Geophys. J. R. astr. Soc. (1979) 58, 593–607

A new statistical approach to the alignment of time series

R. M. Clark *Department of Mathematics, Monash University, Clayton, Victoria, Australia 3168*
R. Thompson *Department of Geophysics, University of Edinburgh, Edinburgh EH9 3JZ*

Received 1978 December 1

Summary. Much research in the Earth Sciences is centred on the search for similarities in waveforms or amongst sets of observations. For example, in seismology and palaeomagnetism, this matching of records is used to align several series of observations against one another or to compare one set of observations against a master series. This paper gives a general mathematical and statistical formulation of the problem of transforming, linearly or otherwise, the time-scale or depth-scale of one series of data relative to another. Existing approaches to this problem, involving visual matching or the use of correlation coefficients, are shown to have several serious deficiencies, and a new statistical procedure, using least-squares cubic splines, is presented. The new method provides not only a best estimate of the 'stretching function' defining the relative alignment of the two series of observations, but also a statement, by means of confidence regions, of the precision of this transformation. The new procedure is illustrated by analyses of artificially generated data and of palaeomagnetic observations from two cores from Lake Vuokonjarvi, Finland. It may be applied in a wide variety of situations, wherever the observations satisfy the general underlying mathematical model.

1 Introduction

Measurement of the similarity between waveforms and subsequent record matching is used in many branches of the Earth Sciences, particularly seismology. Anstey (1964) describes early uses of correlation techniques applied to diverse subjects including echo ranging, well logs, record stacking and the determination of apparent velocities across seismometer arrays.

An example in the well logging field occurs in analysing palaeomagnetic and palaeolimnological data from lake sediments when it is important to compare results from several cores (Thompson & Berglund 1976). In such a situation, it is sensible to choose one particular core for which a time-scale is constructed (e.g. by radiocarbon dating), and then to match all other records against the chosen core, that is, to transform the depth-scale of each core to that of the chosen master core.

In this paper, we (1) give a mathematical and statistical formulation of the problem of 'shifting and stretching' the depth-scale of one core relative to another, (2) discuss some of the deficiencies in existing methods for doing this, and (3) describe a new statistical procedure, which we illustrate on both artificially generated and actual data.

2 Mathematical and statistical formulation

We consider the problem of transforming the depth-scale of a given core (core 1) relative to that of a proposed master core (core 2) on the basis of measurements on some *scalar* (e.g. magnetic susceptibility) which, for generality, we refer to as the 'response'. The fundamental assumption is that the variation of the response with time is identical in the two cores, and can be represented by some function H. Of course, H is not only unknown but unobservable; what we do observe is the variation of the response with depth in the two cores. This is related to H as follows. If we assume, as seems reasonable, that the depth in each core is a strictly monotonic increasing function of time, denoted by ϕ_1 and ϕ_2 for cores 1 and 2 respectively, then the variation of the response with depth d in each core is given by functions F_1 and F_2 respectively, where

$$F_1(d) = H(\phi_1^{-1}(d)) \tag{1}$$

$$F_2(d) = H(\phi_2^{-1}(d)). \tag{2}$$

Here, ϕ_1^{-1}, for example, denotes the inverse of the function ϕ_1; $\phi_1(t)$ denotes the depth in core 1 at time t, while $\phi_1^{-1}(d)$ denotes the time corresponding to depth d (in core 1). Under our assumptions of monotonicity, both inverse functions ϕ_1^{-1} and ϕ_2^{-1} are well defined and monotonic increasing.

We now denote by $g(d)$ the depth in core 2 which corresponds to the same time as depth d in core 1 (see Fig. 1). That is, $g(d) = \phi_2(t)$ if and only if $d = \phi_1(t)$ or, equivalently,

$$\phi_2^{-1}(g(d)) \equiv \phi_1^{-1}(d). \tag{3}$$

This implicitly defined function g defines the correct transformation of the depths in core 1 relative to core 2, since it follows, from equations (1), (2) and (3), that

$$F_2(g(d)) \equiv F_1(d) \quad \text{(for all } d \text{ in core 1).} \tag{4}$$

In other words, the response in core 1, plotted against the transformed depth $g(d)$, is identical to the response in core 2 plotted against the actual depth d. Our task is to find this 'stretching' function g or, equivalently, its inverse g^{-1}.

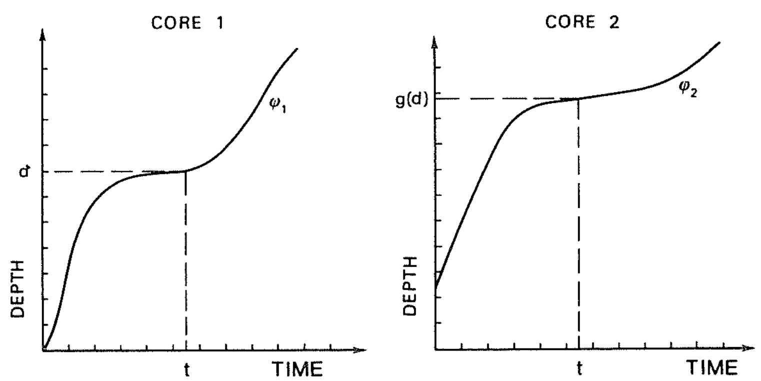

Figure 1. Schematic representation of the functions ϕ_1, ϕ_2 and the stretching function g.

Two further points need to be made before we proceed to statistical aspects of the problem. First, the assumed monotonicity of the functions ϕ_1 and ϕ_2 implies that g and its inverse g^{-1} are both well defined and monotonically increasing. Furthermore, it can be easily shown that

$$F_2(d) \equiv F_1(g^{-1}(d)). \tag{5}$$

That is, the stretching of core 1 relative to core 2 is specified, as expected intuitively, by the inverse of the function specifying the stretching of core 2 relative to core 1. Hence it does not really matter which core we call core 1.

Secondly, while the rates of sedimentation $\dot{\phi}_1(t)$ and $\dot{\phi}_2(t)$ in the two cores would have varied, in general, with time t, it may not be unreasonable, in certain cases, to assume that the sedimentation rates have been proportional to one another. In other words, one might assume, as a special case, that

$$\dot{\phi}_2(t) = \beta\dot{\phi}_1(t) \tag{6}$$

for some constant β. It then follows that $\phi_2(t) = \alpha + \beta\phi_1(t)$ and, furthermore, that

$$g(d) = \alpha + \beta d. \tag{7}$$

Thus the stretching function g is simply a linear function, with slope equal to the parameter β in equation (6). Conversely, it can be shown that equation (7) holds only if equation (6) holds. In short, linearity of g implies proportionality of the derivatives $\dot{\phi}_1$, $\dot{\phi}_2$ and vice versa. The assumption (6) of proportional rates of sedimentation would not appear to be unreasonable when the two cores in question come from the same lake, and has the appealing consequence that the stretching function g can be specified completely by its two parameters α and β. In this special case of linear stretching, we refer subsequently to α as the *lag* and β as the *stretching factor* (for core 1 relative to core 2).

In practice, all of the aforementioned functions are unknown. We have no direct information on the implicitly defined functions ϕ_1, ϕ_2, H and g, but we do have a relatively large number of measurements of the response (y_{1i}, y_{2i}) at various depths (d_{1i}, d_{2i}) in the two cores. Since these observed responses are subject to inevitable errors of measurement, we therefore assume that our two sets of paired measurements $\{(d_{1i}, y_{1i})\}$ and $\{(d_{2i}, y_{2i})\}$ satisfy the model

$$\begin{aligned} y_{1i} &= F_1(d_{1i}) + e_{1i}, \quad i = 1, 2, \ldots, n_1 \\ y_{2i} &= F_2(d_{2i}) + e_{2i}, \quad i = 1, 2, \ldots, n_2. \end{aligned} \tag{8}$$

Here, d_{ji} denotes the measured depth of the ith sample in core j, and y_{ji} the corresponding observed 'response'. The $\{e_{ji}\}$ denote random variables representing the overall error of measurement in the response, including instrumental noise, sample orientation and sediment variability. As in Clark & Thompson (1978), we assume that the depths $\{d_{ji}\}$ are known exactly, with *no* errors of measurement. Subsequently, it will be assumed that the errors $\{e_{ji}\}$ are independent and normally distributed with zero mean and some constant variance σ^2. (The validity of our various assumptions is discussed in Section 7.)

The functions F_1 and F_2 are still of unknown form, but it seems not unreasonable that both functions will be 'smooth' in some sense, e.g. at least differentiable. Our task, then, is given the observations satisfying model (8), to estimate the stretching function g defined implicitly by the relationship (4) between the two unknown but 'smooth' functions F_1 and F_2.

3 Criteria for a core-stretching method

Before considering any specific methods for doing this, it is salutary to consider what properties we should expect or demand of any core-stretching procedure. We propose the following fairly self-evident criteria:

(1) Although the method may involve certain arbitrary decisions (e.g. our decision in Section 5 to use spline functions), the method should be objective and reproducible.

(2) It should make as few assumptions as possible, and those assumptions which are made should be capable of verification from the available data.

(3) It should not matter which core is labelled as 'core 1' and which as 'core 2'.

(4) As well as giving a best estimate of the stretching function g, the method should give a statistically valid estimate of the *precision* of the estimate. That is, we should know to what extent our estimate of g is likely to differ from the correct function.

(5) The statistical procedure should not only be valid but *efficient*.

(6) The method should not require knowledge of the error variance σ^2, although an independent estimate of it may be necessary in order to check some of the assumptions.

Not all of these criteria are of equal importance. For example, a method which violates criteria 1 and 3 may still be acceptable in practice, if the dependence on the labelling of cores or on subjective judgements is negligible in practical terms. Criterion 4 is clearly crucial; there is little point in having an estimate at all unless we have a *reliable* assessment of its accuracy. Our desire to satisfy this criterion led us to reject the two existing methods described in the next section, and to develop our alternative procedure described in Section 5.

4 Existing methods

We now discuss two existing methods of estimating the stretching function g, namely (1) visual matching, and (2) cross-correlation.

4.1 VISUAL MATCHING

In this method, the plots of response versus depth for each core are scanned visually, and key features (such as local extrema) common to both plots are identified. Since these common features are due presumably to corresponding features in the underlying function H, the function g may be estimated directly by plotting against one another the estimated depths in each core corresponding to successive features (Stober & Thompson 1977).

This method is clearly subjective and not reproducible. On the other hand, it has much intuitive appeal, since g is estimated directly from its definition. We recommend that, despite its obvious limitations, this method should be used as a preliminary analysis to our more complicated method, for three reasons. First, the relative ease with which the common features can be identified will give an indication of the precision to which g is capable of being estimated by any method. Secondly, visual matching may indicate whether the function g is linear and, thirdly, it provides a valuable check on possible errors in computer programing.

4.2 CROSS-CORRELATION

This method is based on the observation that, if there were no errors of measurement, the responses at *corresponding* depths in the two cores, when plotted against one another, would lie exactly on the straight line $Y = X$. Thus if the function g defining the correspondence between depths is correct, the correlation coefficient of the corresponding paired responses should be very close to 1.

In practice, this simple idea is complicated by the need to *estimate* the response at certain depths, because of unequal spacing and the arbitrary nature of the stretching function g. To fix ideas, suppose that instead of estimating g, we want to test whether or not the data are compatible with some given pre-assigned stretching function g_0. We start by defining, for each observed depth d_{1i} in core 1, the corresponding depth $d_{2i}^* = g_0(d_{1i})$ in core 2, noting that this depth will not necessarily equal any of the *actual* depths $\{d_{2i}\}$. If y_{2i}^* denotes the estimated response at depth d_{2i}^* in core 2, derived presumably by some form of interpolation of the observed responses $\{y_{2i}\}$, the paired responses $\{(y_{1i}, y_{2i}^*), i = 1, 2, \ldots, n_1\}$ satisfy the equations

$$\begin{aligned} y_{1i} &= F_1(d_{1i}) + e_{1i} \\ y_{2i}^* &= F_2(d_{2i}^*) + e_{2i}^* \end{aligned} \qquad i = 1, 2, \ldots, n_1 \tag{9}$$

where e_{2i}^* is an appropriate error term.

Now *if* the given stretching function g_0 is correct, it follows from equation (4) that $F_2(d_{2i}^*) = F_2(g_0(d_{1i})) = F_1(d_{1i}) = \theta_i$ say, for $i = 1, 2, \ldots, n_1$. Hence y_{1i} and y_{2i}^* should be identical, if it were not for errors of measurement. The ordinary product-moment correlation coefficient of the observations $\{(y_{1i}, y_{2i}^*)\}$, which we denote by $r(g_0)$, should then be very close to 1.

In principle, we may compute the correlation coefficient $r(g_0)$ for a large number of trial stretching functions g_0, and take, as our best estimate of the actual function g, that trial function for which $r(g_0)$ is a maximum. In the special case of linear stretching, this procedure is quite straightforward, since g_0 is specified completely by the two parameters α and β of equation (7).

While this method is also intuitively appealing, it has two serious deficiencies. First, we recall that the correlation between any two sets of observations is 1 whenever those observations lie on *any* straight line; however, our aim is to test whether the observations $\{(y_{1i}, y_{2i}^*)\}$ lie on (or are close to) the *particular* line $y_1 = y_2^*$. Thus, a high correlation coefficient for the observations $\{(y_{1i}, y_{2i}^*)\}$ does *not* necessarily mean that the corresponding g_0 is correct, or even approximately correct. As a rather extreme example, if there were no errors of measurement, and if the function F_1 were linear, the correlation between $\{y_{1i}\}$ and $\{y_{2i}^*\}$ would be identically 1 for *every* possible linear stretching function g_0 i.e. for *any* α and β.

Secondly, despite the fairly strong assumptions concerning the error-terms $\{e_{1i}\}$ and $\{e_{2i}\}$ in equation (8), it is not possible to construct a statistical test to test whether the observed correlation coefficient, $r(g_0)$, is significantly different from the hypothetical correlation (ρ) of 1. The familiar t-test (Kendall & Stuart 1969, section 16.28) and z-test (Kendall & Stuart 1969, section 16.33) cannot be applied, since the former is designed to test the hypothesis $\rho = 0$ while the latter is undefined when $\rho = 1$. More importantly, both of these tests assume that the correlation coefficient is computed from a random sample from a single bivariate normal distribution. Instead, our observations $\{(y_{1i}, y_{2i}^*)\}$ satisfy a *linear functional relationship* (Kendall & Stuart 1973, chapter 29) in which successive pairs (y_{1i}, y_{2i}^*) come from different bivariate distributions, whose *means* lie on a straight line. Although Kendall & Stuart (1973, section 29.21) give an alternative t-test for this situation, this cannot be applied in our case. This modified t-test assumes that both sets of error-terms are uncorrelated, with equal variances, whereas our $\{y_{2i}^*\}$, being interpolated from the

$\{y_{2i}\}$, will be correlated amongst themselves, and have unequal variances. Thus, to sum up, it is not possible to decide statistically whether $r(g_0)$ is sufficiently close to 1 and, even if $r(g_0)$ were equal to 1, this would not necessarily guarantee that the corresponding g_0 was correct.

Three further points should be noted. First, if the above trial-and-error method is used to estimate g, the corresponding maximized correlation coefficient is bound to be large, simply because we have searched for a maximum. Box & Newbold (1971, fig. 1 and table 2) give examples in which the maximum correlation coefficient (maximized over various lags α, after de-trending) between two *completely random* series of 50 observations, is typically about 0.5. Secondly, for any given g_0, the correlation between $\{y_{1i}\}$ and $\{y_{2i}^*\}$ is not necessarily the same as that between $\{(y_{1i}^*, y_{2i}), i = 1, 2, \ldots, n_2\}$, where y_{1i}^* denotes the estimated response at the depth $d_{1i}^* = g_0^{-1}(d_{2i})$ in core 1 corresponding to depth d_{2i} in core 2. Hence the method depends on the labelling of the cores. Thirdly, the method used to interpolate the $\{y_{2i}\}$ to find each estimated y_{2i}^* will affect both the numerical value of $r(g_0)$ and the correlation structure of the $\{y_{2i}^*\}$. In some cases, it may not be possible to give a reasonable estimate y_{2i}^*, if, for example, the transformed depth $g_0(d_{1i})$ lies outside the range of the $\{d_{2i}\}$.

5 An alternative method

The idea behind our alternative method is quite simple. Suppose for the moment that our aim is to test whether or not the data are compatible with some given stretching function g_0. Imagine, then, that we superimpose the two observed records of response versus depth, *but* with the depths in one core (say core 1) converted to the corresponding depth-scale for the other core, as given by g_0. If the given stretching function g_0 is correct, both sets of data, $\{(g_0(d_{1i}), y_{1i})\}$ and $\{(d_{2i}, y_{2i})\}$, should lie on the same curve, because of equation (4). If g_0 is incorrect, the two plots will exhibit two separate curves. We can use a standard statistical test for testing whether or not the underlying curve is the same for both sets of data by making the two additional assumptions described below.

We assume, first, that for a *suitable* choice of p functions $M_1, M_2, \ldots, M_p$, the function F_2 can be written

$$F_2(x) = \sum_{j=1}^{p} \lambda_j M_j(x) \tag{10}$$

where the unknown coefficients $\{\lambda_j\}$ will be subsequently estimated by least-squares. If g_0 is correct, then combining equations (4), (8) and (10), the superimposed data $\{(d_{2i}^*, y_{1i}), (d_{2i}, y_{2i})\}$ must satisfy the equations

$$\begin{aligned} y_{1i} &= F_2(g_0(d_{1i})) + e_{1i} = \sum_j \lambda_j M_j(d_{2i}^*) + e_{1i}, \qquad i = 1, 2, \ldots, m_1 \\ y_{2i} &= F_2(d_{2i}) + e_{2i} = \sum_j \lambda_j M_j(d_{2i}) + e_{2i}, \qquad i = 1, 2, \ldots, m_2 \end{aligned} \tag{11}$$

where $d_{2i}^* \equiv g_0(d_{1i})$. (For reasons given below, it may be necessary to use subsets of only m_1 and m_2 observations from cores 1 and 2 respectively.) If, on the other hand, g_0 is incorrect, and there are two separate curves, we assume, secondly, that both of these curves are of the same form as equation (10), but with possibly different coefficients. Hence an alternative model for the data would be

$$y_{1i} = \sum_j \lambda_{1j} M_j(d_{2i}^*) + e_{1i}, \qquad i = 1, 2, \ldots, m_1$$
$$y_{2i} = \sum \lambda_{2j} M_j(d_{2i}) + e_{2i}, \qquad i = 1, 2, \ldots, m_2. \tag{12}$$

The essential point to note is that equation (11) is the special case of equation (12) in which $\lambda_{1j} = \lambda_{2j} = \lambda_j$, $j = 1, 2, \ldots, p$.

We now estimate the parameters of equations (11) and (12) by least-squares, denoting the corresponding Residual S.S.'s by R_1^2 and R_0^2 respectively. If g_0 is correct, both equations (11) and (12) should fit the data equally well, so that the difference $(R_1^2 - R_0^2)$ should be fairly small. If g_0 is incorrect, equation (12) should fit very much better than equation (11), so that the difference $(R_1^2 - R_0^2)$ will be 'large'. The crucial question is, how large? Since both equations (11) and (12) are linear models, it follows by standard statistical theory (Searle 1971, section 3.6; Seber 1977, section 4.1) that, under our assumptions, the F-statistic

$$F = \frac{(R_1^2 - R_0^2)/p}{R_0^2/(m_1 + m_2 - 2p)} \tag{13}$$

has the F-distribution on p, $m_1 + m_2 - 2p$ df. Thus the data are compatible with the given stretching function g_0 if the observed F-ratio is not significant.

The above argument holds for *any* hypothetical stretching function g_0, whether linear or not. Furthermore, the method can be readily extended to provide joint confidence limits for the lag α and stretch β in the special case of linear stretching. One simply computes the F-ratio (13) for a whole series of trial values of (α, β). By a standard statistical argument (Kendall & Stuart 1973, section 23.26), it follows that a joint 95 per cent confidence region for (α, β) is given simply by the set of values (α, β) for which the observed F-ratio (13) is *not* significant at the 5 per cent level. This confidence region provides a reliable assessment of the accuracy to which the linear stretching can be estimated from the available data, thus satisfying our most important criterion.

While the formal linear model structure of equations (11) and (12) holds whatever functions $\{M_j\}$ are used, our procedure is justified only if F_1 and F_2 can be expressed in the form (10). All that we know about F_1 and F_2, apart from the relationships discussed in Section 2, are that they are likely to be 'smooth' in some sense, e.g. differentiable. This suggests that they may be closely approximated by *cubic splines.* (For readers unfamiliar with cubic splines, they may be thought of as a series of piecewise cubic polynomials. These are carefully joined at points, known as knots, so that the function and its first and second derivatives are continuous. The four coefficients of each section are derived uniquely from the data. In this case of cubic B-splines, a least-squares minimization of goodness of fit is used.) Cubic splines are known to be the most successful and adaptable approximating functions for empirical functions such as F_1 and F_2 and, in particular, are preferable to polynomials (Wold 1974; Greville 1969). Accordingly, we recommend that the functions $\{M_j\}$ be taken as cubic B-splines defined on certain specified knots, so that F_1 and F_2 are effectively approximated by least-squares cubic splines with given knots. The user is now left with the choice of the number and location of the knots; this question is discussed in Section 7 and by Wold (1974). *Note* that the *same* knots must be used in both equations (11) and (12). Most subroutines for fitting cubic splines, such as the NAG subroutine E02BAF, require the knots to be internal to the data, and consequently it will be necessary, in such cases, to use only a subset of the observations, preferably the maximal subset corresponding to the 'intersection' or overlap of the sets $\{d_{2i}\}$ and $\{d_{2i}^*\}$. The parameter p in equations (11), (12) and (13) must be set equal to the number of internal knots *plus 4* (Cox 1972).

In certain situations, a single 'best' estimate of g will be required as well as confidence limits. In the case of linear stretching, we suggest that the best estimate of (α, β) be taken as that minimizing R_1^2. (This use of a least-squares criterion can be justified on general theoretical and statistical grounds.) Since α and β appear in equation (11) in an implicit non-linear manner (through d_{2i}^*), there is no simple equation for their least-squares estimates. However, given that confidence limits are to be found anyway, R_1^2 can be computed, at no extra effort, on the same grid of trial values (α, β), and R_1^2 can then be minimized by usual methods.

6 Applications

6.1 ARTIFICIALLY GENERATED DATA

In order to check the performance of our method, we generated several sets of data in which both the response curve F_2 and the actual stretching function g were known. To avoid the possibility of cheating, the sets of data were generated in random order, and the code identifying the various sets was not broken until the whole simulation experiment was completed. Only linear stretching was considered, and our main concern was the construction of confidence regions for (α, β) using equation (13). For brevity, we discuss three related cases, in which the stretching function was the same, namely

$$g(d) = 20 + 1.2d. \tag{14}$$

Each of three cases comprised two sets of observations $\{(d_{1i}, y_{1i}), (d_{2i}, y_{2i})\}$ (corresponding to two 'cores') satisfying equations (1), (2), (4) and (14), with

$$F_1(x) = 5.0 + \sin(0.12x + 1.5) + 0.25\sin(0.56x) \tag{15}$$

and $n_1 = n_2 = 120$. The 'depths' were approximately equally spaced over the range (1, 120) in 'core' 1 and (11, 142) in 'core' 2. The error-terms $\{e_{1i}, e_{2i}\}$ were generated as independent, pseudo-normal random variables with zero mean and equal standard deviation within each case, namely 0.15, 0.25 and 0.35 in cases 1, 2, 3 respectively. Thus the three cases corresponded to 'high', 'medium' and 'low' signal/noise ratios, in the sense that the standard deviation was respectively 0.6, 1.0 and 1.4 times the amplitude of the high-frequency component of F_1 and F_2. The pooled data for case 2 are shown, with the correct stretching, in Fig. 4.

Joint confidence regions for the lag α and stretch β of an assumed linear stretching function were obtained as described in the preceding section, using the NAG subroutine E02BAF to fit the required cubic splines by least-squares. The number of knots was chosen after some initial experiments using the methods described in the next section. Within each case, the same number of equally spaced internal knots was used for all trial values of α and β, while only those data points in the overlapping sections of the two cores were used. The resulting 95 per cent confidence regions are plotted in Fig. 2. As expected, the confidence region is smaller where the signal/noise ratio is higher, and all three confidence regions cover the point $\alpha = 20, \beta = 1.2$ defining the stretching function actually used.

Figs 4 and 5 demonstrate the remarkable sensitivity of our procedure to quite small departures from the correct stretching. These diagrams show respectively the superimposed data $\{(d_{2i}^*, y_{1i}), (d_{2i}, y_{2i})\}$ for the correct linear stretching ($\alpha = 20, \beta = 1.2$) and a slightly

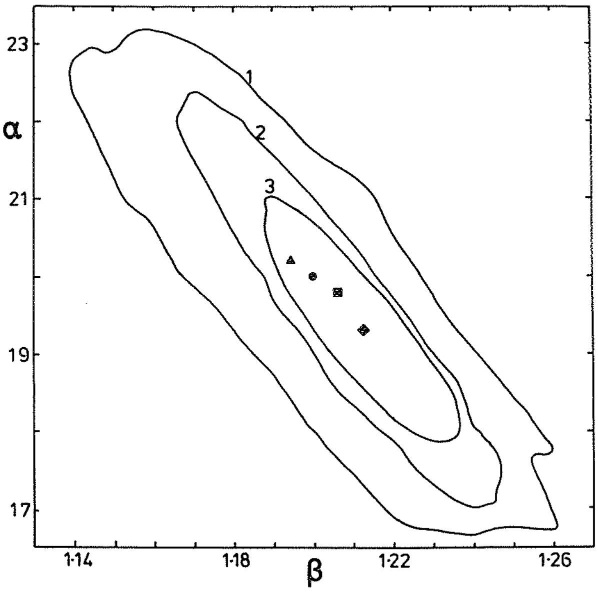

Figure 2. Joint 95 per cent confidence regions and best estimates a, b for α, β of artificial data. Circle marks 'correct match' stretching function of equation (14). Best estimate for case 1, triangle; case 2, diamond; and case 3, square.

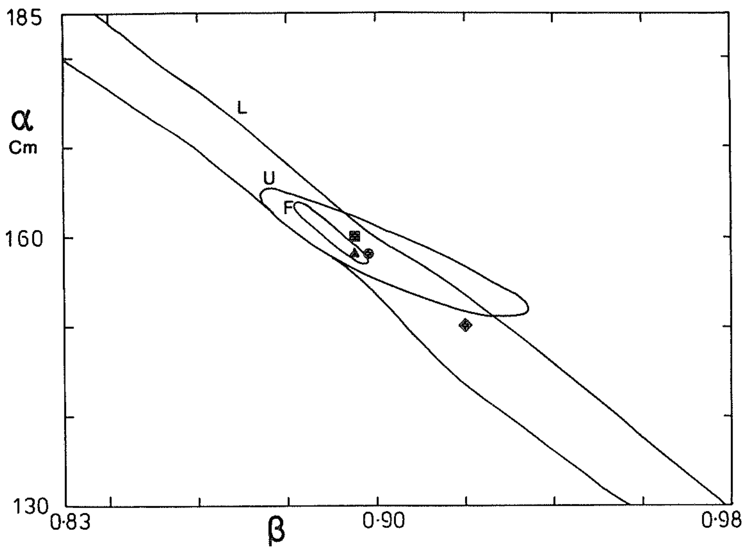

Figure 3. Joint 95 per cent confidence regions and best estimates a, b for Vuokonjarvi real data. Diamond marks α, β given by linear regression of d_2 versus d_1 graph of Stober & Thompson (1977). Best estimate for full set (F), triangle; upper set (U), square; and lower set (L), circle.

incorrect stretching ($\alpha = 17$, $\beta = 1.2$), together with the two least-squares cubic splines corresponding to equation (12). Clearly, if the two fitted splines had not been plotted, it would have been extremely difficult to decide by inspection of the plotted points which, if either, of the diagrams corresponded to the correct stretching. However, closer inspection of Fig. 5 shows that the curve corresponding to core 1 (knots plotted as squares) is almost everywhere a short distance to the left of the curve for core 2, implying that the transformed

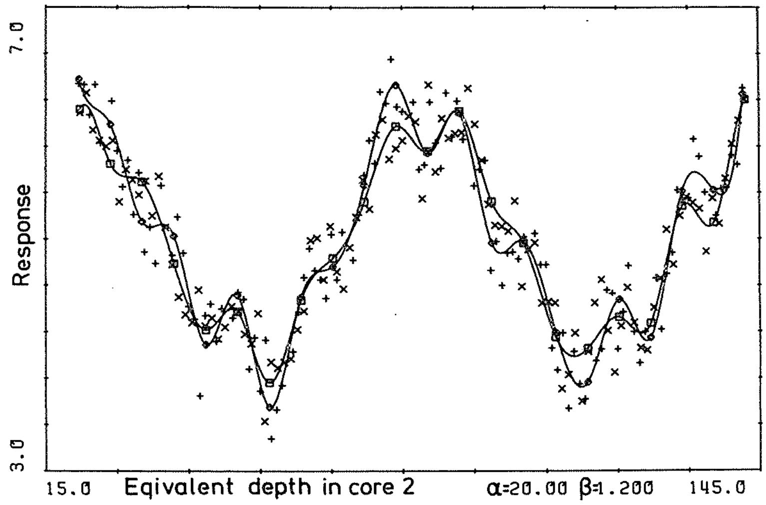

Figure 4. Case 2 of artificially generated data, with $\alpha = 20$, $\beta = 1.2$ (correct match). Data for core 2, +; data for core 1 (after transformation to core 2 depth scale), X. Cubic splines fitted to each set of data separately. Squares and diamonds mark location of knots (on curves for core 1 and core 2 respectively).

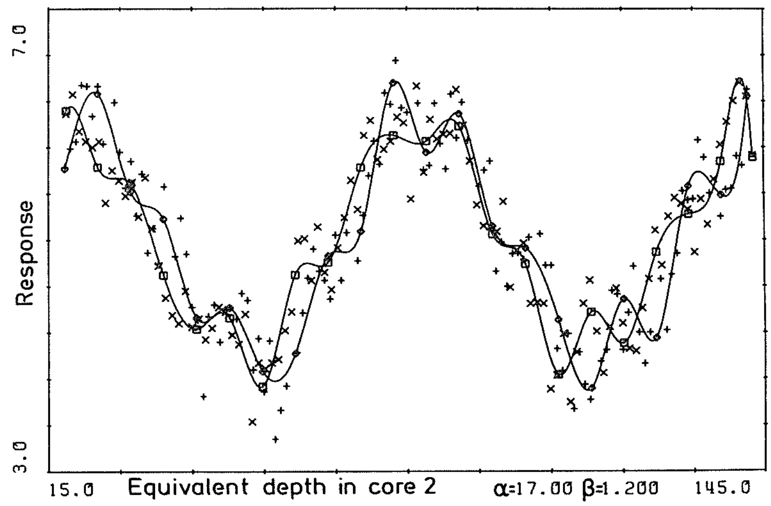

Figure 5. As for Fig. 4 but $\alpha = 17$ (poor match).

depths d_{2i}^* have not been moved far enough to the right, i.e. α is too small. This consistent error in the trial stretching of Fig. 5 is borne out by the corresponding F-ratios of 1.09 for Fig. 4 and 3.62 for Fig. 5. This latter figure is extremely significant on the relatively high df of 36 and 147. In comparison, the cross-correlation method is less sensitive, the correlation coefficient $r(g_0)$ being 0.907 for Fig. 4 and 0.835 for Fig. 5.

6.2 ACTUAL DATA

The method was then applied to two cores from Lake Vuokonjarvi, Finland (Stober & Thompson 1977) comprising 127 and 128 observations respectively. The response variable chosen was the susceptibility, on the grounds that it was: (1) independent of the palaeomagnetic directions under investigation, (2) a scalar and (3) contained physically meaningful and high-frequency components. Visual matching of 13 features of the susceptibility record in both cores produced a preliminary estimate (Stober & Thompson 1977, fig. 4) of the stretching function g. This was obtained by cubic spline interpolation of the 13 plotted depth-pairs, without assuming that g was linear. Nevertheless, this preliminary estimate (Fig. 3) was extremely close to the straight line $g(d) = 150 + 0.92\,d$.

Since both cores were taken close to one another in the same lake, the assumption of proportional sedimentation rates (6) seemed reasonable. We therefore applied the method of Section 5, assuming g to be linear, and again approximating F_1 and F_2 by cubic splines with equally spaced knots in the pooled data. The resulting 95 per cent confidence region for α and β is shown in Fig. 3 and is very tight. To check its reliability the data of core 1 was divided into an upper and lower set of 4.5 m. The residual variance of the lower set appeared to be higher than in the upper set. The upper set was roughly twice the length of the lower set. Core 2 was then separately aligned with the upper and lower groups. The best alignments (minimum R_1^2s) and confidence regions for α, β for these two separate cases are also plotted in Fig. 3. Excellent agreement is found between all the alignment estimates. The closeness of the alignments of the full and subsets and the smallness of the confidence region of the full set shows that the necessary transformation of one depth-scale to another can be determined to an adequate precision for practical purposes, and confirms that palaeomagnetic variations are reproducible in these two cores. The initial estimate of (α, β), based on visual matching, lies outside the full and upper confidence regions, highlighting the difficulty of correctly identifying, by visual means, the features of a palaeomagnetic record and the depths at which these occur.

7 Discussion of assumptions

Our various assumptions fall into two categories, i.e. those concerning: (1) the *model* and (2) the *method* of analysis. We discuss these separately.

7.1 THE MODEL

Our assumptions concerning the probability distribution of the errors $\{e_{1i}\}$ and $\{e_{2i}\}$ in our basic model (8) are the same as those made and discussed in Clark & Thompson (1978). Although the response variable y considered in that paper was declination, the same general comments are likely to hold for other palaeomagnetic parameters, such as susceptibility or intensity. The assumption of normality can be justified by appeal to the Central Limit Theorem (Cramér 1946, p. 231) and, in any case, it is known (Cochran 1947; Boneau 1960) that the F-test is robust under moderate departures for normality. The distribution of the F-statistic (13) is, however, sensitive to heterogeneity of the error variances. If the error variances are not constant but their *ratios* are known, this problem is easily overcome by using weighted least-squares fitting of the cubic splines. (The subroutine E02BAF has such a facility.)

Returning to the basic mathematics, the assumption that the functions ϕ_1 and ϕ_2 are *strictly* monotonic is essential, for it is only then that the inverse functions $(\phi_1^{-1}, \phi_2^{-1})$, the response functions (F_1, F_2) and the stretching function g itself can be unambiguously defined. The functions ϕ_1 and ϕ_2 need not necessarily be smooth. For example, if it happened that there was no disposition of sediment for several consecutive years, the functions ϕ_1 and ϕ_2 would have a 'flat-spot' at the corresponding time-interval (assuming both cores come from the same lake). It would still be possible to define monotonic inverse functions ϕ_1^{-1} and ϕ_2^{-1}, but these functions, and F_1 and F_2, would have a jump at the depth at which deposition stopped. These jumps would, in fact, help to pinpoint the stretching function g, since they would be very obvious features of the palaeomagnetic record.

The assumption of proportional rates of sedimentation seems reasonable for two cores from adjacent parts of the same lake. Fortunately, the assumption of linear stretching can be easily tested. If the actual stretching function g is distinctly non-linear, but we *assume* that it is linear, the F-ratio (13) will be significantly large for *every* possible choice of α and β, and so we have an immediate and clear indication that our assumption of linearity is incorrect. It should be noted that linear stretching is merely a special case, and that our F-statistic (13) may be used to test *any* proposed monotonic stretching function, linear or not.

On the other hand, it may happen that the correct stretching function g is very close to a straight line, but not exactly linear. It may then turn out that the F-ratio, computed under the assumption of linear stretching, is not significant, at least for certain αs and βs. The appropriate interpretation in this case is that the data $\{y_{1i}, y_{2i}\}$ are not sufficiently numerous or precise for any departure from linearity of g to be detected. In such a case, it is reasonable to act as if the function g were really linear, since this gives us the simplest adequate explanation of the data.

7.2 THE METHOD

Provided F_1 and F_2 are smooth, their approximation by cubic splines is not unreasonable. But the choice of knots is important. If there are too many knots, much of the degrees of freedom will be used up in estimating unnecessary parameters, and it is even possible that the least-squares equations will have no unique solution (Schoenberg & Whitney 1953). If there are too few knots, the cubic splines will not be an adequate representation of the functions F_1 and F_2. For example, some of the fine detail of F_1 and F_2, which may well be important in defining g, may be lost. Further, the error terms in equations (11) and (12) may contain systematic as well as random components, thereby changing the null distribution of equation (13) to a non-central F-distribution.

We recommend therefore that various choices of knots be tried following Wold's (1974) guidelines. If one has an estimate of the error variance σ^2, it is possible to test the adequacy of the cubic spline approximation to F_1 and F_2 corresponding to any given set of knots. If s^2 denotes an independent estimate of σ^2, with ν df, then the ratio $R_0^2/((m_1 + m_2 - 2p)s^2)$ should have the F-distribution on $m_1 + m_2 - 2p$, ν df. (In the unlikely event that σ^2 is known exactly, the ratio R_0^2/σ^2 would have the χ^2-distribution on $m_1 + m_2 - 2p$ df.) Both these statements hold even if the *wrong* g_0 is used in equation (12), since R_0^2 is computed by fitting two separate curves. In the likely event that several sets of knots give an adequate approximation, as judged by either criterion, one would normally use the smallest such set of knots, corresponding to the simplest adequate approximation to the data.

If the error variance cannot be estimated independently, one may use the technique of cross-validation (Stone 1974; Clark 1977) to determine an appropriate set of knots. The aim here is, effectively, to determine an appropriate value of p in the approximation (10) to F_2, and this choice should be largely independent of the estimation of g. We recommend that cross-validation be applied to core 2 *only*, first, for simplicity of computation and, secondly, because the F-ratio (13) is computed after converting all depths to the core-2 scale. If equally spaced knots are used, the appropriate number of such knots would be that corresponding to the smallest cross-validation mean-square-error (CVMSE). (If the subse-

Table 1.

m	a	b	s_1^2	s_0^2	CVMSE	χ^2	df
1	21.0	1.180	0.1826	0.1867	0.392	609.4	204
8	20.4	1.180	0.0884	0.0909	0.115	276.4	190
16	19.8	1.200	0.0728	0.0728	0.094	202.7	174
20	19.3	1.213	0.0609	0.0594	0.078	157.7	166
24	19.7	1.204	0.0570	0.0557	0.183	140.8	158
32	19.6	1.207	0.0587	0.0578	0.220	131.3	142
64	20.6	1.190	0.0580	0.0562	0.162	70.1	78
Correct values	20.0	1.200	0.0625	0.0625	–	df	–
Cross-correlation	19.9	1.186	–	–	–	–	–

m, number of internal knots. a, best estimate of α. b, best estimate of β. s_1^2, residual *mean* square for single spline at $(a, b) = R_1^2/(m_1 + m_2 - p)$. s_0^2, residual *mean* square for two separate splines at $(a, b) = R_0^2/(m_1 + m_2 - 2p)$. CVMSE, cross-validation mean square. χ^2, R_0^2/σ^2. df, degrees of freedom for $\chi^2 = m_1 + m_2 - 2p$.

quent computation of equation (13) uses only a subset of the data, this number of knots should, of course, be reduced accordingly.)

Table 1 summarizes the effect of changing the number of internal knots for case 2 of our artificially generated data. The point-estimates (a, b) of (α, β), obtained by minimizing R_1^2, are all very similar and, for example, the estimate b of the stretching factor differs by no more than 2 per cent from the correct value. The χ^2-values (based on R_0^2, and using the known value of σ^2) imply that for these data the minimum number of equally spaced interval knots to achieve an adequate fit to the data is 20. This figure is confirmed by cross-validation, as it also corresponds to the minimum CVMSE. Experiments with varying numbers of knots showed that both the confidence regions for (α, β) and the corresponding least-squares estimate of F_1 and F_2 were affected only slightly by increasing the number of knots beyond the optimum number of 20. On the other hand, we may expect the confidence region to increase considerably in size as the number of knots decreases, because of the increasing scatter of the observations about the fitted curves, as measured by s_0^2.

For case 2 of our artificially generated data, the use of 20 equally spaced knots corresponds to roughly six data points between knots on average in each core, in good agreement with the recommendation by Wold (1974) that, all other things being equal, the number of data points between knots should be at least 4 or 5. In general, the optimum knot spacing will be different for different data-sets; for the Vuokonjarvi data, cross-validation suggested a high number of knots, so following Wold (1974) 30 internal knots, i.e. an average of four data points between knots, were used. In most practical situations, the ratio R_0^2/σ^2 cannot be evaluated, since the exact value of σ^2 will be generally unknown; the corresponding χ^2-values of Table 1 were computed to demonstrate the good agreement between cross-validation and the more traditional statistical methods for determining the number of knots.

Finally, if the functions F_1 and F_2 contain jumps (due to jumps in ϕ_1^{-1} or ϕ_2^{-1}), we recommend that F_1 and F_2 be represented by a series of disjoint cubic splines, with as far as possible the same jumps. This representation is of the same form as equation (10), and so the preceding theory still applies. Although we have not yet done any experiments on this point, our conjecture is that one needs to allow for only those jumps which are large relative to the standard deviation σ, i.e. those which are obvious from a plot of the data.

8 Conclusions

The new statistical approach presented in this paper meets the major criteria listed in Section 3 for a core-stretching method. The new method is clearly reproducible, efficient (since least-squares curve-fitting is used), and produces valid confidence regions for the stretching function g. The representation of the function F_2 by cubic splines would appear to be reasonable, given the flexibility and adaptability of cubic splines for approximating empirical functions; the adequacy of this approximation can be tested from the data, provided an independent estimate of the error variance σ^2 is available. If σ^2 cannot be estimated, cross-validation may be used to determine an appropriate set of knots for the cubic splines, and a confidence region for the stretching function g may still be found using the F-statistic (13). The method may be used to investigate both linear and non-linear stretching functions, although the computations are simpler and the results easier to interpret if g is assumed to be linear. In such a case, the labelling of the cores is irrelevant; if F_2 can be represented by a cubic spline, so too can F_1 (with the same relative knot spacing), and vice versa.

Since all these major criteria are met, the new approach is clearly preferable to the existing methods of visual matching and cross-correlation. A possible disadvantage for some workers is the amount of computing involved. However, most computer installations will be able to supply efficient subroutines for fitting cubic splines by least-squares, and the amount of computer time needed is not excessive. For example, once the number of knots was chosen, production of the confidence region in Fig. 3 using the full data set from Lake Vuokonjarvi required 90 s of central-processor time on the University of Edinburgh System 4/75, at a nominal cost of £4.00.

The comparison of waveforms and the concept of record-matching to align two or more sequences of observations arises in many fields. For example, the calibration of floating tree-ring chronologies using radiocarbon dates (Clark & Sowray 1973) and the construction and extension of tree-ring chronologies (Baillie & Pilcher 1973) both involve observations satisfying the same basic statistical structure as assumed in this paper. In both these cases, the aim is to estimate simply the *lag*, α, between the two sequences; in other words, the stretching function g is, by definition, simply $g(d) = \alpha + d$. It follows that the method described in this paper may be applied to these problems, and indeed in any situation with data satisfying equations (4) and (8). In addition, our method can be easily generalized to perform k stretchings simultaneously.

Application of our new approach to the susceptibility versus depth records of the two cores from Lake Vuokonjarvi shows that the transformation from one depth-scale to another can be determined to a high precision. The directional measurements from both cores may now be superimposed using a common depth-scale, and these pooled data may be used both to test the repeatability of directional measurements from core to core and to estimate, more accurately than otherwise, the past variations in the geomagnetic field direction at Lake Vuokonjarvi. Work on these further problems is currently in progress, and we expect to report shortly on the results of this further investigation.

References

Anstey, N. A., 1964. Correlation techniques – a review, *Geophys. Prosp.*, **12**, 355–382.

Baillie, M. G. L. & Pilcher, J. R., 1973. A simple cross-dating program for tree-ring research, *Tree-Ring Bull.*, **33**, 7–14.

Boneau, C. A., 1960. The effects of violations of assumptions underlying the *t*-test, *Psych. Bull.*, **57**, 49–64.

Box, G. E. P. & Newbold, P., 1971. Some comments on a paper by Coen, Gomme & Kendall, *J. R. Stat. Soc. A*, **134**, 229–240.

Clark, R. M., 1977. Non-parametric estimation of a smooth regression function, *J. R. Stat. Soc. B*, **39**, 107–113.

Clark, R. M. & Sowray, A., 1973. Further statistical methods for the calibration of floating tree-ring chronologies, *Archaeom.*, **15**, 255–266.

Clark, R. M. & Thompson, R., 1978. An objective method for smoothing palaeomagnetic data, *Geophys. J. R. astr. Soc.*, **52**, 205–213.

Cochran, W. G., 1947. Some consequences when the assumptions for the Analysis of Variance are not satisfied, *Biometrics*, **3**, 22–38.

Cox, M. G., 1972. The numerical evaluation of B-splines, *J. Inst. Math. Appl.*, **10**, 134–149.

Cramér, H., 1946. *Mathematical Methods of Statistics*, Princeton University Press.

Greville, T. N. E., 1969. Introduction of spline functions, in *Theory and Applications of Spline Functions*, pp. 1–36, ed. Greville, T. N. E., Academic Press, New York.

Kendall, M. G. & Stuart, A., 1969. *The Advanced Theory of Statistics*, Vol. I, Distribution theory, Griffin, London.

Kendall, M. G. & Stuart, A., 1973. *The Advanced Theory of Statistics*, Vol. II, Inference and relationship, Griffin, London.

Schoenberg, I. J. & Whitney, A., 1953. On Polya frequency functions III. The positivity of translation determinants with an application to the interpolation problem by spline curves, *Trans. Am. Math. Soc.*, **74**, 246–259.

Searle, S. R., 1971. *Linear Models*, J. Wiley, New York.

Seber, G. A. F., 1977. *Linear Regression Analysis*, J. Wiley, New York.

Stober, J. C. & Thompson, R., 1977. Palaeomagnetic secular variation studies of Finnish lake sediment and the carriers of remanence, *Earth planet. Sci. Lett.*, **37**, 139–149.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions, *J. R. Stat. Soc. B.*, **36**, 111–147.

Thompson, R. & Berglund, B., 1976. Late Weichselian geomagnetic 'reversals' as a possible example of the reinforcement syndrome, *Nature*, **263**, 490–491.

Wold, S., 1974. Spline functions in data analysis, *Technometrics*, **16**, 1–11.