

August 21, 2006

On the alignment of multiple time series fragments

BY KANCHAN MUKHERJEE

Department of Mathematical Sciences, The University of Liverpool,

Liverpool L69 7ZL, U.K.

k.mukherjee@liverpool.ac.uk

ROBERT H. SHUMWAY

Department of Statistics, University of California, Davis, California 95616 U.S.A.

rhshumway@ucdavis.edu

AND KENNETH L. VEROSUB

Department of Geology, University of California, Davis, California 95616 U.S.A.

klverosub@ucdavis.edu

SUMMARY

We consider a local least-squares criterion for aligning multiple time series fragments differing by locations and show the consistency of the time-lag estimator and the asymptotic normality of the location estimator. We apply the criterion to the problem of aligning 50 glacial varve fragments and construct a 3000-year surrogate for global temperature.

Some key words: Aligning time series; Asymptotics; Global warming; Lag estimator; Local least-squares; Varve chronology.

1. INTRODUCTION

The current debate over the possible human induced contributions to rising global temperature values has underscored the need to understand historical climatic variability. Consequently, there have been attempts to obtain surrogates for temperature that extend over longer intervals of time than current global temperature records that generally go back about 150 years. These temperatures have risen steeply over the last 25 years and there is a substantial rise of about one degree Centigrade over the past 100 years.

In order to obtain longer records from a wider array of phenomena with a broader geographic distribution, it is necessary to turn to biological recorders, such as tree-rings and coral growth bands, and geological recorders, such as glacial varves. Glacial varves are laminated sediments that record annual depositional cycles in certain lakes that are fed by glaciers. The thickness of each layer represents the amount of silt and sand deposited by a melting glacier over a period of one year. In this sense, changes in varve thickness are an indication of yearly temperature changes. Therefore, reconstruction of a record of such thickness changes, known as a varve chronology, can be useful as a potential long-term proxy for paleoclimate.

As a result of the vagaries of geological depositional and erosional processes, an entire glacial varve sequence is seldom found at any particular site and the complete varve chronology must be pieced together from different outcrops or drill cores. Figure 1 shows some of the log-transformed varve thickness series collected from sites in Connecticut and Massachusetts by Antevs (1922, 1928) and Verosub (1979a, 1979b). The Connecticut series, labeled Conn. in Fig. 1(a), is a master series of length 871 years, that started at 15,600 years before present (BP), as determined by carbon dating, according to Ridge & Larson (1990). The notation 15,600 years BP means approximately that many years before 1990. The Broadbrook series is an individual site and represents those collected by Verosub. There are a total of 50 series in the database, 12 from the Antevs paper tapes collected from Connecticut, Massachusetts,

New Hampshire, New York and Vermont and 28 series collected from the same states by Verosub. The series lengths ranged from short series with only 25 years to the 871-year master Connecticut series. The total time span represented by all series, taken as a whole, is 3000 years, and this is the length of the signal that we would like to construct by aligning these 50 time series fragments.

INSERT Fig. 1 HERE

2. LAG ESTIMATION FOR TIME SERIES FRAGMENTS

Suppose $\{X_t, Y_t\}_{t \geq 1}$ is a stationary ergodic time series. Consider the following model where one observes $\{Y_t; 1 \leq t \leq n\}$ and $\{X_t; 1 \leq t \leq n_X\}$, $n_X \geq n$, satisfying

$$Y_t = \beta + X_{t+J} + \epsilon_t, \quad 1 \leq t \leq n, \quad (1)$$

where β is the unknown location parameter and J , with $0 \leq J < \infty$, is the unknown lag. It is assumed that $J \leq n_X - n$. We wish to estimate β and J . We assume that $\{\epsilon_t; 1 \leq t \leq n\}$ are independent and identically distributed with zero mean and finite variance σ^2 and that $\{\epsilon_t; 1 \leq t \leq n\}$ are independent of $\{X_t; 1 \leq t \leq n_X\}$. Typically, $\{X_t; 1 \leq t \leq n_X\}$ are taken as the observations of the master series whereas $\{Y_t; 1 \leq t \leq n\}$ is the series to be aligned.

The simplified linear model in (1) is motivated by Fig. 1, which shows that the logarithms of the series are dominated by fixed differences, with no apparent scale changes. Although general correlation methods based on the usual linear regression model and various nonlinear metrics can improve upon the above model in certain cases, as shown in a University of Bergen technical report by B. Auestad, D. Tjostheim, R. Shumway and K. Verosub, we concentrate on the simple model here in the development of asymptotics. Note that, for stationarity, we will generally be dealing with increments $\log V_t - \log V_{t-1} = \log(V_t/V_{t-1})$ of the original varve series V_t .

Alignment problems occur because the sequences may only match near the beginning or end of the master series or because varve sequences will sometimes have extra or missing values, called insertions and deletions. This difficulty is compounded by the need to consider a large number of fragments with unknown time delays. R. H. Shumway and K. L. Verosub, in a paper at the Clay Varve Chronology Workshop at Trosa, Sweden, in 1993, considered minimisation of a local least-squares measure of distance between a target series X_t and a potential matching fragment Y_t , where the series can be assumed to differ by a constant (Fig. 1). They proposed the criterion

$$s^2(j, k) = \min_{\beta} \left\{ \frac{1}{n} \sum_{t=1}^n (Y_{t+k} - X_{t+j} - \beta)^2 \right\}, \quad k \geq 0,$$

with the shift relative to a start point $1 + k$ on the matching fragment estimated by

$$\hat{j}(k) = \operatorname{argmin}_j \{s^2(j, k)\} - k.$$

INSERT Fig. 2 HERE

Figure 2 shows the local least-squares matching for just four of the 50 segments based on the local least-squares criterion with window-width $n = 30$. Consistent local matches will appear as horizontal straight lines in the $(k, \hat{j}(k))$ diagram in Fig. 2, where k is the point on the fragment and $\hat{j}(k)$ is the value on the master series. In Fig. 2(a), there is a strong match between the first Massachusetts series and the Connecticut series at $\hat{j}(k) = 450$ on the Connecticut series. This extends the record beyond the $n = 871$ years covered by the master Connecticut series. In Fig. 2(b), the match is good only for the first segment of the second Massachusetts series and the estimated start point is $\hat{j}(k) = 767$ on the Connecticut series. In Fig. 2(c), we see that the first Verosub record, called the Broadbrook series, matches at several different lags in the diagram, corresponding to possible extra varves; the middle value $\hat{j}(k) = 153$ was assigned for the series since it was difficult to tell exactly where the changes occurred. The final match between the 14th Verosub series, called the Will Mansett series,

shows matches with the second Massachusetts series at $\hat{j}(k) = 309$; 308 indicates a missing varve somewhere between $k = 40$ and $k = 60$.

In general, many ‘two-at-a-time’ comparisons were made between the 50 varve sequences in order to align them all with the master or target sequences. For example, two other Massachusetts varve sequences can serve as the target series for points beyond the Connecticut series and the first Massachusetts sequence. In general, many checks need to be made, alternating the specification of the master or target series to check for the possibility of negative lags.

For the local least-squares criterion, we will show consistency for the estimator $\hat{j}(k)$ and asymptotic normality for $\hat{\beta}$ under the assumption that X_t and Y_t are stationary ergodic series differing by white noise.

3. ON THE ASYMPTOTICS OF LAG ESTIMATION

In (1), regressing $\{Y_t; 1 \leq t \leq n\}$ on $\{X_{t+j}; 1 \leq t \leq n\}$, first we obtain the j th least-squares estimator of β , for $0 \leq j \leq n_X - n$, as

$$\hat{\beta}_j = \operatorname{argmin} \left\{ \sum_{t=1}^n (Y_t - X_{t+j} - u)^2; u \in \mathbb{R} \right\} = \bar{Y} - \bar{X}_j, \quad (2)$$

where $\bar{Y} = \sum_{t=1}^n Y_t/n$ and $\bar{X}_j = \sum_{t=1}^n X_{t+j}/n$. Next we estimate the lag J by minimising the residual sum of squares. Therefore, \hat{J} is defined as

$$\hat{J} = \operatorname{argmin} \left[\sum_{t=1}^n \{Y_t - X_{t+j} - (\bar{Y} - \bar{X}_j)\}^2; 0 \leq j \leq n_X - n \right]. \quad (3)$$

Finally, we use $\hat{\beta}_j = \bar{Y} - \bar{X}_j$ to estimate β .

We make the following assumptions about the covariance structure of the increment of the stationary process $\{X_t; 1 \leq t \leq n\}$. Define $d_{t,j} = X_{t+J} - X_{t+j}$, $D_{t,j} = (X_{t+J} - X_{t+j})^2$ and $X_{t,j} = X_{t+J} - X_{t+j} - (\bar{X}_J - \bar{X}_j)$, $0 \leq t \leq n$, $0 \leq j \leq n_X - n$.

Condition 1. For some δ_1 and δ_2 such that $\delta_1 > \delta_2 > 0$, $E(X_J - X_j)^2 \geq \delta_1$ for all $j \neq J$ and $|E(d_{s,j}d_{t,j})| \leq \delta_2$ for all $s \neq t$.

Condition 2a. We assume that $n^{-1} \sum_{j=0, \neq J}^{n_X - n} E(X_J - X_j)^2 = o(1)$.

Condition 2b. We assume that $n^{-1} \sum_{j=0, \neq J}^{n_X - n} E(X_J - X_j)^4 = o(1)$.

Condition 3. We assume that $\sum_{j=0, \neq J}^{n_X - n} g_{nj} = o(n^2)$, where $g_{nj} = \sum_{1 \leq s < t \leq n} \text{cov}(D_{sj}, D_{tj})$.

Condition 4. We assume that $n_X - n = o(n)$.

Conditions 2a, 2b and 3 are the moment conditions in the overlapping zone of the two series. Condition 4 states that the ratio of the numbers of observations in the master series and the matching series should be close to one. Note that, in Conditions 2a, 2b and 3, the sum can also include $j = J$ since the corresponding summands are zero. However, to keep conformity with Condition 1, we prefer to exclude $j = J$ in writing the sum.

Remark 1. Conditions 1-3 are satisfied when, for example, $\{X_t; 1 \leq t \leq n_X\}$ are independent and identically distributed with zero mean and finite fourth moment. Also, under a finite fourth moment assumption on X_1 , Condition 4 implies Conditions 2a and 2b.

Remark 2. Conditions 1-3 are satisfied when $\{X_t; t \geq 1\}$ is a stationary linear process represented by

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i},$$

where $\{Z_t; t \in \mathcal{Z}\}$ is a white noise process with finite fourth moment and the sum with respect to i is over the set of integers \mathcal{Z} with $\sum_i |\psi_i| < \infty$. We assume that $E(Z_t) = 0$, $\text{var}(Z_t) = \sigma_Z^2$ and $E(Z_t^4) = \eta \sigma_Z^4$ with $0 < \eta < \infty$. Thus, in particular, our results are applicable to stable autoregressive moving average processes. We verify this remark in the Appendix.

The first theorem states the consistency of the lag estimator.

THEOREM 1. *Consider model (1) and suppose that Conditions 1-4 hold. Then, for all $\epsilon > 0$,*

$$\lim_{n \rightarrow \infty} \text{pr}(|\hat{J} - J| \leq \epsilon) = 1.$$

The next result gives the asymptotic normality of $\hat{\beta}_j$.

THEOREM 2. *Consider model (1) and suppose that Conditions 1-4 hold. Then*

$$n^{1/2}(\hat{\beta}_j - \beta) \rightarrow N(0, \sigma^2),$$

in distribution, as $n \rightarrow \infty$.

4. SMALL SAMPLE PERFORMANCE

In this section we report some simulations that indicate that Theorem 1 may be true even for more general combinations of master series and fragment lengths, n_X and n , than would be implied by Condition 4. The varve-matching application involves series lengths that depart substantially from this assumption and so such simulations give more insight into this matter. To examine the performance of the matching procedure for other combinations, we fitted a first-order moving average to the first differences of the logarithms of several of the varve series,

$$x_t - x_{t-1} = w_t - \theta w_{t-1}$$

say, where the w_t were taken as independent and identically distributed normal random variables with variance σ_w^2 . In general, we obtained $\theta = 0.8$ and we used this to generate a series for X_t for use in model (1). Note that the first-order moving average model results from taking the first difference $y_{tj} - y_{t-1,j}$ in our final merging model. In simulating the values of $\{Y_t\}$ in (1), we took $\sigma^2 = 0.25$ and $\beta = 4$. The variance of the first-order moving average was taken as 0.25 so that the equivalent signal-to-noise ratio in the data generated from the model (1) was unity. Table 1 shows the results from 500 repetitions for each combination of n_X and n . For convenience in the simulations, the true match between the fragment and the master series used a value of J such that the midpoint $n/2$ of the fragment matched the midpoint of the master series $n_X/2$; alternative placements did not change the results.

INSERT TABLE 1 HERE

The results show that, with the window-width $n = 30$ in the applications, \hat{J} is expected to perform very well for most values of n_X that are in the range of the master series used in this study. The simulations also show a gradual reduction of the success probability $\text{pr}(\hat{J} = J)$ as the length of the master series increases. One can object to the optimistic nature of the assumptions, which take the signal as a first-order moving average and draw the noise values from a normal distribution, but these are sensible for the problem under consideration.

5. DISCUSSION

The primary purpose of aligning the fragments is to reconstruct a long varve record in order to obtain a potential surrogate for paleoclimatic temperature. A sensible model that results from the above matching exercise might be

$$Y_{jt} = \beta_j + X_t + \epsilon_{jt},$$

where Y_{jt} denotes the logarithm of the varve thickness of the j th fragment and the merged logarithm of the varve signal X_t is a random walk. The observed series Y_{jt} is assumed to start at the correct time delay. Estimation of the scale factors and unknown variances $\sigma_j^2 = \text{var } \epsilon_{jt}$, as well as the additional variance of the random walk, can be done using the EM algorithm, as developed for the missing data case by Shumway & Stoffer (1982). This leads to the merged varve signal in Fig. 3 based on 50 varve fragment series. In the present case, over 90% of the varve observations are missing.

INSERT Fig. 3 HERE

Here we remark that maximising the likelihood of the above model with respect to the time delays as well as the variance parameters and the scale factors would be an alternative option to the sequential two-by-two least-squares approach considered in this paper. The

number of additional parameters, 500 times the number of possible delays in this case, made the ‘two-by-two’ procedure an easier route. It is also probable that the likelihood function will not be identifiable in this highly parameterised approach.

ACKNOWLEDGEMENT

R.H. Shumway and K.L. Verosub were supported, in part, by a grant from the U.S. National Science Foundation. The authors would like to thank the editor and three referees for their careful reading and many valuable suggestions that led to significant improvement of the paper.

APPENDIX

Technical details

This section contains sketches of proofs of the theoretical results; for detail proofs, see a University of Liverpool technical report by K. Mukherjee and R. H. Shumway.

Verification of Remark 2. In this section, we verify that Conditions 1-3 are satisfied when $\{X_t; t \geq 1\}$ is a stationary linear process represented by

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i},$$

where $\{Z_t; t \in \mathcal{Z}\}$ is a white noise process with finite fourth moment and the sum with respect to i is over the set of integers \mathcal{Z} with $\sum_i |\psi_i| < \infty$. We assume that $E(Z_t) = 0$, $\text{var}(Z_t) = \sigma^2$ and $E(Z_t^4) = \eta\sigma^4$. Under these conditions, it can be shown that, with $\gamma(k) = E(X_0 X_k)$,

$$\begin{aligned} & E(X_u X_{u+p} X_v X_{v+q}) \\ &= \gamma(p)\gamma(q) + \gamma(v-u)\gamma(v-u+q-p) + \gamma(v-u+q)\gamma(v-u-p) \\ &+ (\eta-3)\sigma^4 \sum_i \psi_i \psi_{i+p} \psi_{i+v-u} \psi_{i+v-u+q}; \end{aligned} \tag{A1}$$

see for example Shumway & Stoffer (2000, Eq. 1.142). Clearly, for every $k \geq 0$, $\alpha(k) := \sum_j |\psi_j \psi_{j+k}| < \infty$. However, we also need to assume that $\sum_k \alpha(k) < \infty$. This is satisfied for ARMA(p, q) processes.

To verify Condition 3, we use the stationarity of $\{X_t\}$ in the second and third equality and (A1) with $p = 0$, $p = 0 = q$ and $u = v$ at the fourth equality below to obtain that, for $s \leq t$,

$$\begin{aligned}
\text{cov}(D_{s,j}, D_{t,j}) &= E(D_{s,j}D_{t,j}) - \{E(X_J - X_j)^2\}^2 \\
&= E(D_{s,j}D_{t,j}) - [2\{\gamma(0) - \gamma(|J - j|)\}]^2 \\
&= E(X_s^2 X_t^2) + E(X_t^2 X_s^2) + E(X_{s+J}^2 X_{t+j}^2) + E(X_{s+j}^2 X_{t+J}^2) \\
&\quad - 2\{E(X_{s+J}^2 X_{t+J} X_{t+j}) + E(X_{s+j}^2 X_{t+J} X_{t+j}) + E(X_{t+J}^2 X_{s+J} X_{s+j}) + E(X_{t+J}^2 X_{s+J} X_{s+j})\} \\
&\quad + 4E(X_{s+j} X_{s+J} X_{t+j} X_{t+J}) - \{4\gamma^2(0) + 4\gamma^2(|J - j|) - 8\gamma(0)\gamma(|J - j|)\} \\
&= 2\gamma^2(0) + 4\gamma^2(t - s) + 2(\eta - 3)\sigma^4 \sum_i \psi_i^2 \psi_{i+t-s}^2 \\
&\quad + \gamma^2(0) + 2\gamma^2(|t - s + j - J|) + (\eta - 3)\sigma^4 \sum_i \psi_i^2 \psi_{i+|t-s+j-J|}^2 \\
&\quad + \gamma^2(0) + 2\gamma^2(|t - s + J - j|) + (\eta - 3)\sigma^4 \sum_i \psi_i^2 \psi_{i+|t-s+J-j|}^2 \\
&\quad - 2\{\gamma(0)\gamma(|J - j|) + T_1\} + \gamma(0)\gamma(|J - j|) + T_2 \\
&\quad + \gamma(0)\gamma(|J - j|) + T_3 + \gamma(0)\gamma(|J - j|) + T_4\} \\
&\quad + 4\{\gamma^2(|j - J|) + \gamma^2(t - s) + \gamma(|t - s + J - j|)\gamma(|t - s + j - J|)\} \\
&\quad + (\eta - 3)\sigma^4 \sum_i \psi_i \psi_{i+|t-s+J-j|} \psi_{i+t-s} \psi_{i+|t-s+j-J|} \\
&\quad - 4\gamma^2(0) - 4\gamma^2(|J - j|) + 8\gamma(0)\gamma(|J - j|),
\end{aligned}$$

where, for example, T_1 is defined by

$$\begin{aligned}
\gamma(0)\gamma(|J - j|) + T_1 &= E(X_{s+J}^2 X_{t+J} X_{t+j}) \\
&= \gamma(0)\gamma(|J - j|) + 2\gamma(t - s)\gamma(|t - s + J - j|) + (\eta - 3)\sigma^4 \sum_i \psi_i^2 \psi_{i+t-s} \psi_{i+|t-s+j-J|}.
\end{aligned}$$

Hence, cancelling the terms involving $\gamma^2(0)$ and $\gamma^2(|J - j|)$ and $\gamma(0)\gamma(|J - j|)$, we obtain

$$\begin{aligned}
& \text{cov}(D_{s,j}, D_{t,j}) \\
&= 4\gamma^2(t-s) + 2(\eta-3)\sigma^4 \sum_i \psi_i^2 \psi_{i+t-s}^2 \\
&\quad + 2\gamma^2(|t-s+j-J|) + (\eta-3)\sigma^4 \sum_i \psi_i^2 \psi_{i+|t-s+j-J|}^2 \\
&\quad + 2\gamma^2(|t-s+J-j|) + (\eta-3)\sigma^4 \sum_i \psi_i^2 \psi_{i+|t-s+J-j|}^2 \\
&\quad - 2(T_1 + T_2 + T_3 + T_4) \\
&\quad + 4\{\gamma^2(t-s) + \gamma(|t-s+J-j|)\gamma(|t-s+j-J|) \\
&\quad + (\eta-3)\sigma^4 \sum_j \psi_i \psi_{i+|t-s+J-j|} \psi_{i+t-s} \psi_{i+|t-s+j-J|}\}.
\end{aligned}$$

The above involves terms of the form $\gamma^2(t-s)$, $\gamma^2(|t-s+j-J|)$, $\gamma(|t-s+J-j|)\gamma(|t-s+j-J|)$, $\gamma(t-s)\gamma(|t-s+J-j|)$, $\sum_i \psi_i^2 \psi_{i+t-s}^2$, $\sum_i \psi_i^2 \psi_{i+|t-s+j-J|}^2$, $\sum_i \psi_i^2 \psi_{i+t-s} \psi_{i+|t-s+j-J|}$ and $\sum_i \psi_i \psi_{i+|t-s+J-j|} \psi_{i+t-s} \psi_{i+|t-s+j-J|}$. We next show that sum of the above terms over $0 \leq j \leq n_X - n$, $1 \leq s \leq t \leq n$ is $o(n^2)$. As representative proofs, we only show that

$$\sum_{j=0}^{n_X-n} \sum_{1 \leq s \leq t \leq n} \gamma^2(t-s) = o(n^2), \tag{A2}$$

$$\sum_{j=0}^{n_X-n} \sum_{1 \leq s \leq t \leq n} \sum_i \psi_i \psi_{i+|t-s+J-j|} \psi_{i+t-s} \psi_{i+|t-s+j-J|} = o(n^2). \tag{A3}$$

Equation (A2) follows if we note that $|\gamma(k)| \leq C_1$, so that

$$\sum_{1 \leq s \leq t \leq n} \gamma^2(t-s) \leq C_1 \sum_k |\gamma(k)| < \infty.$$

Similarly (A3) follows if we note that $\sum_i |\psi_i \psi_{i+|t-s+J-j|} \psi_{i+t-s} \psi_{i+|t-s+j-J|}| \leq C\alpha(t-s)$. Hence

$$\begin{aligned}
& \left| \sum_{j=0}^{n_X-n} \sum_{1 \leq s \leq t \leq n} \sum_i \psi_i \psi_{i+|t-s+J-j|} \psi_{i+t-s} \psi_{i+|t-s+j-J|} \right| \\
& \leq C \sum_{j=0}^{n_X-n} \sum_{1 \leq s \leq t \leq n} \alpha(t-s) \leq \sum_k \alpha(k)(n_X - n) = o(n^2).
\end{aligned}$$

Note that Conditions 2a and 2b are satisfied by the finiteness of the fourth moment. Also, since $\{X_t; 1 \leq t \leq n\}$ is stationary with autocorrelation function $\{\gamma(k)\}$, Condition 1 implies

that, for all $s \neq t$, $j \neq J$,

$$2\gamma(|t-s|) - \gamma(|t-s+j-J|) - \gamma(|t-s+J-j|) \leq \delta_2 < \delta_1 \leq 2\{\gamma(0) - \gamma(|j-J|)\}.$$

This is equivalent to saying that, for all $u, v \neq 0$,

$$2\gamma(|v|) - \gamma(|v+u|) - \gamma(|v-u|) \leq \delta_2 < \delta_1 \leq 2\{\gamma(0) - \gamma(|u|)\}.$$

For proving Theorem 1, we first summarise several implications of Conditions 1-4 in the form of the following Lemma. For its proof, see the technical report by K. Mukherjee and R. H. Shumway.

LEMMA A1. *In the model (1) suppose that Conditions 1-4 hold. Then,*

$$\text{for all } j \neq J, E\{n^{-1} \sum_{t=1}^n X_{t,j}^2\} \geq (1 - 1/n)(\delta_1 - \delta_2) = \delta_{3n} \text{ say,} \quad (\text{A4})$$

$$\sum_{j=0, j \neq J}^{n_X-n} E\{n^{-2} \sum_{t=1}^n X_{t,j}^2\} = o(1), \quad (\text{A5})$$

$$\sum_{j=0, j \neq J}^{n_X-n} E(\bar{X}_J - \bar{X}_j)^2 = o(1), \quad (\text{A6})$$

$$\sum_{j=0, j \neq J}^{n_X-n} \text{var}(\bar{D}_j) = o(1), \text{ where } \bar{D}_j = n^{-1} \sum_{t=1}^n D_{t,j}. \quad (\text{A7})$$

Note that, in (A5)-(A7), the sum can also include $j = J$ since the corresponding summands are zero. However, to keep conformity with (A4), we exclude $j = J$ from writing the sum.

Proof of Theorem 1. Since \hat{J} and J are integer-valued random variables, it is equivalent to proving that

$$\lim \text{pr}(\hat{J} = J) = 1. \quad (\text{A8})$$

Note that

$$\text{pr}(\hat{J} = J) = \text{pr}\left[\bigcap_{j=0, j \neq J}^{n_X-n} \left\{\sum_{t=1}^n (Y_t - X_{t+j} - \bar{Y} + \bar{X}_j)^2 \geq \sum_{t=1}^n (Y_t - X_{t+J} - \bar{Y} + \bar{X}_J)^2\right\}\right]. \quad (\text{A9})$$

From (1) and (2), $\bar{Y} - \bar{X}_J = \beta + \bar{\epsilon}$, where $\bar{\epsilon} = \sum_{t=1}^n \epsilon_t/n$. Hence $Y_t - X_{t+J} - \bar{Y} + \bar{X}_J = \epsilon_t - \bar{\epsilon}$.

Also, $Y_t - X_{t+j} - \bar{Y} + \bar{X}_j = \epsilon_t - \bar{\epsilon} + X_{t,j}$. Therefore, using $\sum_{t=1}^n X_{t,j} = 0$, we obtain from

(A9) that $\text{pr}(\hat{J} = J) = 1 - \text{pr}[\cup_{j=0, j \neq J}^{n_X - n} \{\sum_{t=1}^n X_{t,j}^2 + 2 \sum_{t=1}^n \epsilon_t X_{t,j} < 0\}]$ and so, for (A8), it is enough to show that

$$\lim_{j=0, \neq J} \sum_{j=0, \neq J}^{n_X - n} \text{pr}(n^{-1} \sum_{t=1}^n X_{t,j}^2 + 2n^{-1} \sum_{t=1}^n \epsilon_t X_{t,j} < 0) = 0. \quad (\text{A10})$$

Choose c such that $(\delta_1 - \delta_2)/2 > c > 0$. Fix any $j \neq J$. Then, with $\delta_{3n} = (1 - 1/n)(\delta_1 - \delta_2)$ as defined in (A4),

$$\begin{aligned} & \text{pr}(n^{-1} \sum_{t=1}^n X_{t,j}^2 + 2n^{-1} \sum_{t=1}^n \epsilon_t X_{t,j} < 0) \\ & \leq \text{pr}(|n^{-1} \sum_{t=1}^n \epsilon_t X_{t,j}| > c/2) + \text{pr}(n^{-1} \sum_{t=1}^n \{X_{t,j}^2 - EX_{t,j}^2\} < c - \delta_{3n}). \end{aligned} \quad (\text{A11})$$

This follows because (A11) can be written, with the obvious identifications, as

$$\begin{aligned} \text{pr}(X + 2Y < 0) & \leq \text{pr}(Y < -c/2) + \text{pr}(X < c) \\ & \leq \text{pr}(|Y| > c/2) + \text{pr}\{X - E(X) < c - E(X)\} \\ & \leq \text{pr}(|Y| > c/2) + \text{pr}(X - E(X) < c - \delta_{3n}), \end{aligned}$$

and $\delta_{3n} < E(X)$ from (A4).

We bound the first probability in (A11) by applying Rosenthal's inequality to the tail probability of the sum of martingale differences; see Corollary 2.1 and Theorem 2.12 of Hall & Heyde (1980). Rosenthal's inequality states that, if $\{D_t; 1 \leq t \leq n\}$ is a sequence of martingale differences with respect to an increasing filtration $\{\mathcal{D}_t; 1 \leq t \leq n\}$, $M_k := \sum_{t=1}^k D_t$ and $p \geq 2$, then there exists a constant $C = C(p)$, such that, for any $\epsilon > 0$,

$$\begin{aligned} \text{pr}(|M_n| > \epsilon) & \leq \text{pr}\left(\max_{1 \leq k \leq n} |M_k| > \epsilon\right) \\ & \leq C\epsilon^{-p} \left[\sum_{t=1}^n E|D_t|^p + E\left\{\sum_{t=1}^n E(D_t^2 | \mathcal{D}_{t-1})\right\}^{p/2} \right]. \end{aligned}$$

Fix $j \neq J$ throughout. Note that, by independence between the X 's and the ϵ 's, $D_t := n^{-1}\epsilon_t X_{t,j}$, $1 \leq t \leq n$, is a sequence of martingale differences with respect to the sigma-field

$\mathcal{D}_t := \sigma\{X_{l,j}, \epsilon_k; 1 \leq l \leq t, 1 \leq k \leq t-1\}$, $1 \leq t \leq n$. Therefore, using Rosenthal's inequality with $p = 2$, we obtain

$$\Pr(|n^{-1} \sum_{t=1}^n \epsilon_t X_{t,j}| \geq c/2) \leq C_1(c) n^{-2} \left\{ \sum_{t=1}^n (E \epsilon_t^2 E X_{t,j}^2) + E(\sigma^2 \sum_{t=1}^n X_{t,j}^2) \right\}. \quad (\text{A12})$$

To bound the second probability in (A11), note that

$$\begin{aligned} & n^{-1} \sum_{t=1}^n (X_{t,j}^2 - E X_{t,j}^2) \\ &= n^{-1} \sum_{t=1}^n [\{(X_{t+J} - X_{t+j})^2 - E(\bar{X}_J - \bar{X}_j)^2\}] - \{(\bar{X}_J - \bar{X}_j)^2 - E(\bar{X}_J - \bar{X}_j)^2\} \\ &= n^{-1} \sum_{t=1}^n (D_{t,j} - E D_{t,j}) - \{(\bar{X}_J - \bar{X}_j)^2 - E(\bar{X}_J - \bar{X}_j)^2\}. \end{aligned}$$

Hence,

$$\begin{aligned} & \Pr\{n^{-1} \sum_{t=1}^n (X_{t,j}^2 - E X_{t,j}^2) < c - \delta_{3n}\} \\ & \leq \Pr\{n^{-1} \sum_{t=1}^n (D_{t,j} - E D_{t,j}) < (c - \delta_{3n})/2\} + \Pr\{(\bar{X}_J - \bar{X}_j)^2 - E(\bar{X}_J - \bar{X}_j)^2 < (c - \delta_{3n})/2\}. \end{aligned}$$

Note that, for all large n , there are some c_1 and c_2 , such that $0 < c_1 < \delta_{3n} - c < c_2$. Hence the above probabilities are eventually bounded by

$$\Pr\{|n^{-1} \sum_{t=1}^n (D_{t,j} - E D_{t,j})| > (\delta_{3n} - c)/2\} + \Pr\{|(\bar{X}_J - \bar{X}_j)^2 - E(\bar{X}_J - \bar{X}_j)^2| > (\delta_{3n} - c)/2\}.$$

If now we apply Chebychev's inequality and Markov's inequality respectively, the sum of the above two probabilities is bounded above by

$$C_2 \text{var}(\bar{D}_j) + C_3 E(\bar{X}_J - \bar{X}_j)^2, \quad (\text{A13})$$

for some constants C_2 and C_3 . Combining (A12) and (A13), we obtain that (A11) is bounded by a constant times

$$n^{-2} \sum_{t=1}^n E(X_{t,j}^2) + \text{var}(\bar{D}_j) + E(\bar{X}_J - \bar{X}_j)^2.$$

If we sum this over $j \neq J$, (A10) now follows from (A5), (A6) and (A7).

Proof of Theorem 2. For any $x \in R$,

$$\text{pr}\{n^{1/2}(\hat{\beta}_{\hat{J}} - \beta) \leq x\} \leq \text{pr}(\hat{J} \neq J) + \text{pr}\{n^{1/2}(\hat{\beta}_J - \beta) \leq x, \hat{J} = J\}.$$

By Theorem 1, the first probability is $o(1)$. Using (1) and (2), we can easily see that $n^{1/2}(\hat{\beta}_J - \beta) = n^{1/2}\bar{\epsilon}$ and so, by the ordinary central limit theorem, $\text{pr}\{n^{1/2}(\hat{\beta}_J - \beta) \leq x\} \rightarrow \Phi(x/\sigma)$.

9. REFERENCES

- Antevs, E. (1922). The recession of the last ice sheet in New England. *Am. Geog. Soc. Res. Ser.* **11**, 120pp.
- Antevs, E. (1928). The recession of the last ice sheet in New England. *Am. Geog. Soc. Res. Ser.* **17**, 297pp.
- Hall P. & Heyde, C. C. (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.
- Ridge, J. C. & Larson, F. D. (1990). Re-evaluation of Antev's New England varve chronology and new radiocarbon dates of sediments from glacial Lake Hitchcock. *Bull. Geol. Soc. Am.* **102**, 889-99.
- Shumway, R. H. & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* **3**, 253-64.
- Shumway, R. H. & Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. New York: Springer-Verlag.
- Verosub, K. L. (1979a). Paleomagnetism of varved sediments from western New England: Variability of the paleomagnetic recorder. *Geophys. Res. Lett.* **6**, 241-4.
- Verosub, K. L. (1979b). Paleomagnetism of varved sediments from western New England: Secular variation. *Geophys. Res. Lett.* **6**, 244-8.

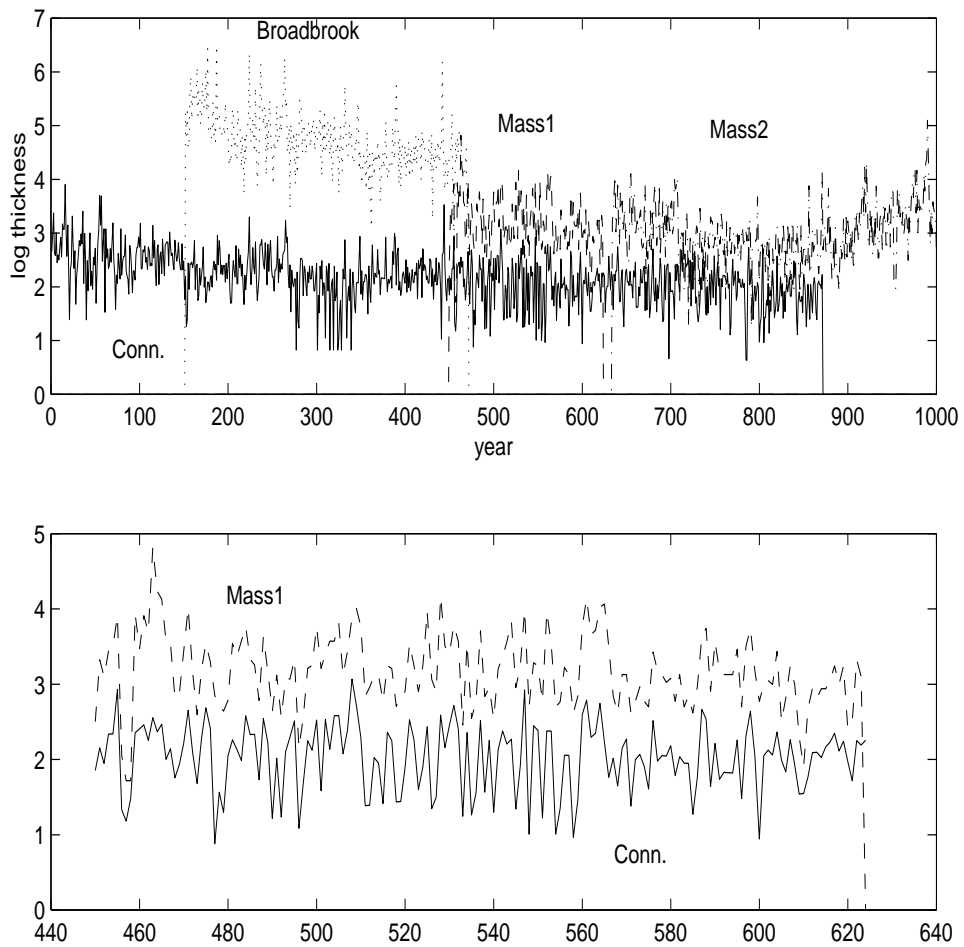


Figure 1: (a) Logarithms of yearly varve thicknesses series from several locations, beginning at 15,600 years before present. Series have been aligned by minimising squared error over possible lags. (b) The best alignment of the Massachusetts fragment and the master Connecticut sequence with a starting point beginning at 450 points into the Connecticut series.

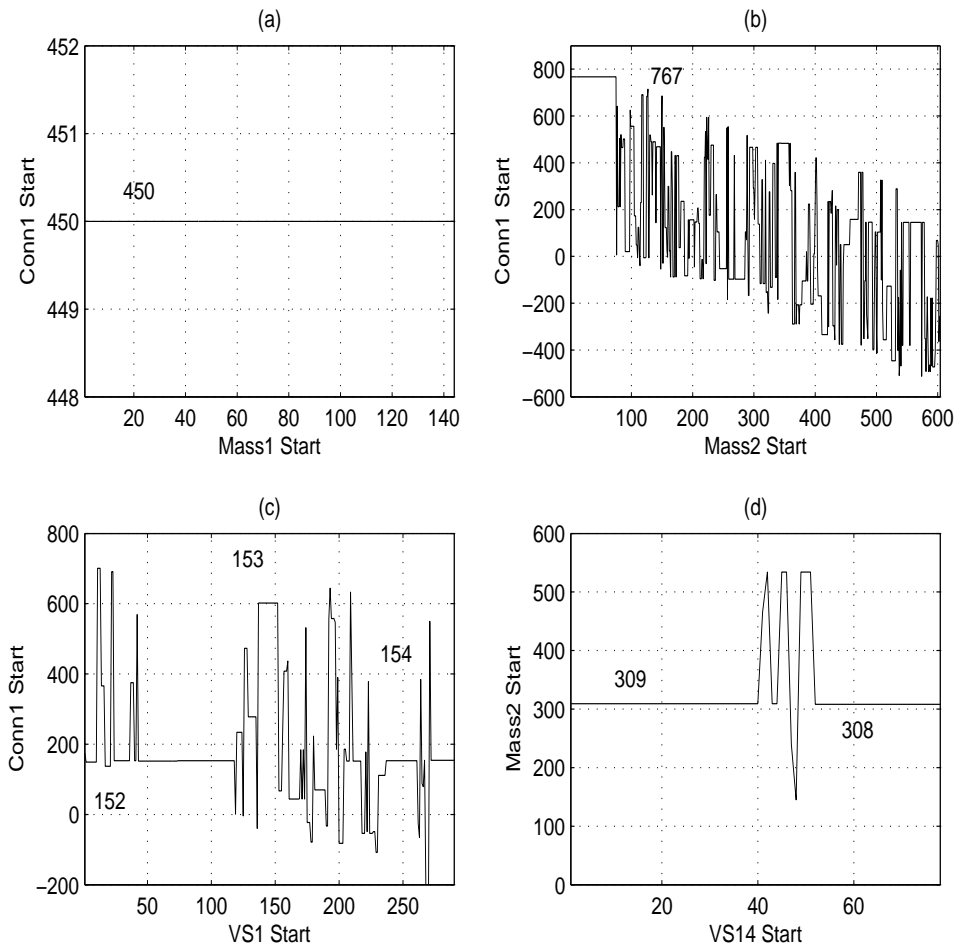


Figure 2: Local least-squares matching of varve segments to the master Connecticut segment with window-width $n = 30$. Best match $\hat{j}(k)$ is on the ordinate. (a) Connecticut-Massachusetts1 plot shows uniform agreement with a starting point at 450. (b) Connecticut-Massachusetts2 plot shows matching at the end only, year 767, of the Connecticut series. (c) Connecticut-Broadbrook plot shows varying start points ranging from 152-154, indicating missing and extra varves. (d) Connecticut-Will Mansett plot shows matches at points 309 and 308 with a missing varve.

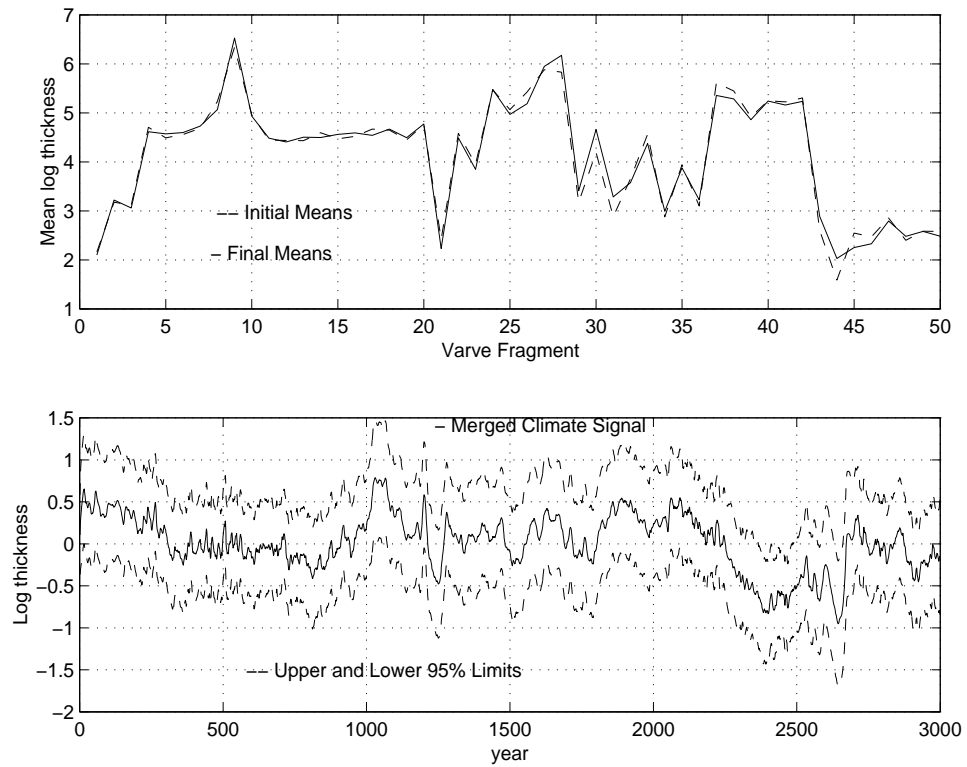


Figure 3: Reconstructed 3000-year varve signal beginning 15,600 years before present. (a) Initial and final means based on 50 estimated varve fragment series. (b) Merged signal along with 95% probability limits.

Table 1: Estimation of J based on simulated observations with various master series of lengths n_X and fragment lengths n . Values shown are $\text{pr}(\hat{J} = J)$ based on 500 repetitions at a signal-to-noise ratio of unity, with a first-order moving average used for the signal.

Fragment	Master series length		
	$n_X = 50$	$n_X = 100$	$n_X = 500$
$n = 10$	0.802	0.712	0.394
$n = 20$	0.992	0.974	0.886
$n = 30$	1.000	0.994	0.988
$n = 40$	1.000	1.000	1.000
$n = 50$	1.000	1.000	1.000