 and variance 0.55^2. The blue lines correspond to a normal distribution function with mean 0 and variance 0.4^2.

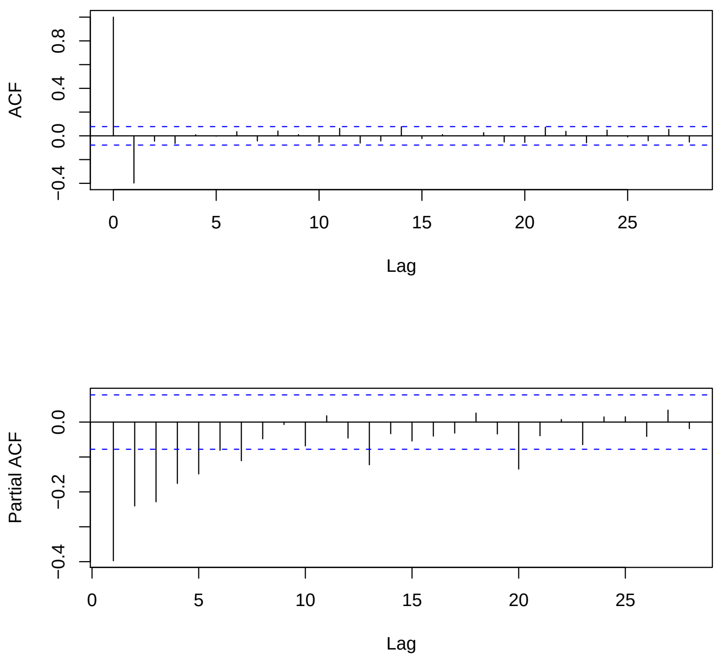

Figure 5: Correlogram (top) and partial correlogram (bottom) of the differenced log data.

2 Methods

The slide numbers in this section all refer to the lecture notes of Davison (2008).
We chose to fit an ARIMA (autoregressive integrated moving average) model to our data because these models are quite flexible and therefore often provide good fit without using too many parameters (slide 151). An ARIMA(p, d, q) model for a time series $\{Y_t\}$ can be written as

$$\phi(B)(I - B)^d Y_t = \theta(B)\varepsilon_t,$$

where B is the backshift operator (defined by $B^k Y_t = Y_{t-k}$ for $k = 0, 1, \ldots$) and $\{\varepsilon_t\}$ is Gaussian white noise. If $\{X_t\}$ denotes the series of the dth difference of Y_t, i.e. $X_t = (I - B)^d Y_t$ for all t, an ARIMA(p, d, q) model of $\{Y_t\}$ is basically the same as an ARMA(p, q) model of $\{X_t\}$. Thus the series $\{X_t\}$ has to be stationary, whereas $\{Y_t\}$ does not. The operators $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average operators of orders p and q respectively (slides 145 and 146). In fact, the ARIMA(p, d, q) model of the observations $\{Y_t\}$ generalises some important classes of models:

- If $d = 0$, the model reduces to an ARMA(p, q) model of $\{Y_t\}$ (defined on slide 148).
- If $p = 0$, we have an MA(q) model of $\{X_t\}$.
- If $q = 0$, we have an AR(p) model of $\{X_t\}$.

The crucial part of the fitting of an ARIMA(p, d, q) model is the choice of the orders p, d and q. A way of how to do this is described on slide 169. In particular the empirical autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary series $(I - B)^d Y_t$ are useful to get an initial guess of p and q, due to the properties of the (P)ACF of AR(p) and MA(q) processes.

To fit an ARIMA model to our data, we used the function `arima` of the `R` package (R Development Core Team, 2007). It fits the model using the maximum likelihood method. The computations are done by using a state-space representation of the ARMA process and a Kalman filter. The basic ideas of this can be found on slides 304-305 but for more details, one should refer to the help of the function `arima` and references therein. The results provided by the `arima` function are estimates and estimated standard errors of the coefficients of $\phi(B)$ and $\theta(B)$ as well as the maximised likelihood and the value of the AIC (Akaike information criterion, see slide 197). Note that the function `arima` does not include an intercept into the model if $d \geq 1$, whereas this is possible if $d = 0$.

To check the significance of the coefficients, we use the asymptotic normality of the maximum likelihood estimator which yields the 95% confidence interval $\hat{\theta} \pm 1.96 \cdot SE(\hat{\theta})$ (SE: standard error) for the true parameter value θ (slide 115). To compare nested models we use the likelihood ratio test described on slide 116.

If the model fits the data well, the residuals should be Gaussian white noise. To test this, we use some common time series plots, such as the (partial) correlogram and their confidence bands (based on theorem 10, slide 51), the cumulative periodogram (definition 12c on slide 79; slide 85) and the Ljung-Box test (slide 69). From these plots it can be seen whether the residuals might be white noise (in fact we only see when they are *not* white noise). To check normality of the residuals, we use a normal Q-Q plot.

In addition to the residuals being white noise, their absolute values or squares should also be uncorrelated if the model fits the data well. When there is an ARCH (autoregressive conditionally heteroskedastic) structure in the observations, the observations themselves are white noise but their squares are positively correlated. When fitting an ARMA model to such data, there remains some structure in the squared residuals because the variance of the ARMA model is constant in time. Our differenced log data, however, seem to be homoskedastic, such that the fit of ARIMA models should not leave any particular structure in the squared residuals.

Another possibility to check the fit of a model is to predict an observation from it by using all previous observations, that is predict Y_n based on the observations $Y_0, \ldots, Y_{n-1}$ under the model fitted to the complete dataset. Explained in words, we use the first 100 observations to predict the 101st one, then we use the true 101 first observations to predict the 102nd observation and so forth. On slide 207, the computation of predictions from an ARMA model can be found. Though conceptually quite easy, many steps of computation are needed. Moreover the general linear representation of the process (possibly infinite linear combination of white noise realisations) is needed to compute the prediction error (theorem 60 on slide 206), which might be rather complicated. Assuming that the predictions are normally distributed, a 95% confidence interval for the predicted observation Y^*_{n+h} $(h > 0)$ can be computed as suggested on slide 208, that is $Y^*_{n+h} \pm 1.96 \cdot \sqrt{PE_{n+h}}$ (PE: prediction error).

It is much easier to use the function `predict` for the class `arima` in the `R` package. This function computes predictions and their standard errors based on a fitted ARIMA model and given observations. Again the computations are done by using a Kalman filter on a state-space representation of the ARIMA model, for details one should refer to the help of the function `KalmanForecast` in `R`. Note that the function `predict` does not take into account the uncertainty due to the model selection and the estimation of the parameters. The first uncertainty can be accounted for by comparing predictions from different models and the second is small according to the help of the function `predict.Arima` in `R`.

3 Model-fitting

In the first part we found that the differenced log data looked stationary. We will denote this series by x_t, whereas the log series is denoted by y_t. In this section we fit different ARMA(p, q) models to the series x_t and choose the best one(s) using the methods described previously. As the correlogram of x_t shows a correlation at lag 1 (figure 5), we need to fit at least an MA(1) model. Since higher-order models might fit the data as well, we present results from fitting different MA and ARMA models. Instead of fitting ARMA(p, q) models to the differenced log series x_t, we fit ARIMA(p, d, q) models to the log series y_t with difference order $d = 1$ by using the `arima` function of `R`. Since the mean of x_t is about 0 it does not matter that there is no intercept in the fitted ARIMA models. To simplify the reading, however, we write about ARMA(p, q) models presuming that they are fitted to the series x_t.

Table 1 shows the final log likelihood and AIC values for all models we tried to fit. We attempted to fit MA models up to order 4, and ARMA models up to order $(2, 2)$. For the sake of completeness, we also put the results from AR models up to order 3 and the simple model of independent and identically distributed normal observations.

p	d	q	Log likelihood	AIC
0	1	0	−548.91	1099.83
1	1	0	−494.56	993.13
2	1	0	−475.67	957.34
3	1	0	−458.71	925.41
0	1	1	−440.72	885.44
0	1	2	−432.77	871.55
0	1	3	−430.90	869.80
0	1	4	−430.90	871.80
1	1	1	−431.44	868.88
1	1	2	−430.95	869.91
2	1	1	−430.95	869.89
2	1	2	−430.72	871.45

Table 1: Log likelihood and AIC values for different ARIMA$(p, 1, q)$ models fitted to the log data.

From the AIC value, we see immediately that the assumption of independent observations is wrong and also that the AR models do not fit well. This is confirmed by the residuals, which are not white noise for any of these models. Therefore we focus on the MA and ARMA models.

MA models

We notice that when we change the order q from 3 to 4, the log likelihood does not decrease anymore, therefore the model of order 4 does not seem to fit better the data than lower-order models. Moreover, the AIC value of model MA(4) is higher than the one of model MA(3). From this we deduce that model MA(4) is not an adequate one.

If we just take into account the AIC value, the model that fits best is the one with smallest AIC, i.e. the MA(3) model.

Let us now have a closer look at the estimations of the moving average parameters for models MA(1), MA(2) and MA(3). They are reported in table 2. Parameter θ_1 is highly

Model	θ_1	θ_2	θ_3
MA(1)	−0.771 (0.034)	–	–
MA(2)	−0.671 (0.038)	−0.159 (0.039)	–
MA(3)	−0.659 (0.040)	−0.114 (0.047)	−0.081 (0.042)

Table 2: Estimations of the moving average parameters after fitting MA(1), MA(2) and MA(3) models to the differenced logarithms of the varve data. The numbers in parentheses are the estimated standard errors of the parameters.

significant in model MA(1), but that may simply be a consequence of the fact that it is the only parameter we include in the model. If we consider the moving average process of order 2, i.e. MA(2), we obtain that the two parameters θ_1 and θ_2 are highly significant as well. But if we increase the order up to 3, the last parameter θ_3 is not significant anymore. A likelihood ratio test between models MA(3) and MA(2) confirms that model MA(3) is not significantly better than the MA(2) model.

Parameter θ_2 happened to be highly significant when we tried to fit an MA(2) process to the differenced log varve data, therefore the MA model that seems to be the most adequate is the MA(2) model, even though the analysis of the correlogram of the differenced log data only showed correlation at lag 1. We also notice that its AIC value is smaller than the one for the MA(1) model.

The MA(2) model can be written as follows:

$$x_t = (1 - 0.671_{0.038}B - 0.159_{0.039}B^2)\varepsilon_t,$$

where x_t are the differenced log series, B is the backshift operator and $\varepsilon_t \overset{\text{iid}}{\sim} (0, 0.48^2)$. The subscripts are the estimated standard errors of the parameters.

A residual analysis confirms that an MA(2) model fits well the varve data (see figures 6 and 7). The correlogram and partial correlogram show nearly no correlation between the residuals and the cumulative periodogram of the latter is within the confidence band for white noise. Moreover the p-values for the Ljung-Box statistic are quite high. The residuals follow a normal distribution with mean 0, according to the normal Q-Q plot, and with estimated variance 0.48^2. We deduce that the residuals resulting from fitting an MA(2) model are white noise. The diagnostic plots of the absolute values of the residuals show that the latter seem to be white noise as well.

A graphical analysis of the residuals of the MA(1) model showed us that the fit was less good than with an MA(2) model. In particular, the cumulative periodogram was outside the confidence band. The residual plots for the higher order MA models were about the same as for the MA(2) model, such that the latter is the MA model that fits the best the data.

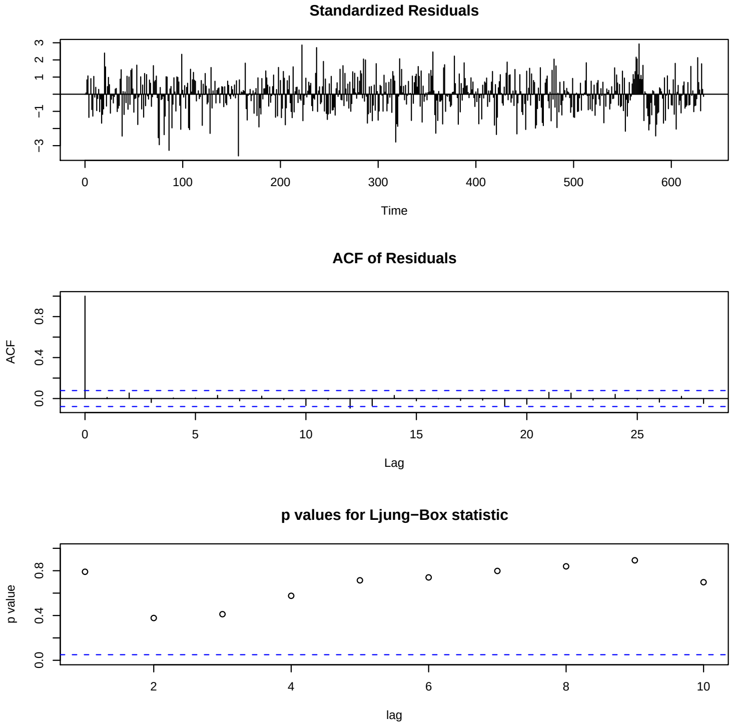

Figure 6: Diagnostic plots of the residuals of the MA(2) model. On top the residuals plot, in the middle their correlogram and below the p-values of the Ljung-Box test.

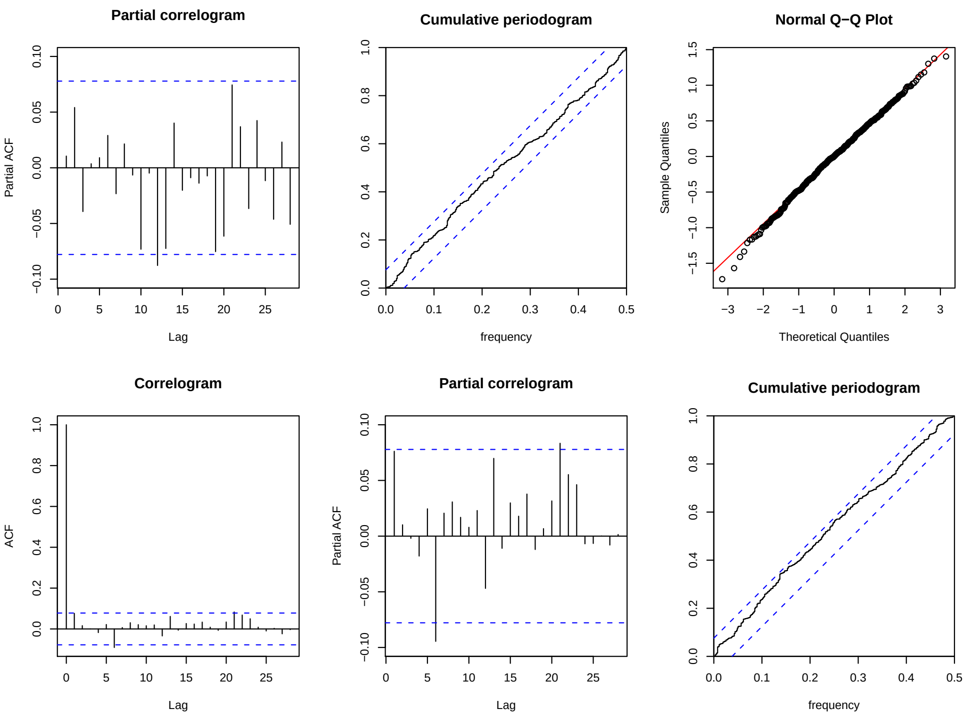

Figure 7: Diagnostic plots of the residuals of the MA(2) model (top) and their absolute values (bottom).

ARMA models

If we take into account only the AIC value of the ARMA models in table 1, the best model would be the ARMA(1,1) model. For the higher order models the AIC increases, which means that they seem to overfit the data. Especially the ARMA(2,2) model has too many parameters, as likelihood ratio tests against the ARMA(2,1) and ARMA(1,2) models show that it does not improve the fit significantly. Likelihood ratio tests of the ARMA(2,1) and ARMA(1,2) models against the ARMA(1,1) model do not show an important improvement of fit either. The second order coefficients ϕ_2 and θ_2 in table 3 are not significant, whereas the first order coefficients ϕ_1 and θ_1 always are. For the sake of parsimony we opt for the ARMA(1,1) model, which can be written as

$$(1 - 0.233_{0.052}B)x_t = (1 - 0.886_{0.029}B)\varepsilon_t, \tag{1}$$

where B is the backshift operator, x_t are the differenced log observations and $\varepsilon_t \overset{\text{iid}}{\sim}(0, 0.48^2)$.

Model	ϕ_1	ϕ_2	θ_1	θ_2
ARMA(1,1)	0.233 (0.052)	–	−0.886 (0.029)	–
ARMA(2,1)	0.244 (0.050)	0.047 (0.047)	−0.904 (0.030)	–
ARMA(1,2)	0.452 (0.221)	–	−1.114 (0.234)	0.186 (0.195)

Table 3: Estimations of the parameters after fitting ARMA(1,1), ARMA(2,1) and ARMA(1,2) models to the differenced logarithms of the varve data. The numbers in parentheses are the estimated standard errors of the parameters.

The residuals of the ARMA(1,1) model also show a good fit. Diagnostic plots are shown in figures 8 and 9. All these plots suggest that the residuals are white noise, although there is a period, starting at year 560, where there might be some positive correlation in the residuals (remember the peak in the original data at this period). The normal Q-Q plot in figure 9 suggests moreover that the residuals are normally distributed, i.e. Gaussian white noise with estimated variance of 0.48^2.

The absolute values of the residuals also seem to be white noise. The significant correlations at lag 6 in both ACF and PACF occur in fact also in the absolute values of the residuals of all other ARMA models fitted and in the MA models as well. We cannot explain this systematic appearance, because there is nothing special in the ACF and PACF of the differenced log data nor of its absolute values (plot not shown). In the PACF this correlation decreases when fitting an MA(1) model to the absolute values of the residuals, whereas it remains significant in the ACF even when fitting an ARMA(1,1) model to the residual series. However, we think that these latter models are too complicated with respect to the small approvement of fit they bring, all the more considering that the absolute values of the residuals from the ARMA(1,1) model can also be seen as white noise with the significant correlations at lag 6.

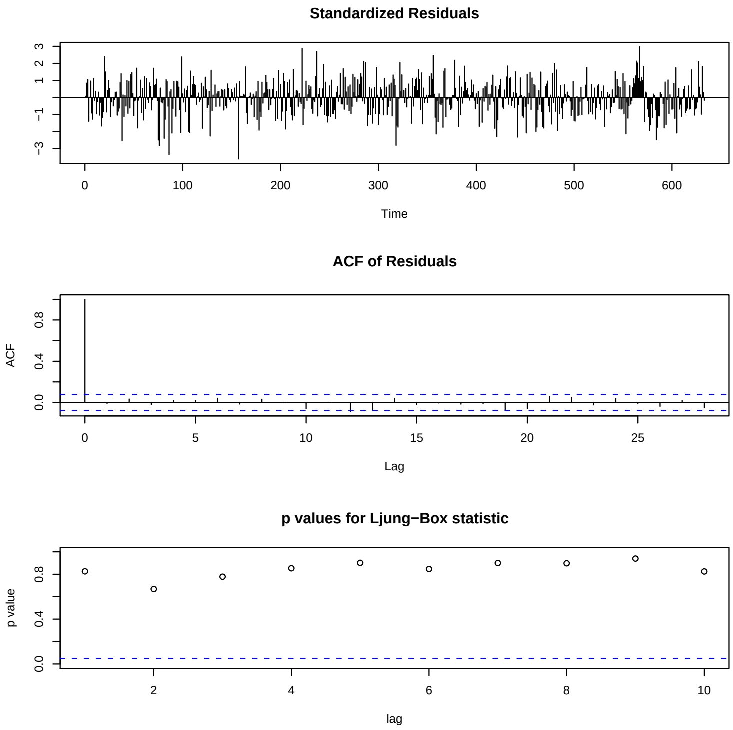

Figure 8: Diagnostic plots of the residuals of the ARMA(1,1) model. On top the residuals plot, in the middle their correlogram and below the p-values of the Ljung-Box test.

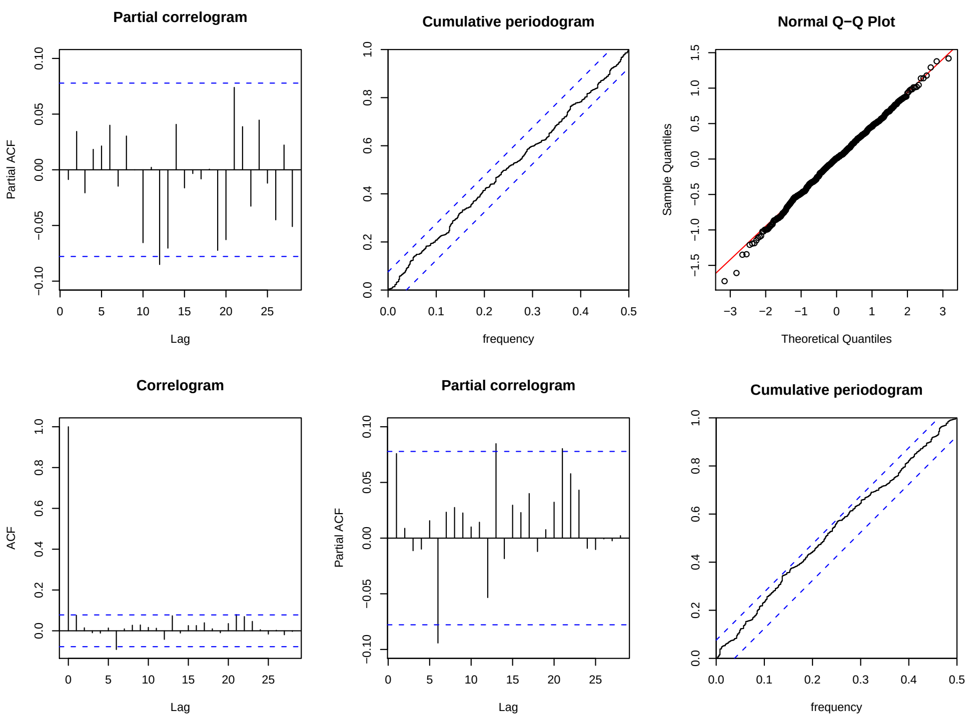

Figure 9: Diagnostic plots of the residuals of the ARMA(1,1) model (top) and their absolute values (bottom).

The analysis of residuals of the other ARMA models fitted yields the same conclusion as for the ARMA(1,1) model. Both the residuals and their absolute values can be considered as white noise. In fact the plots of figures 8 and 9 are almost the same for the other models, which leads us to choose the ARMA(1,1) model as best ARMA model.

Since the fits of the MA(2) and the ARMA(1,1) models are very similar and both have two parameters, we conclude by considering the smaller AIC value of the ARMA(1,1) given by (1) with Gaussian white noise is overall the best model for our data.

One-step-ahead predictions

As confirmation of the fit, we used the estimated ARMA(1,1) model and the log observations $y_0, \dots y_{n-1}$ to construct a 95% confidence interval for observation y_n as described in section 2. We did this for $n = 100, \dots, 633$. The result is shown in figure 10.

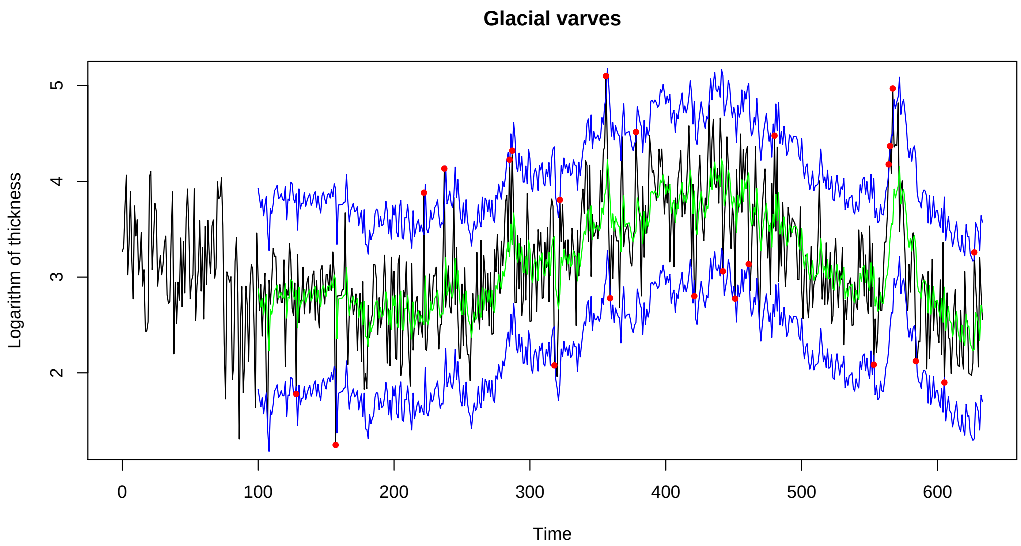

Figure 10: Plot of the log data and one-step-ahead predictions (green) of the observations at years 100 to 633 using the fitted ARMA(1,1) model (1) based on all previous observations. The blue curves are the bounds of a 95% confidence interval for the predictions. The red points are observations that lie outside the confidence interval.

There are 23 of 534 observations that lie outside their confidence interval. For a 95% confidence interval, this rate of 4.3% of the observations outside the confidence interval seems reasonable. In general, the green line of the actual predicted observations follows the important patterns in the observations quite well, although the peaks are less distinct and occur somewhat delayed. However, the model fails to predict peaks in the further future than just one step ahead. Especially the peak aroud year 570 lies outside the prediction confidence interval when predicting 15 observations based on $y_0, \dots, y_{562}$. If we base the prediction on observations $y_0, \dots, y_{567}$ (y_{567} is the first maximum of the peak), there is just the second maximum that lies outside the confidence band.

The same predictions from other models that fit the data also quite well (MA(2), MA(3), ARMA(2,1) and ARMA(1,2)) are very similar, whereas the confidence intervals are much larger for models that fit the data worse (e.g. AR(1)).

Interpretation of the model

The interpretation of the ARMA(1,1) model is complicated by the data transformation and the differencing. One way to interpret it is to use that

$$x_t = y_t - y_{t-1} = \log v_t - \log v_{t-1} = \log \frac{v_t}{v_{t-1}} \approx \frac{v_t - v_{t-1}}{v_{t-1}},$$

where v_t are the original, untransformed observations. Thus the series x_t is approximately the relative change of the varve thickness from one year to the next. According to the ARMA(1,1) model

$$x_t = 0.233x_{t-1} + \varepsilon_t - 0.886\varepsilon_{t-1},$$

this change consists – apart from a random variation ε_t – of the change in the year before (x_{t-1}) and the random variation in the year before (ε_{t-1}). The correlation between two successive observations x_t and x_{t-1} is negative (formula in example 52 of Davison, 2008), which indicates that the rate of change tends to oscillate, i.e. there seems to be no systematic increase or decrease in the change rates.

Another way of interpretation is to consider the series y_t, which is modelled as

$$y_t = 1.233y_{t-1} - 0.233y_{t-2} + \varepsilon_t - 0.886\varepsilon_{t-1}.$$

That is, the logarithm of the varve thickness can be explained by the thicknesses of the two previous varves and the random variation of the past two years. Finally, the original data $\{v_t\}$ is modelled as

$$v_t = v_{t-1}\left(\frac{v_{t-1}}{v_{t-2}}\right)^{0.233}\frac{\xi_t}{(\xi_{t-1})^{0.886}},$$

where $\{\xi_t\}$ is lognormal white noise. The current varve thickness can thus be seen as the thickness of the previous varve multiplied to some amount by the change of thickness from two years ago and a random factor from the two past years.

It seems plausible to us that the thickness of a varve is correlated with the thickness of the past varves because the climate, which is an important factor influencing the formation of varves, does not change quickly from one year to another but evaluates rather fluently over several years. Therefore climate indicators such as temperature or rainfall of successive years cannot be considered as independent.

4 Discussion

There are different ARIMA models that seem to fit the log transformed data quite well, the ARIMA(1,1,1) given by (1) being the best of them. Since the correlation at lag 6 in the absolute values of the residuals occurs in all fitted ARIMA models, it seems to be more than just random. We tried to fit different ARMA models and also an ARCH(1) model (see slide 234 of Davison, 2008) to the absolute values of the residuals from the orginal model but it did not give useful results. This problem might thus be examined further in a project of bigger scope.

The good fit of the ARIMA(1,1,1) model is confirmed by successive one-step-ahead predictions. However the high peaks in the data cannot be well predicted based only on less recent observations, which suggests that either these peaks are due to some special events or the ARIMA models are too restricted for these data. Shumway and Stoffer (2006) suggest in example 5.1 on page 274 that the log transformed series might have a long-term memory and therefore they fit an ARIMA model with fractional differencing. Another alternative would be to construct a structural model (Durbin and Koopman, 2001) that is more specific to the data. We tried to fit a local trend model to the log transformed data but the result was not satisfactory (residuals not white noise). A structural model has thus to be more elaborated and possibly it would be good to learn more about varves before starting this work. These two approaches might thus be subject of further examinations of these data.

References

Davison, A. C. (2008) Time series. Lecture notes, Mathematics Section, EPFL.

Durbin, J. and Koopman, S. J. (2001) *Time Series Analysis by State Space Methods.* Oxford University Press.

R Development Core Team (2007) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Shumway, R. H. and Stoffer, D. S. (2006) *Time Series Analysis and its Applications, with R Examples.* Second edition. Springer.

Shumway, R. H. and Verosub, K. L. (1992) State space modeling of paleoclimatic time series. In *Proc. 5th International Meeting on Statistical Climatology*, pp. 139–144. Toronto.