[RADIOCARBON, VOL. 37, NO. 3, 1995, P. 851–859]

COMPARING ^{14}C HISTOGRAMS:[1] AN APPROACH BASED ON APPROXIMATE RANDOMIZATION TECHNIQUES

TOM DYE

Historic Preservation Division, Department of Land and Natural Resources, 33 South King Street, Honolulu, Hawai'i 96813 USA

ABSTRACT. An approximate randomization technique for comparing ^{14}C histograms is described and illustrated with an example from Hawai'i. The technique determines intervals of significant difference in the histograms, rather than providing a single number summary.

INTRODUCTION

Techniques for constructing histograms from a suite of radiocarbon dates range greatly in sophistication and the ease with which they can be constructed. At the simple end are histograms of the frequencies of conventional ^{14}C ages per suitably chosen interval of the ^{14}C time scale. More sophisticated techniques take into account the standard deviations of the ^{14}C ages, express the results on the calendar time scale and employ some mechanism for minimizing the effects on the histogram of fluctuations over time in the production of ^{14}C, which manifest themselves as wiggles in the calibration curves (Stolk *et al.* 1994).

When the criteria for including ^{14}C dates in a suite are chosen suitably, variations in the height of the ^{14}C histogram constructed from the suite can be interpreted as changes in the intensity of a cultural trait. Rick (1987) provides a good explanation of the factors involved in selecting ^{14}C dates to investigate population change and his discussion can be generalized to other situations. In principle, it is possible to compile ^{14}C date suites to investigate a wide variety of natural and cultural phenomena. Examples include population change (Rick 1987; Dye and Komori 1992; McFadgen, Knox and Cole 1994), the spread of agriculture (Allen 1992) and the evolution of settlement patterns and land use (Streck 1992; Williams 1992; McFadgen, Knox and Cole 1994).

Most of these studies have interpreted patterns in individual ^{14}C histograms as reflecting the direction and pace of cultural change, but there is a growing interest in comparing histograms to gain insight into systemic relations between cultural traits over time (Williams 1992; Dye 1994). Indications of differences in the shapes of ^{14}C histograms are easy to identify through visual comparison, and this is often a useful first step in an exploratory data analysis. At some point in most analyses, however, it is useful to move beyond indication to determine whether or not observed differences are significant (Mosteller and Tukey 1977). Traditional techniques for comparing histograms, based on the chi-square distribution or the Kolmogorov-Smirnov test, yield a single measure that indicates whether two curves are significantly different, but they do not specify where significant differences occur. It is this information that is potentially most important in interpretation.

I present here an approximate randomization method for determining if the observed differences between two ^{14}C histograms are significant, or if they are due to chance factors associated with sample size. After the approximate randomization technique and its application to the determination of differences in ^{14}C histograms is discussed, the technique is illustrated with a worked example from Hawai'i.

[1]"Histogram" is established in the literature as the term that describes a fairly wide variety of graphical displays that show the distribution of a ^{14}C date suite over time. A proposed revision of terms is under discussion (Stolk *et al.* 1994: 1, note 1).

https://doi.org/10.1017/S0033822200014934 Published online by Cambridge University Press

Randomization Techniques

Approximate randomization techniques take their name from the fact that their results approximate those of the preferred exact randomization techniques (Noreen 1989). An exact randomization test generates the distribution of the test statistic by calculating its value for all permutations of the data, then compares the actual value of the test statistic with this distribution. An actual value of the test statistic that falls near a tail of the distribution is unlikely to result from chance factors alone, and so is significant. Unfortunately, exact randomization tests are only practical for small data sets and cannot approach problems the size of the example below. There are 75! permutations of the example data, a number so large that calculating the test statistic this many times is a task well beyond the capability of the most powerful computers of the present and foreseeable future. In cases such as this, the distribution of the test statistic can be approximated by a large number of randomly chosen permutations and the actual value of the test statistic compared with this approximate distribution. In this way, approximate randomization techniques can yield useful results with a modest amount of computing effort.

An appealing feature of randomization techniques is that any test statistic can be used, and the analysis is not limited to a test statistic whose distribution is known, or assumed to be known, beforehand. In practical terms, this means effort can be spent identifying the test statistic that makes the most sense in a particular situation, rather than in complex derivations of statistical distributions. In the case of comparing ^{14}C histograms, the choice of a test statistic can be guided by the research question under consideration and can vary from one instance to another. Because a test statistic is a single-valued function of the data (Noreen 1989), the application of approximate randomization techniques to the comparison of ^{14}C histograms requires that the statistic be computed at several points. The distance between the points at which the statistic is computed determines the precision with which intervals of significant difference between histograms can be identified. In the example below, the test statistic is computed at each calendar year that two ^{14}C histograms overlap; computing the test statistic at 10-yr, 50-yr or other intervals would produce identical results at the tested points, but would reduce the precision with which intervals of significant difference are identified. Although the test statistic is computed many times for each permutation of the data, in the example below the test statistic is chosen so that its calculation at each year is independent of the statistic at other years.

Implementation

The implementation here generally follows Noreen (1989), who provides a useful source code in Basic, Pascal and Fortran. Figure 1 provides a schematic illustration of the technique. The basic data are two conventional ^{14}C date suites, $z = (z_1 \pm \sigma_1, z_2 \pm \sigma_2, ..., z_m \pm \sigma_m)$ and $y = (y_1 \pm \sigma_1, y_2 \pm \sigma_2, ..., y_n \pm \sigma_n)$, selected such that z and y both relate to phenomena of interest (Fig. 1B). For convenience, their union is denoted as $w = z \cup y$. Calibrated histograms constructed from z and y take the general form $H = (x_t, x_{t+1}, ..., x_{t+p})$, where t is a date on the calendar time scale, and x is an expression of intensity (Stolk *et al.* 1994). Calibration makes use of a calibration curve, $\mu(\theta)$, in which values intermediate between the knots on the bidecadal atmospheric curve (Stuiver and Pearson 1993) or the marine curve (Stuiver and Braziunas 1993) are interpolated with a cubic spline function, and which might be smoothed (Törnqvist and Bierkens 1994). This results in calibrated histograms $H_z = f(z, \mu(\theta))$ and $H_y = f(y, \mu(\theta))$ (Fig. 1C).

To compare the shapes of H_z and H_y independent of the suite sizes m and n, the histograms are normalized so that the sum of the x_t is equal to unity. The actual test statistic, $v = f(H_{zt}, H_{yt})$, is then computed for each common value of t, $r = t_z \cap t_y$, yielding $v = (v, v_{r+1},\ldots, v_{r+p})$, where p is the number of elements in r (Fig. 1D).

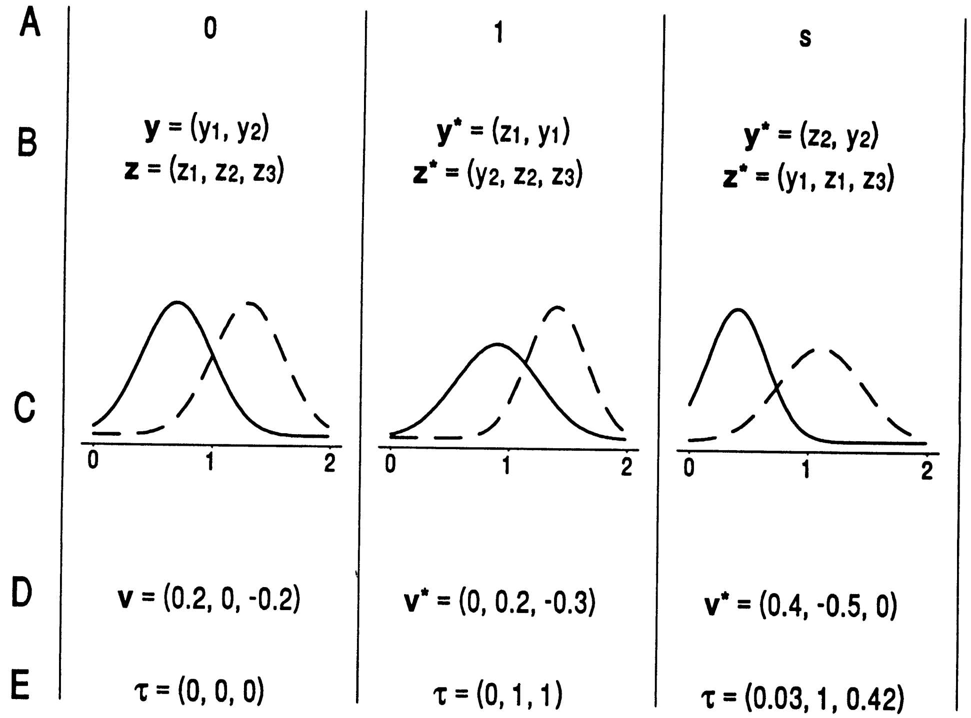

Fig. 1. Schematic illustration of the approximate randomization technique for comparing two ^{14}C histograms. The left panel shows the original data, the middle panel shows the first shuffle, and the right panel shows the last shuffle. A. Shuffle counter. B. Original (y, z) and simulated (y^*, z^*) ^{14}C date suites. The individual ^{14}C dates are shown in abbreviated form; $y_1 = y_1 \pm \sigma_1$, *etc.* C. Calibrated histograms; H_y = solid line, H_z = dashed line. The x-axis shows years 0, 1, and 2. D. The actual (v) and simulated (v^*) test statistics for years 0, 1, and 2. E. Estimated significance of the actual test statistic. The 0 values in the left panel indicate no estimate of significance. In the right panel the actual statistic at year 0 is significant at $\alpha < 0.05$.

The distribution of v at each r is approximated through repeated application of a procedure known as shuffling (Noreen 1989). Shuffling yields two simulated samples, $z^* = (z^*_1 \pm \sigma^*_1, z^*_2 \pm \sigma^*_2, \ldots, z^*_m \pm \sigma^*_m)$ and $y^* = (y^*_1 \pm \sigma^*_1, y^*_2 \pm \sigma^*_2,\ldots, y^*_n \pm \sigma^*_n)$, where z^* is constructed by randomly selecting m ^{14}C dates from $w = z \cup y$, and y^* is assigned the remaining n ^{14}C dates (Fig. 1B). Histograms, $H^*_z = f(z^*, \mu(\theta))$ and $H^*_y = f(y^*, \mu(\theta))$ (Fig. 1C), are made from the shuffled samples, and a simulated test statistic, $v^* = f(H^*_{zt}, H^*_{yt})$ is calculated for each $t \in r$, yielding $v^* = (v^*_r, v^*_{r+1},\ldots, v^*_{r+p})$ (Fig. 1D). This shuffling procedure is repeated s times yielding $V^* = (v^*_1, v^*_2,\ldots, v^*_s)$, where a sufficiently large value of s is chosen to ensure that the V^*_r approximate the permutation distributions of the test statistics.

The significance of the actual test statistic, τ, is assessed by calculating the proportion of $V^*_r \geq v_r$,

$$\sum_{k=1}^{s} \frac{j_k}{s} \tag{1}$$

https://doi.org/10.1017/S0033822200014934 Published online by Cambridge University Press

where

$$j_k = \begin{cases} 1 \; if \; v^*_{kr} \geq v_{kr} \\ 0 \; if \; v^*_{kr} < v_{kr} \; . \end{cases} \tag{2}$$

This yields $\tau = (\tau_r, \tau_{r+1}, \ldots, \tau_{r+p})$ (Fig. 1E). How well τ_r approximates the exact significance can be assessed with a confidence level, ϕ, the derivation and calculation of which is described by Noreen (1989).

AN EXAMPLE FROM HAWAIʻI

In the late 1980s, earth-moving during construction of a highway produced a 3-km-long, 100-m-wide stratigraphic "trench" through the inland portion of the traditional land division of Kaneʻohe on the island of Oʻahu (Williams 1992). Revealed in this trench at depths from 10–150 cm below surface were 107 traditional Hawaiian habitation and agricultural features that had been buried by sediments eroded from the steep mountain slopes above, a depositional event tied to the introduction of grazing herbivores in the late 18th century. A total of 75 ^{14}C dates was processed from 12 habitation features and from 27 agricultural features of three types. The 30 dates from habitation sites were collected from earth ovens, hearths or occupation layers within formal domestic features such as stone platforms, terraces or pavements. The 45 agricultural dates include: 14 dates from charcoal flecks picked out of pondfield soils; 13 dates from dispersed charcoal within non-irrigated garden features or colluvial sediments; and 18 dates from isolated earth ovens interpreted by the excavator as having been used by cultivators as they worked their fields.

One goal of Williams' analysis was to chart the history of traditional Hawaiian land use in inland Kaneʻohe leading up to the establishment of the Contact-era settlement pattern, in which permanent habitations were dispersed among agricultural fields. The theoretical importance of permanent inland settlement was pointed out by Hommon (1976, 1986), who recognized it as a precondition for the development of the traditional Hawaiian land and social division called the *ahupuaʻa*. The *ahupuaʻa* community differed from analogous groups elsewhere in Polynesia at the time of European contact, and is assumed to have developed its unusual form in Hawaiʻi. This involved: 1) the development of ephemeral kindreds out of the corporate descent groups characteristic of Polynesia; 2) establishment of a marked social and economic self-sufficiency; and 3) a rupture of the traditional kinship relationship between chiefs and commoners. Hommon theorized that the physical manifestation of these interrelated processes was the establishment of social networks along a primary coastal/inland axis from a previous pattern of integrated coastal communities. Thus, the development of several important features of Contact-era Hawaiian society might be traced by the progress of inland settlement.

Although the ^{14}C dates collected by Williams do not constitute a formal random sample of traditional Hawaiian habitation and agricultural features in the inland portion of Kaneʻohe, it is difficult to identify any systematic bias in their collection. The path of the highway was determined by modern political and engineering considerations very far removed from the topographic features that shaped the traditional Hawaiian landscape of the region. The archaeologist attempted to date a large number of each type of feature, and was not biased in the choice of samples by considerations of their likely antiquity. Thus, with respect to the research questions concerning the nature and timing of inland settlement in Kaneʻohe, the data are likely to be unbiased.

Visual comparison of the shapes of the ^{14}C histograms for habitation and agricultural dates (Fig. 2) shows that they are different over most of the x axis. The habitation histogram shows an early rise *ca.* AD 800 caused by a single ^{14}C date that is >400 ^{14}C yr older than the next oldest date. This outlier might be due to the inbuilt age of the sample. Very little data exist on the maximum ages of forest species in Hawai'i, and old wood is a potential problem for all archaeological samples that are not selected from species known to be short-lived, from twigs or small sticks, or from the outer rings of trees (McFadgen 1982). At *ca.* AD 1200, the histogram begins its rise to a peak in the early 15th century, after which a very modest first-order increase occurs. The agriculture ^{14}C-date suite also contains outliers, but these are too old to show on the histogram (Fig. 2). The agriculture histogram begins to rise *ca.* AD 1000, peaks sharply at *ca.* AD 1400, declines rapidly for a century, then more slowly over the next 300 yr. Williams' analysis of these histograms focused on the indications that burning associated with agriculture began earlier than the evidence for habitation, and that the two histograms showed opposite trends in the late prehistoric period, with agriculture declining while habitation increased. He concluded that increased use of upland Kane'ohe for agriculture and cooking at isolated earth ovens began *ca.* AD 1000 and was followed 150–200 yr later by the establishment of habitation sites (Williams 1992). He also concluded that the evidence for the relative decline of agricultural and isolated cooking activities after *ca.* AD 1450 (the agriculture histogram actually reaches its maximum value in AD 1418) was due to a decline in biomass after the virgin forest was cleared, and to the abandonment of garden ovens after habitations were established nearby (Williams 1992); there is no evidence that the practice of agriculture declined.

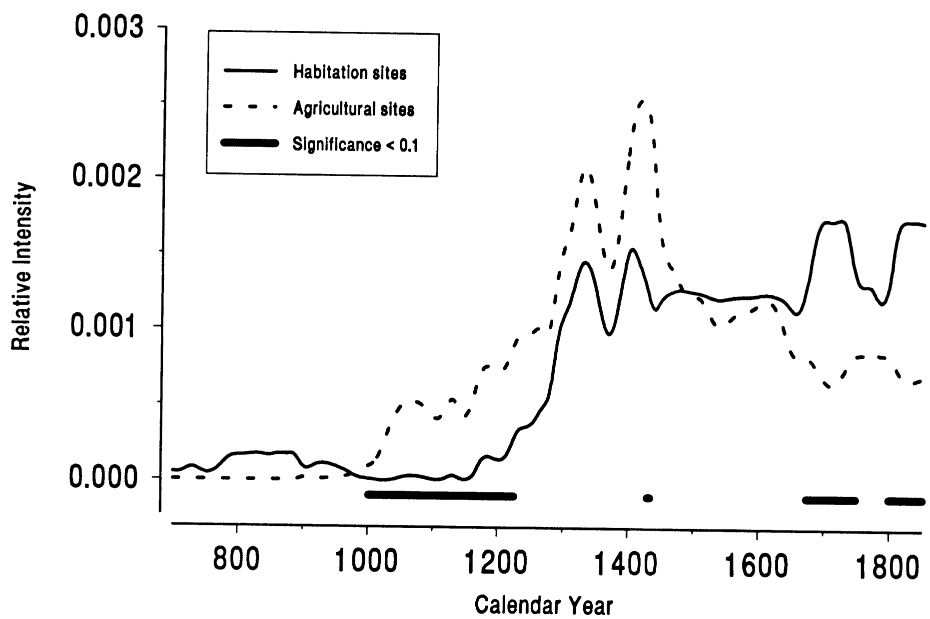

Fig. 2. ^{14}C histograms for habitation and agricultural activities from inland Kane'ohe ahupua'a, normalized so that the area under each histogram is unity and calibrated using Stuiver and Pearson (1993). Periods over which the curves are significantly different (0.1, $\phi > 0.95$) are marked by the heavy horizontal line.

Comparing the Shapes of ^{14}C Histograms for Habitation and Agriculture

The goal of the present analysis is to determine whether the indications of differences upon which Williams based his conclusions are, in fact, significant, or whether they might be due to chance factors associated with sample size. The first step is to choose an appropriate test statistic, one that cap-

https://doi.org/10.1017/S0033822200014934 Published online by Cambridge University Press

tures the differences that Williams believed to be important. For the first interval (1000, 1200), the choice of a test statistic is fairly straightforward. The relative intensity of the agriculture histogram is greater than the habitation histogram over the entire interval and the pertinent question would seem to be whether the differences between the histograms are great enough to be significant. Thus, an appropriate test statistic is simply the difference between the intensities, H_z–H_y.

An appropriate test statistic for the second interval (1418, 1850) is more difficult to choose because the salient characteristic singled out by Williams is the first-order decline of the agriculture histogram relative to the habitation histogram. One possible test statistic is the difference in the annual rate of change, but the histograms show relatively large second-order fluctuations, and these would likely mask the first-order trend. A better choice of test statistic stems from the observation that for the relative decline to be significant, the histograms must show a significant difference either in the early portion of the interval, when the relative intensity of the agriculture histogram is greater than the habitation histogram, and/or in the late portion of the interval, when the relative intensity of the habitation histogram is greater. If there are no significant differences in the interval, then there is no basis for the claim that the intensity of the agriculture histogram declined relative to the habitation histogram. Thus, the difference between the intensities, H_z–H_y, can be used as a test statistic in this situation as well.

The agriculture and habitation histograms were compared following the procedure outlined above, with the number of shuffles, $s = 999$. The four intervals that are significantly different at the 0.1 level, with $\phi > 0.95$, are marked by the bold horizontal line segments in Figure 2. They are (1001, 1225) and (1430, 1434), where the agriculture intensity is greater than the habitation intensity, and (1674, 1748) and (1799, 1850), where the habitation intensity is greater.

These significant intervals can be interpreted to support Williams' conclusions. The inference that habitation lagged agricultural use of the region by 150–200 yr is supported by the first interval of significant difference, which shows that the relatively early rise of the agriculture histogram is not due to sampling error, but instead is a pattern in the archaeological record. Williams assumed that this pattern was due to changes over time in the use of the region, but it is worthwhile to point out that it could also be due to the effects of dating "old wood." If there were some systematic difference in the woods used for burning in the fields and at home, then it is possible that activities occurring at the same time could produce histograms similar to those in Figure 2. This might happen, for instance, if firewood was selected by size, such that the largest logs with the oldest heartwood were left in the fields where they were burned, whereas the smaller branches, more easily transported, were taken home for firewood. This is a potential problem that can be resolved in the future through more careful characterization and choice of dating materials.

The inference that agricultural and isolated cooking activities declined relative to habitation activities after AD 1450 appears to be supported by the last three intervals of significant difference. The intensity of the agriculture histogram is significantly greater than that of the habitation histogram in the brief interval (1430, 1434). The decline of the agriculture histogram brings its relative intensity below that of the habitation histogram at 1484. This difference becomes significant over the interval (1674, 1748), when the relative intensity of the habitation histogram is significantly greater than the agriculture histogram for the first time. This is followed by a 51-yr period during which there is no significant difference between the histograms. The histograms are again significantly different over the last 41 yr (1809, 1850). In general, these three significant intervals indicate that agricultural clearing and the use of garden ovens did decline relative to the use of habitations, but each of the intervals is relatively short, with none of them >75 yr. This is a cause for concern because it is well known that the intensity of ^{14}C histograms can fluctuate due to wiggles in the calibration curve (Stolk *et al.* 1994; Törnqvist and Bierkens 1994). Do the intervals produced by the approximate randomization technique reflect wiggles in the calibration curve?

HOW SMOOTH SHOULD CURVES BE FOR COMPARING ^{14}C HISTOGRAMS?

A practical test of this question was devised using the smoothed calibration curves developed by Törnqvist and Bierkens (1994). The example data were calibrated with five different smoothed curves, including 10-, 20-, 40-, 80-, and 160-yr smooths, and then compared using the approximate randomization technique, as above. In Figure 3, I compare the significant intervals yielded by these tests with the four significant intervals yielded by the original test. The first and second intervals of significant difference (1001, 1225) and (1430, 1434) change very little with different calibration curves. The beginning of the first interval varies by only 6 yr and the end by only 31 yr. The second interval grows slightly with the smoother curves, reaching a maximum of 15 yr, and moves up to 30 yr later. These differences are so small that they are very unlikely to have an effect on interpretation.

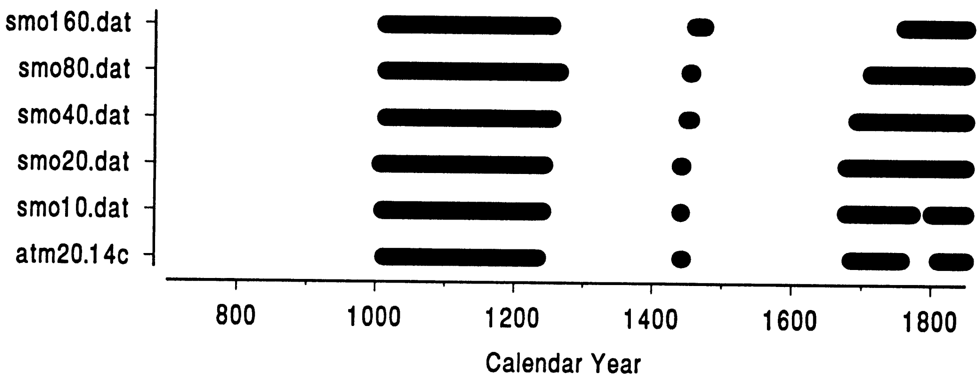

Fig. 3. Comparison of significantly different (0.1, $\phi > 0.95$) intervals of the histograms using calibration curves with different degrees of smoothing. The curve atm20.14c is an unsmoothed curve (Stuiver and Pearson 1993). The other curves (smo10.dat, *etc.*) are distributed with the CALHIS program (Stolk *et al.* 1994).

The other intervals show more pronounced differences. The 51-yr gap between the 3rd and 4th significant intervals shrinks to 25 yr in the histograms produced with the 10-yr smoothed curve, then disappears altogether, yielding a single interval that varies from (1677, 1850) in the histograms produced with the 20-yr smoothed curve to (1760, 1850) in the histograms produced with the 160-yr smoothed curve. The relatively large differences in the beginning of the interval with the very smooth curves appear to be due to boundary effects associated with the end of the calibration curve, and should not pose a problem with suites of older dates. Although the differences in these intervals are relatively large, they are unlikely to have an effect on interpretation. Even the histograms produced with the 80- and 160-yr smoothed curves, which differ most markedly from the histograms produced with the unsmoothed curve, show the same basic pattern of significant intervals, and the greatest differences are due to boundary effects, rather than to the effects of the smoothing *per se*.

The choice of a calibration curve appears to have fairly small effects on the comparison of ^{14}C histograms. Intervals of significance or insignificance shorter than *ca.* 50 yr might be due to wiggles in the calibration curve. The use of calibration curves with a 20-yr smooth or greater can remove these shorter intervals. Clearly, intervals this short should not be overinterpreted. Oversmoothed curves can introduce boundary effects in histograms of relatively late date suites, but should not pose a problem for the comparison of older date suites.

Determination of intervals of significant difference between ^{14}C histograms can also lead to additional avenues of interpretation. Figure 4 shows a finished plot of the example data produced using a 40-yr smoothed calibration curve, which yields three intervals of significant difference. Several interpretations of this figure are possible. One cultural interpretation might posit three periods in the colonization of inland Kane'ohe *ahupua'a* defined by the two long intervals of significant difference and the interval of insignificant difference between them. The first period, from the beginning of the 11th century through the early 13th century, appears to reflect the clearing of native forest for agriculture with little or no associated habitation. Presumably, the farmers lived at the coast and walked to work in their fields, where they cooked occasional meals in isolated earth ovens. This was followed in the mid-13th through mid-17th centuries by a period during which habitations were established among gardens whose boundaries slowly became fixed and in which the clearing of new lands eventually declined. The Contact-era settlement pattern, and by inference the self-sufficient *ahupua'a* community, was in place by the late 17th century, after a 600-yr period of development.

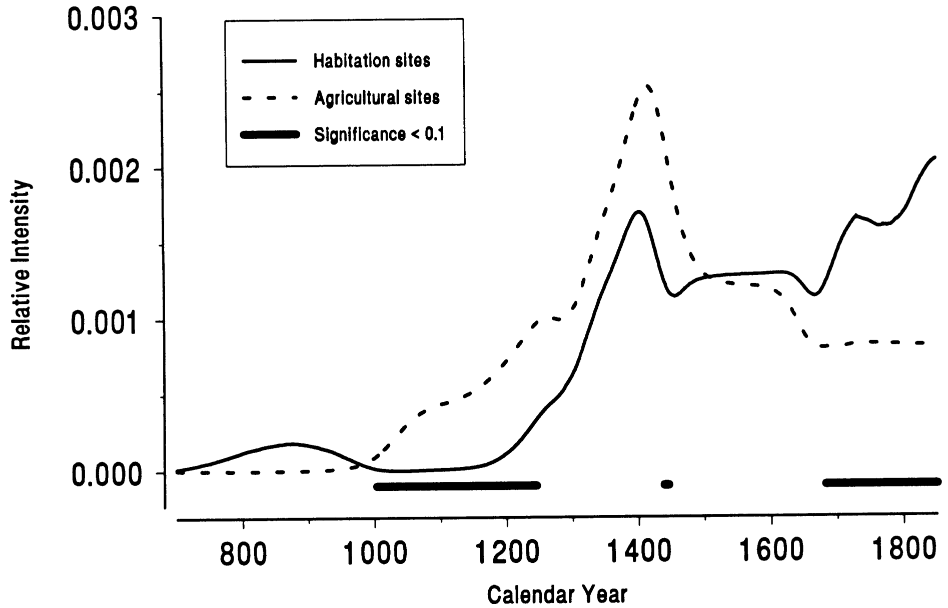

Fig. 4. Comparison of ^{14}C histograms for habitation and agricultural activities from inland Kane'ohe ahupua'a. The histograms were constructed following the procedure outlined by Törnqvist and Bierkens (1994), using the 40-yr smoothed calibration curve. Significant differences (0.1, ϕ >0.95) are shown by the heavy horizontal line.

Conclusion

Approximate randomization techniques appear to be a useful tool capable of moving a comparative analysis of ^{14}C histograms from the level of indication to determination. Determining intervals of significant difference with an appropriate test statistic provides an objective foundation for comparing ^{14}C histograms and guards against overinterpretation of histograms constructed from small date suites. Use of a calibration curve with a moderate amount of smoothing reduces the occurrence of short intervals of significance and insignificance, which should be interpreted with caution in any case.

Acknowledgments

R. J. Hommon, B. F. Leach, D. H. Thomas and S. S. Williams graciously offered useful comments on earlier drafts. Errors remain the author's.

References

Allen, J. 1992 Farming in Hawai'i from colonisation to contact: Radiocarbon chronology and implications for cultural change. *New Zealand Journal of Archaeology* 14: 45–66.

Dye, T. 1994 Population trends in Hawaii before 1778. *The Hawaiian Journal of History* 28: 1–20.

Dye, T. and Komori, E. 1992 A pre-censal population history of Hawai'i. *New Zealand Journal of Archaeology* 14: 113–128.

Hommon, R. J. (ms.) 1976 The Formation of Primitive States in Pre-Contact Hawaii. Ph.D. dissertation, The University of Arizona: 356 p.

____1986 Social evolution in ancient Hawai'i. *In* Kirch, P. V., ed., *Island Societies: Archaeological Approaches to Evolution and Transformation*. Cambridge, Cambridge University Press: 55–68.

McFadgen, B. G. 1982 Dating New Zealand archaeology by radiocarbon. *New Zealand Journal of Science* 25: 379–392.

McFadgen, B. G., Knox, F. B. and Cole, T. R. L. 1994 Radiocarbon calibration curve variations and their implications for the interpretation of New Zealand prehistory. *Radiocarbon* 36(2): 221–236.

Mosteller, F. and Tukey, J. W. 1977 *Data Analysis and Regression: A Second Course in Statistics*. Behavioral Science: Quantitative Methods. Reading, Massachusetts, Addison-Wesley: 588 p.

Noreen, E. W. 1989 *Computer Intensive Methods for Testing Hypotheses: An Introduction*. New York, John Wiley and Sons: 229 p.

Rick, J. W. 1987 Dates as data: An examination of the Peruvian preceramic radiocarbon record. *American Antiquity* 52(1): 55–73.

Stolk, A., Törnqvist, T. E., Hekhuis, K. P. V., Berendsen, H. J. A. and van der Plicht, J. 1994 Calibration of ^{14}C histograms: A comparison of methods. *Radiocarbon* 36(1): 1–10.

Streck, C. F., Jr. 1992 Prehistoric settlement in the upland portions of the island of Hawai'i. *New Zealand Journal of Archaeology* 14: 99–111.

Stuiver, M. and Braziunas, T. F. 1993 Modeling atmospheric ^{14}C influences and ^{14}C ages of marine samples to 10,000 BC. *In* Stuiver, M., Long A. and Kra, R. S., eds., Calibration 1993. *Radiocarbon* 35(1): 137–189.

Stuiver, M. and Pearson, G. W. 1993 High-precision calibration of the radiocarbon time scale, AD 1950–500 BC and 2500–6000 BC. *In* Stuiver, M., Long A. and Kra, R. S., eds., Calibration 1993. *Radiocarbon* 35(1): 1–23.

Törnqvist, T. E. and Bierkens, M. F. P. 1994 How smooth should curves be for calibrating radiocarbon ages? *Radiocarbon* 36(1): 11–26.

Williams, S. S. 1992 Early inland settlement expansion and the effect of geomorphological change on the archaeological record in Kane'ohe, O'ahu. *New Zealand Journal of Archaeology* 14: 67–78.