A comment on Steele's (2010) "radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities"

Briggs Buchanan [a,b,1], Marcus Hamilton [c,d,e], Kevan Edinborough [f], Michael J. O'Brien [b], Mark Collard [a,b,*,1]

[a] Human Evolutionary Studies Programme and Department of Archaeology, Simon Fraser University, Canada
[b] Department of Anthropology, University of Missouri, USA
[c] Department of Anthropology, University of New Mexico, USA
[d] Department of Biology, University of New Mexico, USA
[e] Santa Fe Institute, USA
[f] Institute of Archaeology, University College London, UK

ARTICLE INFO

ABSTRACT

We show that Steele's (2010) criticisms of Hamilton and Buchanan (2007) and Buchanan et al. (2008) do not hold water and demonstrate that his re-analyses of Hamilton and Buchanan's (2007) and Buchanan et al'.s (2008) datasets are flawed. In the process, we highlight some important issues for researchers interested in using radiocarbon dates to reconstruct population movements and demography. Most notably, we explain why OLS regression is preferable to RMA regression when estimating diffusion velocity, and demonstrate that the summed probability distributions yielded by CalPal are more reliable as guides to past demographic change than those produced by Calib and OxCal.

© 2011 Published by Elsevier Ltd.

1. Introduction

In a recent paper in this journal, Steele (2010) criticized Hamilton and Buchanan's (2007) "Spatial gradients in Clovis-age radiocarbon dates across North America suggest rapid colonization from the north" and Buchanan et al.'s (2008) "Paleoindian demography and the extraterrestrial impact hypothesis." He also reported re-analyses of their data that yielded results that are markedly different from the ones they obtained. Here, we show that Steele's criticisms of Hamilton and Buchanan (2007) and Buchanan et al. (2008) are without foundation and that the results of his re-analyses are inaccurate. While our primary goal is to set the record straight regarding Hamilton and Buchanan (2007) and Buchanan et al. (2008), a number of the points we make and the analyses we present have implications for the use of radiocarbon dates to reconstruct population movements and demography in archaeology more generally.

2. Steele (2010) on Hamilton and Buchanan (2007)

2.1. Hamilton and Buchanan's aims, methods, and findings

Hamilton and Buchanan carried out two analyses. First, they used calibrated radiocarbon dates from Clovis-age sites to test six competing models for the spread of Early Paleoindian populations across North America. Four of the models posited external origins and two postulated internal, pre-Clovis origins. Hamilton and Buchanan began by identifying origin points that are consistent with the models. They then used great-circle arcs to measure the distance of each site from the point of origin for each model. Next, they binned the calibrated dates using concentric bins of a constant width. The earliest dates in the bins were then correlated with the distance to the point of origin for each model. In their second analysis, Hamilton and Buchanan estimated the velocity of the Early Paleoindian expansion. To do so, they used a method previously employed by Fort and colleagues (e.g. Fort and Mendez, 2002; Fort et al., 2004a,b; Pinhasi et al., 2005). This method estimates the velocity of an expanding population as the inverse slope of the ordinary least squares (OLS) regression of calibrated dates by distance from origin. Because of the small sample sizes involved,

Hamilton and Buchanan also used a resampling method to estimate the slope.

Hamilton and Buchanan found that the ice-free corridor model had the highest correlation coefficient of the six models ($r = -0.73$). Using standard OLS regression, they estimated the velocity of the wavefront to be 7.56 km per year. The resampling method yielded a lower estimate of 5.13 km per year. Hamilton and Buchanan compared these estimates to velocities that have been estimated for other prehistoric population expansions into unoccupied landmasses and to a velocity generated from hunter-gatherer ethnographic data. They found the estimates for the Early Paleoindians to be relatively fast. The high velocity of the Early Paleoindian diffusion, they suggested, can be accounted for by a combination of demographic processes, habitat preferences, and mobility biases.

2.2. Steele's criticisms and reanalysis

Steele criticized Hamilton and Buchanan's study on three counts. First, he claimed that they over-extended Clovis by including dates from sites that have not produced diagnostic Clovis artifacts. Second, he argued that their post-calibration midpoint values for the dates from three sites—Debert, Hedden, and Vail—are wrong. Third, he asserted that they should have employed reduced major axis (RMA) regression rather than OLS regression. Steele's rationale was that RMA takes into account error in both the independent and dependent variables whereas OLS allows for error only in the dependent variable, and there is reason to think that error exists in both variables in the type of analysis carried out by Hamilton and Buchanan.

Steele carried out three sets of analyses in an attempt to demonstrate that Hamilton and Buchanan's findings are dependent on what he considered to be their insufficiently rigorous approach to selecting radiocarbon dates. In the first, he estimated the slope and velocity of the Clovis diffusion using Hamilton and Buchanan's complete dataset but with the older calibrated midpoints for the dates from Debert, Hedden, and Vail. He used the Markov chain Monte Carlo routine in OxCal (Bronk Ramsey, 2009) to randomly draw a calendar date from each calibrated date range and subjected the resulting set of randomly selected dates to RMA with the point of origin of the diffusion set at Edmonton. This analysis was repeated 999 times, and then the mean slope and median velocity was calculated. Next, Steele removed the dates from Debert, Hedden, and Vail and re-ran the analyses. Subsequently, he repeated the analyses after removing three more dates (from Big Eddy, Casper, and Hiscock) that he argued are questionable because they were not "approved" by Waters and Stafford (2007) in their review of Clovis radiocarbon dates.

All the correlations Steele obtained in the first set of analyses were significant, and the mean velocity of the diffusion was 5.71 km per year. In the second set of analyses, 80% of the correlation coefficients were significant, and the mean velocity of the diffusion wave was 9.3 km per year. In the third set of analyses, 52% of the correlation coefficients were significant, and the mean velocity of the diffusion wave was 11 km per year. Steele concluded from the progressive decline in the number of significant correlations from the first set of analyses to the third that Hamilton and Buchanan's findings were dependent on the inclusion of problematic dates.

Needless to say, if Steele's criticisms of Hamilton and Buchanan's study were correct and his re-analyses of their datasets reliable, there would be reason to be skeptical about Hamilton and Buchanan's findings. Such is not the case, however.

2.3. Problems with Steele's criticisms and reanalysis

The claim that Hamilton and Buchanan's study is problematic because they included dates from sites that have not produced diagnostic Clovis artifacts is based on a misunderstanding of the goal of the study. As explained above, Hamilton and Buchanan's objective was to identify the best-fit gradient for the earliest dated occupations across North America, not the best-fit gradient for only Clovis occupations. Thus, including dates for Early Paleoindian sites that have not produced diagnostic Clovis artifacts was a perfectly valid course of action for Hamilton and Buchanan to have followed.

Steele's second criticism—that Hamilton and Buchanan used the wrong calibrated dates for Debert, Hedden, and Vail—is also problematic. Twenty of the uncalibrated dates in Hamilton and Buchanan's sample return single calibrated ranges in the calibration program they used. However, the uncalibrated dates from Debert, Hedden, and Vail intercept the calibration curve in two places and therefore yield two calibrated ranges. In each case Hamilton and Buchanan used the midpoint from the younger range. Hence, their midpoint values for Debert, Hedden, and Vail are different from the ones Steele employed, but they are not wrong. Significantly, Hamilton and Buchanan's conclusions would not have been different if they *had* used the older post-calibration midpoint values for Debert, Hedden, and Vail. Using the older age estimates for these sites produces a higher correlation coefficient for the ice-free corridor model ($r = -0.75$; Table 1) and a faster estimated velocity for the diffusion (11.9 km per year). Hence, even if Hamilton and Buchanan had used the midpoints that Steele prefers, they would still have concluded that the ice-free corridor model is the best-fit model and that the Clovis-age diffusion was comparatively fast.

Steele's claim that Hamilton and Buchanan should have employed RMA regression rather than OLS regression is flawed on two counts. First, by definition, OLS recovers a linear functional relationship between x and y variables of the form $y = y_0 + \beta x + \epsilon$, where ϵ is an error term. With OLS the relationship between variables is asymmetric and the direction of causality is clear. RMA not only changes the assumption of how errors are structured in the data, but by doing so changes the underlying meaning of the regression model. In RMA, by dividing the error term ϵ between the axes, the relationship between the variables effectively becomes $y + \epsilon/2 = y_0 + \beta x + \epsilon/2$, which has a different meaning to the OLS equation (Smith, 2009). The RMA model is symmetrical, and therefore there is no clear direction of causality. As such, contrary to what Steele contends, RMA regression is actually *less* appropriate than OLS regression for analyses of the type carried out by Hamilton and Buchanan.

Second, reanalysis of Hamilton and Buchanan's data using RMA yields results that are consistent with the results they reported. The velocity of the diffusion wave is lower when the RMA method is used (5.3 km). But the estimate falls within the confidence limits reported by Hamilton and Buchanan and is still faster than the comparative velocities discussed by Hamilton and Buchanan.

Table 1

Results from regression analyses of the earliest Clovis-age occupations in bins correlated with the distance to the point of origin for six models of the colonization of North America. Original correlation coefficients (r) and p-values from Hamilton and Buchanan (2007). New r- and p-values obtained using the older midpoints for Debert, Hedden, and Vail.

Model	Original r	Original p	New r	New p
North	−0.73	<0.01	−0.75	0.004[a]
South	−0.55	0.22	−0.53	0.22
East	0.47	0.97	0.47	0.97
West	−0.68	0.05	−0.61	0.04
Meadowcroft	0.44	0.96	0.44	0.96
Cactus hill	0.39	0.93	0.39	0.93

[a] significant after Bonferroni correction ($p = 0.05/6 = 0.00833$).

Hence, even if Hamilton and Buchanan had used RMA regression, their conclusions would have been identical.

Needless to say, Steele's use of RMA in and of itself casts doubt on the results of his analysis of Hamilton and Buchanan's data. However, the problems with the analysis do not stop there. There are at least three more.

One is that Steele's reanalysis does not replicate the modeling technique used by Hamilton and Buchanan. Steele fitted a regression line through the entire dataset, not through the earliest occupations. This procedure addresses the question of whether there is a spatiotemporal gradient in the average occupation times for Clovis-age sites across North America rather than the question that Hamilton and Buchanan addressed, which is whether there is a gradient in the *earliest* occupations across North America. To address the latter question, it is necessary to bin the dates and analyze only the earliest date in each bin. Thus, Steele's results are not comparable to Hamilton and Buchanan's.

Another problem is that Steele evaluated only a single model. As we explained earlier, Hamilton and Buchanan compared six competing models for the origin of the Clovis-age diffusion. Steele ignored this aspect of Hamilton and Buchanan's study and focused exclusively on the ice-free corridor model. Thus, Steele's results are, once again, not comparable to those reported by Hamilton and Buchanan.

The last problem we will discuss is even more profound than Steele's use of an inappropriate regression technique and his failure to carry out the analysis in such a way that its results can be compared with those obtained by Hamilton and Buchanan. As we explained earlier, Steele's analysis was designed to demonstrate that Hamilton and Buchanan's findings are dependent on what Steele considered to be their insufficiently rigorous approach to selecting radiocarbon dates. To accomplish this, Steele first analyzed Hamilton and Buchanan's dataset using the older midpoints for the dates from Debert, Hedden, and Vail. He then removed those dates and re-ran the analysis. Subsequently, he repeated the analyses after removing three more dates that he considered to be questionable because they were not "approved" by Waters and Stafford (2007). The problem with this approach is that it assumes that the quality of the dataset increases at each step. This is not the case, however. To reiterate, there is nothing wrong with the midpoint values for Debert, Hedden, and Vail employed by Hamilton and Buchanan. They are different from the ones used by Steele, but they are not invalid. The same holds for the dates from Big Eddy, Casper, and Hiscock. As Hamilton and Buchanan explained, Waters and Stafford (2007) excluded these dates from their list of reliable Clovis-age dates without good reason. Thus, contrary to what Steele assumed, removing the dates from the six sites did not produce a more robust dataset. Rather, it produced only a *smaller* dataset. Accordingly, Steele's analysis of Hamilton and Buchanan's dataset does not call into question the reliability of their results. All it shows is that Hamilton and Buchanan's results would have been more ambiguous if they had used fewer dates—a finding that is both trivial and irrelevant.

In sum, Steele's criticisms of Hamilton and Buchanan's study do not withstand scrutiny, and his reanalysis of Hamilton and Buchanan's data did not do what he intended. Contrary to what Steele suggests, therefore, his study does not cast doubt on Hamilton and Buchanan's findings.

3. Steele (2010) on Buchanan et al. (2008)

3.1. Buchanan et al.'s aims, methods, and findings

The study reported by Buchanan et al. was designed to test the main archaeological predictions of Firestone et al.'s (2007)

extraterrestrial (ET) impact hypothesis. Firestone et al. (2007) argued that one or more large extraterrestrial objects impacted or exploded over northern North America 12,900 ± 100 calendar years BP (CalBP). This impact, they suggested, was accompanied by a high-temperature shock wave, changes in pressure that would have resulted in hurricane-force winds, and extensive groundcover burning. Together, these triggered the Younger Dryas cooling event and caused a continent-wide environmental collapse. The latter, in turn, resulted in the extinction of the North American megafauna as well as a population bottleneck and major cultural changes among the Paleoindians.

Buchanan et al. carried out two analyses of a dataset of 1509 radiocarbon dates from North American archaeological sites to test the ET impact hypothesis. The radiocarbon dates spanned 13,000–8000 ^{14}C BP and were obtained from Hamilton and Buchanan (2007), Waters and Stafford (2007), and the Canadian Archaeological Radiocarbon Database (CARD) (Morlan, 2005). Prior to carrying out the analyses, Buchanan et al. removed 203 dates identified as "anomalous" in CARD and then used Calib 5.1 (Stuiver et al., 2005) to pool dates derived from the same occupation. Occupations were defined on the basis of stratigraphic and cultural information given in CARD. In the first analysis, they used the dates to estimate demographic change across the proposed ET impact. They reasoned that if Paleoindians experienced a population bottleneck as a result of an ET impact, then the summed probability distribution (SPD) of the dates should show a major trough at 12,900 ± 100 CalBP. In the second analysis, Buchanan et al. used χ^2 tests to compare the spatial distribution of calibrated dates in the 300 years prior to the ET impact with the spatial distribution of calibrated dates at the time of, and shortly after, the impact. They reasoned that the effects of the impact should have been more pronounced in the northern part of the continent, closer to the proposed zone of impact, than in the southern part, and therefore the southern population would have been less affected than the northern population.

Buchanan et al.'s first analysis did not support the ET impact hypothesis. The SPD exhibited a number of troughs, including one that began at 12,800 calBP, which is within the error range of the date for the impact used by Firestone et al. (2007). However, the trough at 12,800 calBP was not only short but also relatively minor in scale. It lasted only 100 years and was no more pronounced than some of the other troughs in the SPD. As such, Buchanan et al. argued that the trough at 12,800 CalBP was not consistent with a population bottleneck. The spatial analysis of the dates also did not support the impact hypothesis. The χ^2 test revealed no statistical difference in the counts of radiocarbon dates in the six blocks of latitude and longitude between the first and second periods ($\chi^2 = 8.13$, $P = 0.15$). Similarly, no statistical difference was found in the counts of radiocarbon dates in the six blocks of latitude and longitude between the second and third periods ($\chi^2 = 3.83$, $P = 0.57$). Redistributing the blocks using different longitudinal boundaries did not alter the results of the χ^2 test. Buchanan et al. concluded that the results of the two analyses supported neither Firestone et al.'s (2007) original suggestion that the Paleoindians experienced a population bottleneck as a result of an ET impact at 12,900 ± 100 calBP nor a weaker hypothesis in which Paleoindian populations simply migrated south after the proposed impact.

3.2. Steele's criticisms and reanalysis

As with his critique of Hamilton and Buchanan's study, Steele began by criticizing Buchanan et al.'s data. This time his argument was that Buchanan et al. incorrectly pooled multiple dates from discrete cultural and/or stratigraphic layers. To support this claim, Steele showed that several of the groups of dates that Buchanan
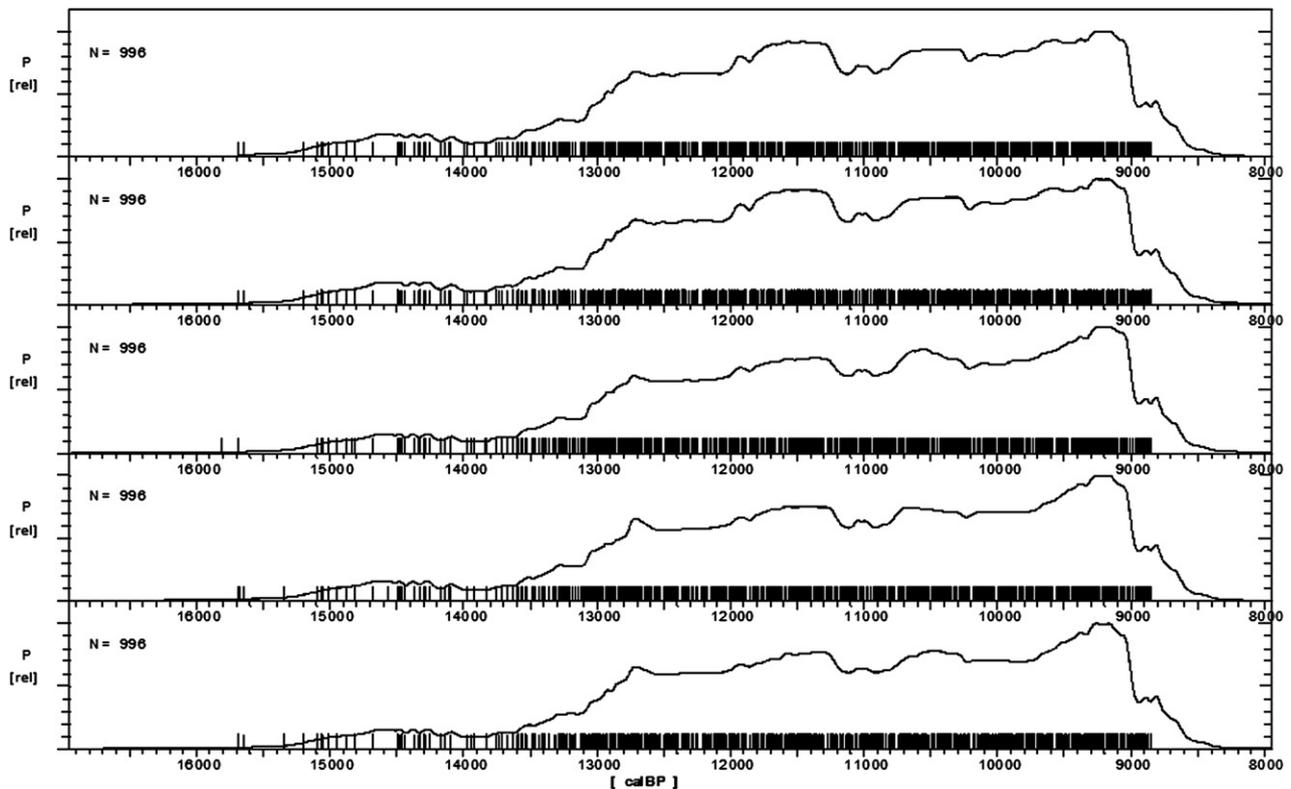
Fig. 1. Summed probability distributions derived from five random samples of 996 radiocarbon dates from Buchanan et al.'s (2008) original dataset.

et al. pooled are statistically different according to Calib's χ^2 test pooling procedure.

Subsequently, Steele criticized the program Buchanan et al. used to generate SPDs, CalPal (Weninger et al., 2007). Steele used CalPal, together with Calib and OxCal, to produce SPDs from Buchanan et al.'s dataset. He found that the SPD yielded by CalPal differed from those returned by Calib and OxCal. The SPDs created with Calib and OxCal exhibit large peaks at approximately 12,900, 11,200, 10,200 and 9500 calBP; the SPD created with CalPal does not. Steele argued that this difference is the result of CalPal employing an "idiosyncratic smoothing algorithm" (p. 2023).

Steele then carried out three analyses of Buchanan et al.'s dataset. He did not explicitly state the goal of the first analysis, but we can infer that it was designed to re-assess Firestone et al.'s (2007) claim that there was an impact-induced population bottleneck at 12,900 ± 100 calBP. To accomplish this, Steele used a bootstrapping-based method to create an SPD. The method repeatedly draws a single calendar-year age from the calibrated age distribution associated with each radiocarbon determination or pooled mean. After sampling each distribution 1000 times, the frequency distribution of events occurring in a series of binned time intervals is plotted. As in the approach employed by Buchanan et al., the major peaks and troughs of the resulting SPD are interpreted as reflecting peaks and troughs of population size.

Steele's second analysis was designed to assess the impact of the onset of the Younger Dryas on demography. In this analysis, Steele calibrated Buchanan et al.'s dates and then assigned them to four 100-year bins spanning 12,900–12,500 calBP. To be assigned to a given bin, a date had to have at least a 10% probability of falling in the relevant 100-year time period. Subsequently, he calculated the number of dates that fell in each of the four bins and compared the totals.

The goal of Steele's third analysis was to "gain an insight into the scale of the problem of 'chronometric hygiene' involved when

compiling and using datasets of this kind to address questions about prehistoric climatic events at an adequate temporal resolution" (p. 2027). To accomplish this, Steele identified the "top five" dates from four time periods: 12,900–12,800 calBP, 12,800–12,700 calBP, 12,700–12,600 calBP, and 12,600–12,500 calBP. The top five dates were defined as those that overlap the most with the time period after calibration. Steele then assessed the validity of the top five dates in each time period.

The results of Steele's first analysis of Buchanan and colleagues' dataset differ markedly from those they obtained. Whereas Buchanan et al. found no evidence for a population bottleneck within the time range proposed by Firestone et al. (2007), Steele identified a trough in his SPD at 12,800 calBP, which he interpreted as evidence of a major population decline.

In the section of the paper that outlines the results of his second analysis of Buchanan et al.'s dataset, Steele compared the 12,600–12,500 calBP time period with the preceding four 100-year time periods. He pointed out that the number of dates in the former is low compared to the latter but suggested that this may be an artifact of the Younger Dryas ^{14}C plateau. The implication of this seems to be that Steele regards the results of his second analysis of Buchanan et al.'s dataset as ambiguous.

Of the 20 dates examined in the third analysis, Steele argued that five had been inappropriately pooled and identified one as residual. He claimed that another three dates had been rendered obsolete by more recent determinations. Steele did not spell out the implications of these findings for Buchanan et al.'s conclusions, but in the discussion section he made it clear that he believed that the analysis undermined Buchanan et al.'s conclusions.

On the face of it, Steele's critique of Buchanan et al.'s study seems even more damning than his critique of Hamilton and Buchanan's study. Once again, however, this is not the case.

3.3. Problems with Steele's criticisms and reanalysis

Steele's assertion that Buchanan et al. incorrectly pooled multiple dates from single occupations is inaccurate. They did not use a statistical test to pool dates because it would have required them to subjectively decide which date to keep in cases where two or more dates from a single occupation are significantly different. Instead, they followed the more conservative course of action of pooling *all* dates from a single occupation. Thus, Steele's "incorrectly pooled dates" are more apparent than real.

We reanalyzed Buchanan et al.'s dataset to ensure that their approach to pooling dates did not bias their results. We removed the 203 dates that are identified as "anomalous" in CARD and then randomly sampled 996 unpooled dates from the resulting set of 1306 unpooled dates (996 is the maximum number of dates that can be analyzed simultaneously in CalPal). Next, we calibrated the 996 randomly sampled dates and generated an SPD from them. Thereafter, we repeated the procedure four times. As can be seen in Fig. 1, the five SPDs are nearly identical to the SPD obtained by Buchanan and colleagues. Critically for present purposes, they also show no evidence of a population bottleneck at 12,900 ± 100 calBP. This strongly suggests that, contrary to Steele's assertion, the method of pooling dates employed by Buchanan et al. did *not* bias their results.

The results of a study reported last year in this journal (Collard et al., 2010) also run counter to Steele's claim that Buchanan et al.'s method of pooling dates biased their results. Collard et al. (2010) compiled a sample of "clean" Clovis and Folsom radiocarbon dates from western North America and then used diffusion analysis and SPD analysis to examine the Clovis–Folsom transition. They found no evidence for a population bottleneck at the time of the putative ET impact or in the subsequent 800 years. Folsom-producing populations appear to have moved into unoccupied territory below 36° North latitude, but above that latitude there was overlap between Clovis and Folsom occupations. Thus, like the results of Buchanan et al.'s analyses, Collard et al.'s (2010) findings are inconsistent with the predictions of the ET impact hypothesis.

Steele's skepticism about CalPal is misplaced. The upper and lower SPDs shown in Fig. 2 were generated from the same dataset and calibration curve with CalPal and Calib, respectively. The dataset comprises 603 artificial uncalibrated dates that cover the time period 13,000–8000 BP in increments of 25 years and have standard errors of ±25 years. Calibration was carried out with the IntCal04 curve. Given that the number of dates per time interval ($n = 3$) does not vary and the dates have the same standard error, the expectation is that an SPD generated from the dataset will be more or less flat and therefore consistent with a constant population size. As can be seen, this expectation is met by the SPD yielded by CalPal but not by the SPD produced by Calib. The former is flat for the most part, whereas the latter contains a number of marked peaks and troughs. We also ran the analysis with uncalibrated standard errors of 50 and 75 years, and with the HULU calibration curve in CalPal, and obtained the same pattern each time. The obvious implication is that, contrary to what Steele contends, the SPDs produced by CalPal are actually *better* for investigating prehistoric demographic change than those produced by Calib and, by extension, OxCal.

The inaccuracy of Steele's claim about the utility of CalPal for demographic reconstruction is further demonstrated by the two SPDs shown in Fig. 3. As before, these were generated from the same dataset and calibration curve with CalPal and Calib. This time, however, the artificial dataset contains a "population spike" at 10,500 BP caused by the addition of 27 dates. Thus, the expectation is that the SPD should be flat, apart from a marked peak at the calibrated equivalent of 10,500 BP. Once again, it is obvious that the expectation is met by the SPD produced by CalPal but not by the SPD yielded by Calib. The former is more or less flat, apart from a large peak at approximately 12,600 BP. In contrast, the Calib SPD contains at least five major peaks. Significantly, the peak that corresponds to the 10,500 BP "population spike" in the uncalibrated dataset is smaller than the four other major spikes. Thus, the SPDs produced by Calib are not simply less robust than those produced by CalPal. Rather, because the "real" peak is smaller than the other major, program-induced peaks, they are positively misleading.

In addition to undercutting Steele's critique of CalPal, the foregoing analysis casts doubt on the results of his SPD analysis of Buchanan et al.'s data. The major peaks and troughs in the SPD we generated from the first artificial dataset with Calib closely match the major peaks and troughs in Steele's SPDs. Given that the artificial dataset was created to mimic a constant population, this indicates that Steele's SPDs cannot be used as a guide to demography. Their major peaks and troughs are either entirely or largely
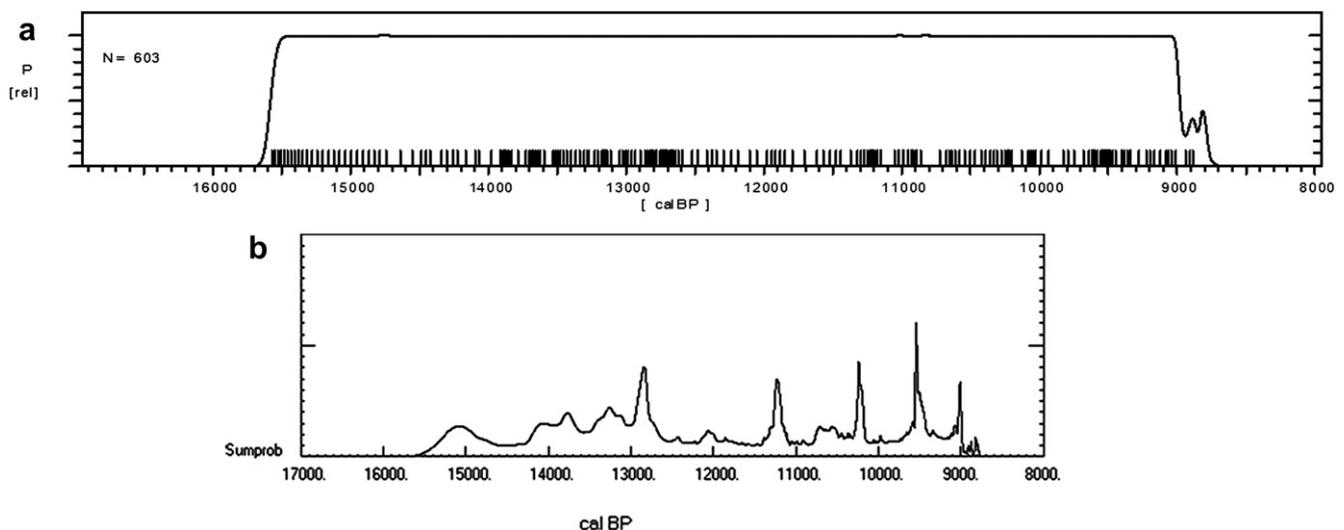

Fig. 2. Summed probability distributions generated from an artificial dataset of 603 uncalibrated radiocarbon dates designed to mimic an unchanging population. Top: Summed probability distribution calculated in CalPal using the IntCal04 calibration curve. Bottom: Summed probability distribution calculated in Calib 5.1 using the IntCal04 calibration curve.
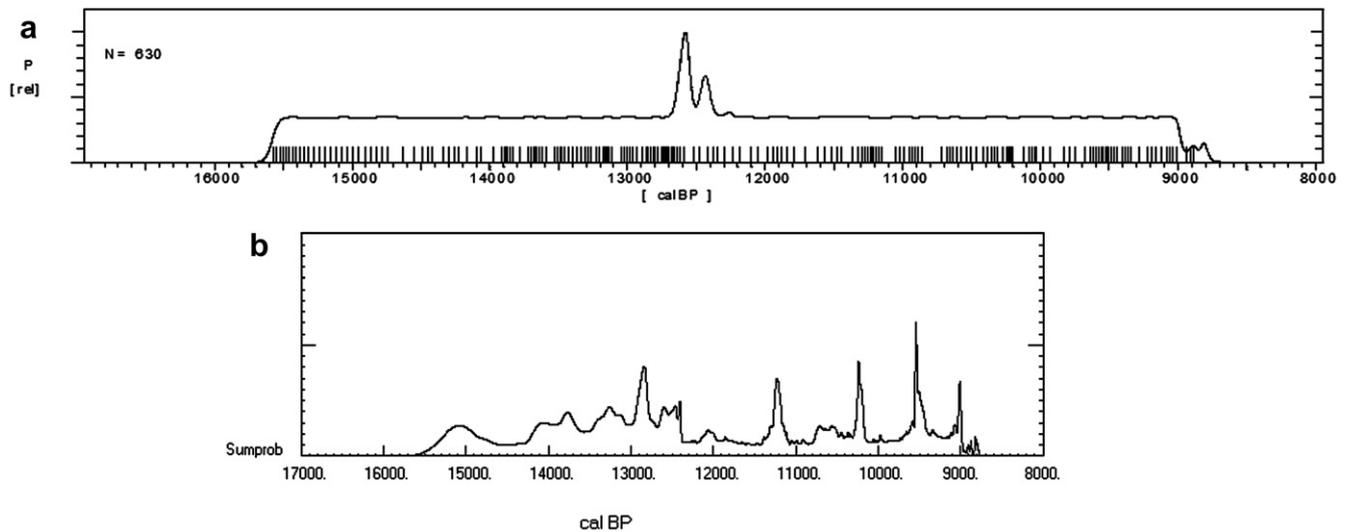
Fig. 3. Summed probability distributions generated from an artificial dataset of 630 uncalibrated radiocarbon dates designed to mimic a population spike. Top: Summed probability distribution calculated in CalPal using the IntCal04 calibration curve and the artificial dataset. Bottom: Summed probability distribution calculated in Calib 5.1 using the IntCal04 calibration curve.

an artifact of the programs used to generate them. The corollary of this is that Steele's claim that his SPDs support the ET impact hypothesis can be discounted.

Use of calibration programs that introduce false peaks and troughs into SPDs is not the only shortcoming of Steele's analysis. His decision to employ bootstrapping to create SPDs is also problematic. It makes sense to use bootstrapping to investigate spatial gradients in radiocarbon dates because such analyses are based on point estimates. However, SPDs take into account the entirety of the calibrated age ranges for a given set of dates, so there is no need to employ bootstrapping. In fact, SPDs generated with bootstrapping can be expected to be misleading. The bootstrap will sample high probability years within each calibrated age distribution more frequently than low probability years, and consequently the peaks and troughs of the SPD will be exaggerated. The obvious corollary of this is that there would be reason to reject Steele's conclusions even if the calibration programs he used were not demonstrably incapable of producing SPDs that can be used to infer past demography.

The problems with Steele's results do not stop there. His second analysis is not designed properly. The problem is the time period covered by the bins. Population size is known to fluctuate in the absence of extraterrestrial impacts. Thus, variation in the number of dated occupations through time is to be expected. The corollary of this is that the relevant test of the impact hypothesis is not whether there was a decline in population at 12,900 ± 100 calBP but whether there was a decline in population at 12,900 ± 100 calBP that *exceeded* population declines at other times in the past. The 400-year time period Steele used in his second analysis is simply too short to allow the predictions of the impact hypothesis to be tested properly. Buchanan et al.'s SPD illustrates this. It contains several declines that are not only as large as the decline at 12,900 calBP but also occur *after* the end of the period considered by Steele.

The results of Steele's third analysis are no more reliable. To reiterate, in the third analysis Steele assessed the "chronometric hygiene" of 20 of Buchanan et al.'s dates. He argued that five were pooled inappropriately (Smith Creek Cave, OTL Ridge, Eppley Rockshelter, Sunshine Locality, and Wilson Butte Cave) and that one is residual (Bolton Spring). He also claimed that three dates have been rendered obsolete by more recent determinations (Sheriden Cave-Layer 3, Folsom, and Sheaman). He concluded from this high proportion of "problematic" dates that Buchanan et al.'s dataset is

unreliable and that the results of their analyses are therefore invalid.

One problem is that many of Steele's judgments regarding the quality of the dates are open to question. To begin with, Steele used Calib's χ^2 test pooling procedure to decide whether the dates had been appropriately pooled. However, as we noted earlier, Buchanan et al. did not use a statistical test to pool dates because it would have required them to subjectively decide which date to keep in cases where dates from a single occupation are significantly different. Instead, they pooled all dates from a single occupation regardless of the significance of the differences among them. Thus, the five dates that Steele rejects as inappropriately pooled are only problematic *if* one accepts Steele's approach to pooling, which we do not.

The judgment that the 10,700 uncalBP date from Bolton Spring is residual is not clear-cut either. Regarding the dating of the cultural stratum at the site, the excavators note that "the fact that no older sources of contamination are known, but younger sources and mechanisms are present, is at least circumstantial evidence that the oldest date of about 10,700 yr BP may be the most representative of the true age of the cultural stratum" (Thorson and McBride, 1988, p. 231). Hence, it is unclear whether the 10,700 uncalibrated date from Bolton Spring is in fact erroneous.

Uncertainty also exists with respect to the dates from Sheriden Cave and Sheaman that Steele claimed have been rendered obsolete by more recent determinations. The newer date from Layer 3 at Sheriden Cave cited by Steele is statistically indistinguishable from the date used by Buchanan et al., according to Calib's χ^2 test pooling procedure ($T = 1.75$, which exceeds the test's significance threshold of $\chi^2 = 5.99$, df = 1, $\alpha = 0.05$). As such, it is not obvious that the older date should be regarded as problematic. The newer date for Sheaman that Steele cites was published by Haynes et al. (2004). However, the validity of the date in question has been disputed (e.g., Waters and Stafford, 2007), so again, it is not obvious that the date used by Buchanan et al. should be regarded as problematic.

The only date in the subsample examined by Steele that is indisputably problematic is the one from Folsom. Needless to say, this gives a much lower error rate than the one yielded by Steele's analysis (1/20 versus 9/20).

Even if Steele's error rate were defensible, the way it was generated means that it would be in appropriate to conclude that

the results of Buchanan et al.'s analyses are invalid. The problem here is Steele's decision to focus on dates from the 400 years between 12,900 and 12,500 CalBP. As we explained earlier, in order to test the ET impact hypothesis it is necessary to compare any decline in population that occurred at $12,900 \pm 100$ CalBP with population declines that occurred at other times in the past. The corollary of this is that to show that Buchanan et al.'s results are invalid it would have been necessary to assess the reliability of dates from across the 7000 years covered by the dataset and demonstrate that the problematic dates are distributed in such a way that the impact-induced population bottleneck at $12,900 \pm 100$ CalBP is probably masked. At 400 years, the time period Steele used is too short to allow the impact of "bad" dates on Buchanan et al.'s results to be assessed properly.

To summarize this section, then, Steele's criticisms of Buchanan et al.'s study are incorrect, and his reanalysis of Buchanan et al.'s data is flawed. Thus, contrary to what Steele suggests, his study does not cast doubt on Buchanan et al.'s findings.

4. Conclusions

We have shown that Steele's (2010) criticisms of Hamilton and Buchanan (2007) and Buchanan et al. (2008) do not hold water and have demonstrated that his re-analyses of Hamilton and Buchanan's (2007) and Buchanan et al.'s (2008) datasets are flawed. Given these findings, there is no reason to reject Hamilton and Buchanan's conclusion that the ice-free corridor model best explains the available Clovis-age radiocarbon dates or Buchanan et al.'s conclusion that when radiocarbon dates are used as a demographic proxy, there is no support for the hypothesis that the Paleoindians experienced a bottleneck as a result of an extraterrestrial impact above the Great Lakes at $12,900 \pm 100$ CalBP. In addition, we have highlighted some important issues for researchers interested in using radiocarbon dates to reconstruct population movements and demography. Most notably, we have explained why OLS regression is preferable to RMA regression when estimating diffusion velocity, and demonstrated that the summed probability distributions yielded by CalPal are more reliable as guides to past demography change than those produced by Calib and OxCal.

Acknowledgments

References

Bronk Ramsey, C., 2009. Bayesian analysis of radiocarbon dates. Radiocarbon 51, 337–360.

Buchanan, B., Collard, M., Edinborough, K., 2008. Paleoindian demography and the extraterrestrial impact hypothesis. Proc. Natl. Acad. Sci. USA 105, 11651–11654.

Collard, M., Buchanan, B., Hamilton, M.J.H., O'Brien, M.J., 2010. Spatiotemporal dynamics of the Clovis–Folsom transition. J. Archaeol. Sci. 37, 2513–2519.

Firestone, R.B., et al., 2007. Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. Proc. Natl. Acad. Sci. USA 104, 16016–16021.

Fort, J., Mendez, V., 2002. Wavefronts in time-delayed systems. Theory and comparison to experiment. Rep. Prog. Phys. 65, 895–954.

Fort, J., Jana, B., Humet, J., 2004a. Multidelayed random walks and application to the Neolithic transition in Europe. Phys. Rev. E 70, 031913.

Fort, J., Pujol, T., Cavalli-Sforza, L.L., 2004b. Palaeolithic populations and waves of advance. Camb. Archaeol. J. 14, 53–61.

Hamilton, M.J., Buchanan, B., 2007. Spatial gradients in Clovis-age radiocarbon dates across North America suggest rapid colonization from the north. Proc. Natl. Acad. Sci. USA, 15625–15630.

Haynes Jr., C.V., Kornfeld, M., Frison, G.C., 2004. Short contribution: new geochronological and archaeological data for the Sheaman Clovis site, eastern Wyoming, U.S.A. Geoarchaeology 19, 369–379.

Morlan, R.E., 2005. Canadian Archaeological Radiocarbon Database. http://www.canadianarchaeology.ca/.

Pinhasi, R., Fort, J., Ammerman, A.J., 2005. Tracing the origin and spread of agriculture in Europe. PLoS Biol. 3, 2220–2228.

Smith, R.J., 2009. Use and misuse of the reduced major axis for line-fitting. Am. J. Phys. Anthropol. 140, 476–486.

Steele, J., 2010. Radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities. J. Archaeol. Sci. 37, 2017–2030.

Stuiver, M., Reimer, P.J., Reimer, R.W., 2005. CALIB 5.0. http://calib.qub.ac.uk/calib/.

Thorson, R.M., McBride, K., 1988. The Bolton spring site, Connecticut; early Holocene human occupation and environmental change in southern New England. Geoarchaeology 3, 221–234.

Waters, M.R., Stafford Jr., T.W., 2007. Redefining the age of Clovis: implications for the peopling of the Americas. Science 315, 1122–1126.

Weninger, B., Jöris, O., Danzeglocke, U., 2007. CalPal-2007: Cologne Radiocarbon Calibration & Palaeoclimate Research Package. http://www.calpal.de/.