Supplementary Methods

The following sections provide further information on the methods used in Williams 2013 (doi: rsb20130486). Here I reproduce extensive parts of Williams (2012) from where many of the methods were adopted, and elaborate on any new processes undertaken as part of this publication.

Radiocarbon data as a proxy for human activity

One of the key aims of Williams (2012) was to determine how reliable the radiocarbon dataset was in providing a proxy for prehistoric human activity. It is a fundamental assumption that radiocarbon dates used in these analyses derive from occupation events.

This assumption is intrinsic to selection of archaeological samples for dating. A direct association is clearly evident for (a) dated hearths and fireplaces, burials, and shell middens but is less secure for (b) detrital charcoal from occupation deposits (which provide the majority of dates in archaeological datasets). The latter are generally assumed to be charcoal from human activity (e.g. from dispersed fireplaces). This is supported by the correlation between charcoal concentration and the density of other occupation debris (such as lithics and faunal bone) observed in most sites (e.g. Smith, 2006: Figure 19). Further support is provided by comparison and statistical correlation of summed probability plots for dated features (group (a) above) and detrital charcoal (Figure S1) showing that both record similar trends in Australian data. Pearson correlation coefficients of these data showed a significant correlation between trends shown in radiocarbon plots for occupation features and detrital charcoal over the last 20,000 years (Table S1 and Figure S1). The correlation was weaker prior to 20,000 years, reflecting the smaller number of dates in these samples, rather than necessarily a decoupling of the relationship.

For this study, I re-explored this issue with the continental wide radiocarbon dataset (n=4575). The aim was again to identify whether the entire dataset or a subset from (a) above would provide the most reliable results for reconstructing prehistoric population. It was also undertaken using several new procedures proposed by Peros et al. (2010) to address similar issues in their dataset. These investigations consisted of three different approaches:

1. Comparison of the entire dataset with a subset of those dates containing laboratory errors <100 years. Peros et al. proposed this approach to determine whether unusually large errors in some data significantly impacted the eventual probability distributions/histograms produced.
2. Comparison of the entire dataset with a subset of dates that could be directly correlated to human activity (e.g. burials, hearths, midden material, etc). This comparison was similar to those undertaken in Williams (2012) described above and was undertaken to address the concerns over the large number of detrital charcoal dates in archaeological sequences.
3. Comparison of the entire dataset with a subset of ‘occupation events’. These events were coined by Peros et al. (2010) to avoid the common issue of archaeological site duplication, and remove artificial peaks from the data due to multiple dates of the same archaeological feature, etc. The method involved the counting of each site only once per 200-year data bin, regardless of the number of times it appeared, and thereby remove multiple dates from the same stratigraphic unit or feature.
Comparison of the overall dataset with filtered subsets (1-3 above) show good correlation (Figure S2). Each subset contains at least 50% of the overall data and demonstrates similar trends. The occupation event subset contains the highest number of dates within a single subset (n=3,711 or 81%) and indicates that archaeological sample duplication is not a significant issue within the data. A Lin’s Concordance Coefficient test between the overall dataset and each subset indicates r values between 0.77 and 0.97, with reduced r values stemming from lack of early data (>20ka) in some subsets, rather than necessarily de-coupling of the relationship. This correlation suggests that removing dates with >100 year laboratory errors, or from detrital charcoal has little effect on the overall shape of the curve. For this reason I used the entire dataset in subsequent analysis.

One further issue that became apparent during the development of this and other publications was the potential impact of major climatic events, such as sea-level change and intense aridity during the Last Glacial Maximum (LGM), on the archaeological record. It is widely understood that large parts of the continental shelf fringing Australia would have been inundated during the Terminal Pleistocene, and any archaeological sites within these areas would have been lost. Similarly, a range of palaeo-climatic records have demonstrated hiatuses or disconformities during the LGM where increased aridity and windiness has led to a scouring or re-working of sediments and which may also have impacted archaeological deposits (e.g. Fitzsimmons et al., 2012). To test whether this had distorted the time-series trends I undertook correlations of the overall dataset with a range of subsets to explore the possible loss of data through climatic change. Specifically, I compared the overall dataset with:

1. Only dates recovered from rockshelters. This site type is thought to be comparatively unaffected by taphonomic loss through time (Johnson & Brook, 2011). Therefore it serves as a control set for any comparisons testing for loss of data through sea-level or post-depositional modification.

2. All dates excluding those from midden sites. This analysis was undertaken to identify how much influence those dates that exploit marine resources and/or are in close proximity to the coast have on the overall dataset. If the influence of these type of data can be determined when they are present (primarily in the Holocene), then the loss of such data in the Pleistocene through continental shelf inundation can be quantified. If correlation between the overall dataset and subset is high, it can be assumed that midden sites (whether present or taphonomically lost) have only minor influence in the broader trends identified.

3. Only dates from bioregions within 20km of a Pleistocene coastline. Due to the shape of the continental shelf, several parts of Australia have remained close to a coastline over the last 50ka, including (by IBRA bioregion) Carnarvon, Swan Coastal Plain, Warren, Esperance Plains, Naracoorte Coastal Plain, South East Corner, Sydney Basin, NSW North Coast, and Tasmanian West. None of these areas reveal evidence of extensive sea-level inundation or coastal evolution through the Terminal Pleistocene/Holocene (e.g. Morse, 1988). If the overall dataset correlates with this subset, it can be assumed that the loss of large parts of continental shelf have a negligible effect on the broader trends in the data.

The results of the above analysis are presented in Figure S3. Each of the subsets show good correlation with the overall dataset (r^2 = >0.95 in all cases), and suggest that systematic distortion of the time series trends is not a significant issue. Trends shown by the overall dataset correlate well with the subset of rockshelters, and (assuming rockshelters are unaffected by taphonomic loss) this implies that the dataset does not contain significant disconformities or hiatuses through this period. Similarly, the analysis shows that removal of midden data has a low influence on the shape of the time series
curves, and implies that sea-level change (i.e. inundation of sites) would have not significantly altered the broader trends in the dataset. This is further confirmed by the analysis of dates from stable coastlines where sea-level change has not been a factor (Figure S3:C), and correlation between the overall data and subset remains high. The same analysis was undertaken but comparing the subset with the entire dataset minus the subset data (Figure S4). The results produced similarly good correlations ($r^2 > 0.92$ in all cases) and supports the interpretations above.

To an extent, these results reproduce findings undertaken by Williams (2012), which demonstrated that a sample of 500 or more dates would reproduce an accurate representation of the overall dataset regardless of issues such as taphonomic loss and climate change.

This current analysis however is limited in the modern Australian landmass – and all population estimates related to this part of Sahul. Papua New Guinea is not included in the analysis, nor are the flooded continental shelves between Australia and Papua New Guinea.

While I acknowledge that climatic change may have led to re-working or loss of some archaeological data, most notably coastal sites along the (now) inundated continental shelf, the analysis here suggests this has not significantly distorted the overall trends shown by the time series data. As proposed in Williams (2012), the large size of the overall dataset provides an averaging effect on the trends that may off-set the impacts of climate change on local and regional archaeological data. For these reasons, I consider the dataset to represent a continuous broad trend in human behaviour over 50ka with only minimal influence from post depositional and climatic change.
Figure S1: Number of radiocarbon dates for the Williams (2012) dataset (solid line), detrital charcoal subset (dot and dashed line) and a subset of known occupation features such as hearths, midden, burials, etc (dashed line), corrected in accordance with taphonomic correction outlined below. Data presented as 3-point moving average (equivalent to 750 years). A statistical analysis of the overall dataset and the two subsets reveal close correlation over the last 20,000 years. This can be seen most clearly in the Holocene where all data shows similar trends, albeit at different magnitudes.

Table S1: Pearson correlation coefficient and significance for various time intervals, comparing radiocarbon data for occupation features and detrital charcoal.

<table>
<thead>
<tr>
<th>Period (cal. yrs BP)</th>
<th>Pearson correlation ($r$)</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 9,999</td>
<td>0.686</td>
<td>0.000</td>
</tr>
<tr>
<td>10,000 – 19,999</td>
<td>0.341</td>
<td>0.000</td>
</tr>
<tr>
<td>20,000 – 29,999</td>
<td>-0.290</td>
<td>0.069</td>
</tr>
<tr>
<td>30,000 – 40,000</td>
<td>0.349</td>
<td>0.025</td>
</tr>
<tr>
<td>Overall</td>
<td>0.341</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Figure S2. Plots showing only those radiocarbon data that: A) demonstrate errors less than 100 years; B) demonstrate a direct link to occupation activities (e.g. hearths, burials, middens, etc); and C) could be identified as ‘occupation events’ after Peros et al. (2010). The insets show linear regression between each subset and the overall uncorrected dataset. A Lin’s concordance coefficient analysis of these data indicate good correlation (r values as follows: A = 0.977; B = 0.770; C = 0.925) and demonstrate that the overall dataset provides a reliable curve for prehistoric activity.
Figure S3. Graphs comparing the overall dataset (black) with: A) a subset of data only from rockshelters (grey) (n=1,971); B) a subset of the entire dataset excluding midden data (grey) (n=3,553); C) a subset of only data in bioregions that would have been in close proximity to the coastline over the last 50ka (n=1,375). The insets show linear regression between each subset and the overall uncorrected dataset. A Pearson correlation coefficient analysis of these data indicate good correlation (r^2 values as follows: A = 0.963; B = 0.997; C = 0.983).
Figure S4. Graphs as presented in Figure S3, but here the subsets are compared against the entire dataset minus the subset data in question. This approach was considered to give a more rigorous correlation than those in Figure S3. A) a subset of data only from rockshelters (grey) (n=1,971); B) a subset of the entire dataset excluding midden data (grey) (n=3,553); C) a subset of only data in bioregions that would have been in close proximity to the coastline over the last 50ka (n=1,375). The insets show linear regression between each subset and the overall uncorrected dataset minus the subset data in question. A Pearson correlation coefficient analysis of these data indicate good correlation ($r^2$ values as follows: A = 0.92; B = 0.96; C = 0.96).
Taphonomic Correction

The potential effects of time-decay on radiocarbon summed probability plots has been highlighted in many studies (e.g. Fiedel & Kuzmin, 2007; Marwick, 2009; Michczyńska et al., 2007; Smith et al., 2008; Steier et al., 2001) including several recent studies (Surovell & Brantingham, 2007; Surovell et al., 2009; Peros et al., 2010). The most influential is by Surovell & Brantingham (2007) who noted that sum probability distributions in very different fields - archaeology, palaeontology and geology - all revealed a monotonically increasing pattern through time. They suggested this was a result of site destruction through time rather than a true representation of past events and showed that this pattern would be produced if an exponential decay function was applied to a dataset where the number of sites was stable over time.

A later paper, Surovell et al. (2009) sought to formally correct sum probability distributions for taphonomic bias through the comparison of taphonomically biased radiocarbon datasets and an independent record of the same events without such bias. They correlated a global volcanism signal from GISP 2 records with a radiocarbon database of global volcanic deposits (Bryson et al., 2006). The record from GISP 2 indicated a constant level of volcanism over the last 40,000 years, while the radiocarbon data showed a positive curvilinear shape reducing through time, which Surovell et al. considered to reflect taphonomic loss. By regressing a power function through the volcanic dataset (when plotted as a frequency distribution), they were able to develop an equation that corrected for exponential loss over time:

\[ n_t = 5.726442 \times 10^6 (t + 2176.4)^{1.3925309} \]  

(S1)

where \( n_t \) = number of radiocarbon dates surviving and \( t \) = time. It should be noted that this equation relates specifically to the volcanic dataset tested, and a further correction is required for other datasets (\( n_t = n_s / n_a \) where \( n_s \) = taphonomically corrected number of radiocarbon dates for the dataset of interest, and \( n_a \) is the actual number of radiocarbon dates present at a specific time \( t \) in the dataset of interest). This equation allowed for exploration of site loss through time, and indicated that within 1,400 years half of the original number of sites would be destroyed, by 10,000 years only 9% of sites would remain, and by 40,000 years only 1.6% (Surovell et al., 2009). Surovell et al. applied Eq. (S1) to archaeological sites in Wyoming, and found that when applying the equation to open sites, they correlated closely with uncorrected closed (rockshelter) sites providing a more reliable indicator of change through time. He considered that closed sites did not exhibit taphonomic loss and their need for correction was equivocal.

Peros et al. (2010) also applied Surovell et al.’s (2009) techniques. Using a novel approach of investigating Canadian one-cent coins between AD 1940 – 2010, he demonstrated a close correlation between documented mint coin production and corrected values from a jar of the same coins.

As part of a wider review of summed probability distributions, Williams (2012) also explored taphonomic loss. Starting with first principles, Williams attempted to replicate Surovell et al.’s results using the Bryson et al. (2006) global volcanic dataset. He showed increasing deviation in distributed residuals in the most recent and oldest time intervals of this dataset, as well as unstable variance (clustering or linearity) that changed over time. He considered that a square root data transformation gave better results, and developed an alternate correction equation to Surovell et al. (2009):

\[ n_t = 2.107 \times 10^7 (t + 2754)^{-1.526} \]  

(S2)
where \( n_t \) = number of radiocarbon dates surviving in the volcanic dataset and \( t \) = time. Williams highlighted, however, that this model did not differ substantially from results produced by Eq. (S1).

As part of Williams (2012) analysis, he found that the standard errors of Eq. (S2) showed increasing uncertainty for periods older than \(~15.0\) ka and that corrections after \(~25.0\) ka were highly variable (e.g. a sample at 10,000 years BP would require a scaling factor of 10.37 with a confidence level of between 9.62 and 11.22, in contrast a sample at 35,000 years BP would require a scaling factor of 54.37 with a confidence level of between 47.01 and 63.60) (Figure S5). He suggested that some caution should be used when interpreting corrected data for time intervals greater than \(~15.0\) ka. He further suggested that any correction of dates \(>40.0\) ka should be rejected, since the non-linear model extrapolates beyond this period due to lack of volcanic data.

Williams (2012) applied Eq. (S2) to a range of archaeological datasets across Australia (Figure S6). He did not divide open and rockshelter (closed) sites, since in Australia these categories are not discrete (many Australian rockshelters have stratified deposits extending beyond the shelter - Smith, 1986; Gould, 1978; Ward et al., 2006) and as dated open sites were otherwise poorly represented in the dataset used. He showed a good correlation with archaeological trends widely identified in the literature - low levels of human activity throughout the Pleistocene, especially during the Last Glacial Maximum (LGM) (Hiscock, 1988, 2008; Smith, 2006; Veth, 1989); increasing human activity in the early – mid Holocene climatic optimum between \(~8.0 – 6.0\) ka, which also correlates with increasing exploitation of shell beds (O’Connor, 1994; Przywolnik, 2005; Sim, 1994; Schrire, 1982), increasing use of marginal upland areas (Veth, 2005) and increasing territoriality (Smith et al., 2009); and a step-wise increase throughout the Holocene (Hiscock, 2008; Mulvaney & Kamminga, 1999) – but does not significantly modify trends identified in uncorrected summed probability plots (e.g. Smith et al., 2008). Eq. (S1) and Eq. (S2), however, gave unrealistic values for time intervals \(>25.0\) ka based on current colonisation models for Australia (Bowdler, 1977; O’Connell & Allen, 2004) and probably reflect the limits of theoretical corrections where independent data on rates of taphonomic loss is not locally available.

Williams (2012) concluded that although taphonomic corrections of the sort proposed by Surovell et al. (2009) have a role in analysing time-series data, several methodological issues needed to be taken into account. Other questions concerned the utility of theoretical taphonomic corrections of this sort. These are essentially a form of data transformation currently unconstrained by independent regional geomorphic data. Their uncritical application may obscure or dampen real trends in occupation, especially if the actual pattern is for a late Holocene increase which Eq. (S1) and (S2) will treat as simply an artefact of taphonomy. Where site visibility is strongly constrained by the age of specific landforms or depositional units (such as fluvial units or lunettes) (e.g. Holdaway et al., 2008, 2009) an assumption of exponential time-dependent taphonomic loss may not be strictly valid. This is particularly the case where Quaternary landforms are well preserved, such as the Willandra Lakes (Bowler, 1998) or the caves of southwestern Tasmania (Cosgrove, 1989), regions where late Pleistocene sites are better preserved and more visible than early-mid Holocene sites. Any taphonomic correction applied to these regions will exaggerate the representation of Pleistocene occupation in time-series plots. Williams (2012) recommended that taphonomic correction should not be routinely applied without some discussion of whether time-dependent taphonomic loss is valid as an \textit{a priori} assumption. The original summed probability plots should always be included for comparison if taphonomically-corrected curves are used.
For this study, I again looked at taphonomic correction for the entire dataset. I manipulated the data for both closed and open sites (Figures S7 and S8, respectively). While Surovell et al. (2009) and Johnson & Brook (2011) both suggest that only open sites require taphonomic correction, a range of publications indicate that rockshelters are also susceptible to post-depositional impacts (see Ulm, 2013: Sections 5 and 6 for discussion). As discussed above, there are also several sites that incorporate open sites and closed rockshelters. For these reasons, I also explored a combination of corrected (open) and uncorrected (closed) sites (Figure S9), and correction of the entire (both closed and open sites) dataset (Figure S10).

The analysis indicates that regardless of how the taphonomic correction is applied, the overall trends in the data remain largely the same. With the exception of closed sites, all taphonomic correction plots show a consistent pattern of slow decline in the Late Pleistocene followed by a peak at ~20ka, a sharp decline between ~18-13ka, and a step-wise increase through the Holocene. Figure S11 provides a summary of all the GRAnn values produced by the taphonomic data. As outlined below the GRAnn in this analysis is developed from the radiocarbon data and is used to infer population change through time, although it can also be interpreted as changes in human activity. Comparing the GRAnn values from each of the datasets shows close correlations; a Pearson’s correlation coefficient analysis between the uncorrected GRAnn curve and the combined (corrected open and uncorrected close site data) and taphonomically corrected (all data) GRAnn curves provide $r$ values of 0.74 and 0.76, respectively. (A Lin’s Concordance coefficient analysis of the same analysis produces $r$ values of 0.70 and 0.66, respectively). Comparison of the uncorrected GRAnn curve with taphonomically corrected open and/or closed sites provide a poorer correlation with $r$ values of 0.59 and 0.43, respectively, and suggest that either a combination of uncorrected and corrected sites and/or a correction of the entire dataset produce results closer to the original data than a subset of either.

Taphonomic correction of rockshelter data provides different trends than other forms of uncorrected or corrected data (Figure S8 and S10: C). The highest corrected values occur in the late Pleistocene with some minor declines at ~36-28ka and ~18-13ka, followed by a very gentle step-wise increase through the Holocene; but in general the results show a fairly constant population/activity over 50ka. This pattern is inconsistent with general archaeological records over the last 50ka (see discussion above), and suggest that taphonomic correction artificially flattens time series in rockshelter (closed) data(as proposed by Johnson & Brook, 2011). Further, correlation analysis between taphonomically corrected closed sites and a range of other corrected and uncorrected data showed relatively poor correlation ($r$ values generally <0.5) suggesting this form of modification does not improve the reliability of the data. Conversely, correlation of the uncorrected closed site data with the corrected open site data (Figure S12) – a method undertaken by Surovell et al. (2009) to demonstrate reliability of the correction method – provided a relatively strong correlation (0.66) and allows greater belief in the shape of a curve when combining the two (i.e. Figure S10).

The results here provide greater reliability in the taphonomic correction methods adopted for the main publication. Figures S7-S10 suggests that regardless of the taphonomic methods adopted the broad trends in the data will remain (Figure S11), and they do not significantly differ from the uncorrected data (when modified to GRAnn values). Statistical analysis similarly shows that they all correlate well (with the exception of the closed site data). The application of the taphonomic correction to the close sites has been explored and the results equivocal, especially when combined with other data (e.g. Figure S10). However, closer correlation could be achieved when the closed site data remained uncorrected and compared with corrected open site data, a method also proposed and adopted by Surovell et al. (2009). This approach conforms to other studies that suggest rockshelter sites are not
exposed to taphonomic loss through time, and that these sort of theoretical corrections should not be applied (e.g. Johnson & Brook, 2011).

This analysis indicates that the combination of corrected open sites combined with uncorrected rockshelter data provided the most robust approach.

As outlined in Williams (2012), taphonomic correction is based on a global volcanic dataset. No Australian correction curve is currently available, and this may introduce some errors in the corrected time series curve. As outlined in previous publications, the Eq. (S2) overcorrects data >25ka for this reason, and I suggest that any corrections for the period >25ka should be considered with caution.

A further limitation (also identified in Williams 2012) is the modification of archaeological data that does not necessarily need correction due to good preservation of older Pleistocene landscapes. This is most noticeable in the Murray Darling Basin and southwest Tasmania, where Pleistocene sites are more common than Holocene sites (see above). As a coarse measure to determine the influence of these landscapes, I have removed the Murray Darling Depression and Tasmanian bioregions and re-ran the analysis (Figure S13). While the Terminal Pleistocene and Holocene trends remain largely unchanged, the most significant difference is removal of the large peak during the LGM. This shows that these are specific regional trends rather than continental trends. In this paper, I have not removed these bioregions, but here, I simply highlight their influence and limitation on the overall time series curve.
Figure S5. Square-root non-linear models (solid line) adjustment or scaling factor of the volcanic dataset through time based on Eq. (S2) and 95% confidence levels (dashed lines). Note the divergence of the scaling from ~15,000 – 40,000 years BP. (Source: Williams, 2012)
Figure S6. Taphonomic correction of archaeological radiocarbon data based on methods developed and applied in Williams (2012) and discussed above: A) Data from the northern two thirds of Australia (Austarch 1, Austarch 2, and IDASQ); B) Queensland coast region; and C) the arid zone (Austarch 1). Dashed lines represent the uncorrected frequency distributions (number of calibrated dates by time interval/overall number of dates of the dataset in question – a coarse version of a summed probability plot) – right hand axis; dark grey line provides the corrected values with associated 95% confidence intervals (pale grey shading); black line presents a 3-point (equivalent to 750-year) moving average. No smoothing or zero-trimming was attempted, since this removed time intervals with an absence of radiocarbon data that may indicate a human response. Climatic periods are shown, including the Last Glacial Maximum (LGM), Antarctic Cold Reversal (ACR), mid-Holocene climatic optimum (MHC) and intensification of El Niño Southern Oscillation (ENSO).
Figure S7. Taphonomic correction of radiocarbon data from open sites only (n=2,568) using methods outlined in Williams (2012). A) Uncorrected number of radiocarbon dates from open sites (white bars) and corrected number of dates (grey bars) with a 3-point (equivalent to 600 years) moving average trend-line; B) plot of GR$_{Ann}$ (population change) developed by applying Eq. (S3) to the corrected open site data; and C) the corrected dataset (grey bars) with interpolated trend line ($df = 25$) used to develop (B) (see below for further discussion).
Figure S8. Taphonomic correction of radiocarbon data from closed sites only (n=2,008) using methods outlined in Williams (2012). A) Uncorrected number of radiocarbon dates from closed sites (white bars) and corrected number of dates (grey bars) with a 3-point (equivalent to 600 years) moving average trend-line; B) plot of GR_{Ann} (population change) developed by applying Eq. (S3) to the corrected closed site data; and C) the corrected dataset (grey bars) with interpolated trend line (df = 25) used to develop (B) (see below for further discussion).
Figure S9. A combination of uncorrected closed sites and taphonomically corrected open sites using methods outlined in Williams (2012). A) Cumulative graph of uncorrected number of radiocarbon dates from open (black bars) and closed (white bars) sites; and a compilation of uncorrected closed sites and corrected open sites (grey bars) with a 3-point (equivalent to 600 years) moving average trend-line; B) plot of GR\textsubscript{Ann} (population change) developed by applying Eq. (S3) to the combination of uncorrected closed site and corrected open site data; and C) the combined dataset (grey bars) with interpolated trend line ($df = 25$) used to develop (B) (see below for further discussion).
Figure S10. Taphonomic correction of radiocarbon data of both closed and open sites using methods outlined in Williams (2012). A) Uncorrected number of radiocarbon dates from both open and closed sites (white bars) and corrected number of dates (grey bars) with a 3-point (equivalent to 600 years) moving average trend-line; B) plot of GR$_{An}$ (population change) developed by applying Eq. (S3) to the corrected dataset; and C) the corrected dataset (grey bars) with interpolated trend line ($df = 25$) used to develop (B) (see below for further discussion).
Figure S11. A summary diagram of all GR$_{Ann}$ plots developed through the various taphonomic corrections after Williams (2012). In this analysis, the GR$_{Ann}$ is considered to reflect population change, but an alternative interpretation can consider these plots to simply represent a change in human activity through time. A) GR$_{Ann}$ values from the entire uncorrected dataset; B) GR$_{Ann}$ values from taphonomically corrected open sites only; C) GR$_{Ann}$ values from taphonomically corrected closes sites only; D) GR$_{Ann}$ values from a compilation of uncorrected closed sites and corrected open sites; E) GR$_{Ann}$ values of the entire dataset following taphonomic correction. Note: Regardless of the taphonomic correction applied or, the broad trends (and in many cases similar values) occur across all plots (A-E); this is most evident in declines at ~32-24ka, 19-12ka, 7-6ka, and peaks at ~20ka, 12-8ka, 2-1ka.
Figure S12. Relative frequency chart of uncorrected closed sites (grey) and taphonomically corrected open sites (black). Visually, there is close correlation through the Terminal Pleistocene and early-mid Holocene with other periods de-coupling. A Pearson’s coefficient correlation between the two datasets provides an $r$ value of 0.52, which increases to 0.66 when comparing only the last 25ka.
Figure S13. Taphonomic correction of radiocarbon data of the overall dataset excluding the Murray Darling Depression and Tasmanian bioregions (n=3,918). A) Uncorrected number of radiocarbon dates from the entire dataset (black bars) and the data used in this analysis (white bars), and corrected number of all dates (grey bars) with a 3-point (equivalent to 600 years) moving average trend-line; and B) Uncorrected number of radiocarbon dates from the entire dataset (black bars) and the data used in this analysis (white bars), and corrected number of open sites combined with uncorrected closed site data (grey bars) with a 3-point (equivalent to 600 years) moving average trend-line.
Palaeo-populations

Peros et al. (2010) recently used radiocarbon data to develop quantitative palaeo-Indian populations for North America. Using more than 25,000 radiocarbon dates spread across the continent, Peros et al. developed a method of converting numbers of radiocarbon dates into an average annual change in population through time (GRAnn). They then applied this equation to a range of founding populations to determine the population of palaeo-Indians through time. Here I use the same approach and methods to develop similar results for prehistoric populations in Australia.

The method developed by Peros et al., first included the calibration of all radiocarbon data. Using the median value, each date was then divided into 200 year data bins of ‘number of dates’. (Data-binning is a form of quantization – mapping a large set of input values into a smaller set and reducing minor observational error). A smoothing spline was then run through the data bins, with subsequent analysis using interpolated values from this spline. The reason for the introduction of the spline and use of interpolated values was two-fold: 1) it removed extreme values and outliers from the data-bins; and 2) most importantly it removed zero values from the data, which are problematic when applying Eq. (S3). In relation to (1), this was controlled by the degrees of freedom (df) used to develop the spline; a lower df reducing the extreme values in the data. Peros et al. adopted a df value of 25 for their analysis, whereas for my analysis I explored a range of df values (15-200) and considered both 25 and 50 to provide a good balance between data variability and coherent results.

Using the values interpolated by the spline, Peros et al. applied an equation to determine annual percentage growth rate (GRAnn) as follows:

$$GR_{Ann} = 0.5((d_2 - d_1)/d_1) \quad (S3)$$

where in a given pair of consecutive 200-yr data bins, $d_2$ is the number of radiocarbon dates in the younger bin, and $d_1$ is the number of dates in the older bin. Each GRAnn value was multiplied by 0.5 to convert to a percentage (i.e. multiple the value by 100) and to produce an annual rate from the 200-year bins (i.e. divide each 200-year bin by 200 to obtain an annual value). In its simplest form Eq. (S3) simply shows the change between each data bin (i.e. number of radiocarbon data per 200 year period divided into annual periods); Peros et al. assumed that the number of radiocarbon dates directly correlated with population, and therefore the changes identified through this equation reflected differences in population growth or decline. Here, I also consider the change in data to reflect a population signal.

While Peros et al. use the GRAnn to re-create quantitative palaeo-Indian populations, they do not elaborate on the methods used to convert the GRAnn into actual population values. Here, I adopt a simple compound interest equation used commonly in the fields of banking and economics (Eq. S4):

$$P = f(1+GR_{Ann})^t \quad (S4)$$

where $P$ is final population, $f$ is initially the founding population followed by the $P$ value from each preceding 200-year data bin, the $GR_{Ann}$ is the relevant Eq. (S3) associated with each 200-year bin, and $t$ is number of years. The equation is then applied to each 200-year data bin and associated GRAnn value through time to create population change from 50ka to Contact (Supplementary Data). So once an initial founding population (see below) is entered into the equation at 50-49.8ka (the first 200-year data bin in this analysis) and the relevant GRAnn applied, the result of this analysis is then placed into the same equation as $f$ for the 49.8-49.6ka data-bin (the second 200-year data bin in this analysis) and...
the relevant GR$_{Ann}$ value applied, and so on until 0ka is reached. Using a hypothetical example: Introducing a founding population of 100 at 50-49.8ka data bin and applying a GR$_{Ann}$ of 5% would equate to a final population value of 105 \[ P = 100(1+0.05)^{200} \] for this data bin. (Note the GR$_{Ann}$ is presented here as a decimal.) Continuing the example, applying population of 105 to the next data bin (49.8-49.6ka) with a GR$_{Ann}$ of -10% would result in a final population of 94.5 \[ P = 105(1+-0.10)^{200} \]. This approach is applied to each data bin until 0ka is reached to produce the final population figures outlined in the main publication and the Supplementary Data.

As outlined in the introduction of the main publication, there are a range of founding populations and colonisation dates for the migration of people into Australia. With regards to colonisation dates, the most conservative is currently considered to be 46ka (O’Connell & Allen, 2012), while my dataset does not extend >50ka. Therefore, these two values delineated the maximum and minimum colonisation dates I used as the start of my population calculations.

For founding populations, early researchers considered a small family group or band (<50) was considered likely, whereas recent DNA analysis suggests numbers in the hundreds and probably low thousands were required (see main publication for discussion). I therefore used a range of founding populations (50, 500, 1000, 2000, 3000, 5000) to apply Eq. (S4) (Supplementary Data). As with Peros et al., the aim of the palaeo-population analysis was to provide a more reliable indication of the populations at time of Contact. Demographic studies (outlined in the main publication) indicate that values of between 300,000 to 1 million were the most likely, and therefore the founding populations applied to Eq. (S4) were used to reproduce values that fell within this range at 0ka; I found that values in excess of 5,000 people at 50 and 46ka reproduced Contact populations well in excess (>2 million) of recorded values, and this therefore provided a maximum founding population. Conversely, a founding population of 50 produced very low numbers at time of Contact; and this value therefore formed the lowest founding population tested. The remaining values provided a range between 50 and 5,000 with which to best reproduce Contact populations in accordance with the observed range above.

In addition, as outlined above, I manipulated the degrees of freedom for the smoothing spline of the data developed through Eq. (S3). This had the effect of increasing/decreasing the extreme values in the data, and provided large differences in the palaeo-populations produced by Eq. (S4). As with the founding populations, I found that $df$ values of $>$50 caused very large and unrealistic populations at Contact, whereas $<$25 failed to see any increase in populations through time. These two $df$ values, therefore provided the maximum and minimum values for analysis and results presented in the main publication.

I explored palaeo-populations with both uncorrected and corrected radiocarbon data. As found by Peros et al. in their study, the corrected data failed to reproduce reliable palaeo-populations; they found that initial colonising populations would have to have been well in excess of documented or postulated values to reach Contact population estimates. In my analysis, I required a founding population of over 30,000 to reproduce observed contact values (Supplementary Data). This value is, similarly well in excess of any reasonable values considered by existing literature, and therefore only the uncorrected data is included in the main publication.
Additional Figures

Figure S14. The overall dataset calibrated and data-binned into 200 year intervals and compared with the summed probability plot. Close correlation can be observed between the two techniques.
Figure S15. The uncorrected 200-year data-binned dataset divided by latitude: A) 9 - 20.3° S (n=1,282); B) 20.3 – 31.6 ° S (n=1,256); and C) 31.6 – 42.9 ° S (n=2,037).
Figure S16. The uncorrected 200-year data-binned dataset divided by longitude: A) 113 - 123° E (n=551); B) 123 – 133 ° E (n=484); C) 133 – 143 ° E (n=1,153); and D) 143 - 153° E (n=2,387).
Figure S17. A plot of population estimates from 50 - 0ka based on taphonomically corrected data (corrected open combined with uncorrected closed site data). Each graph was developed by implementing founding populations at 50ka and applying Eq. (3). A) Population estimates based on GR_{Ann} values developed using a spline with a \( df = 25 \); and B) as A but using a spline with \( df = 50 \).
Figure S18. The uncorrected 200-year data-binned dataset divided by site types: rockshelters (n=1,971); middens (n=1,071); and open sites (n=828).
Figure S19. Comparison of the GR_{Ann} from the uncorrected dataset (A) and the taphonomically corrected dataset (B) with EPICA ice core deuterium/hydrogen ratios (C) (EPICA Community Members, 2004); sea-level change records (D) (Siddall et al., 2003); and El Nino Southern Oscillation records (E) (Rein et al., 2005). Note changes in the GR_{Ann} concurrent with cooling at ~26ka, ~38ka, LGM (~18ka), ACR (~14ka), mid-Holocene Climatic Optimum (~8-5ka) and ENSO (~4ka).
References


Veth, P.M. 1989. *Islands in the Interior. The Dynamics of Prehistoric Adaptations within the Arid Zone of Australia.* International Monographs in Prehistory: Michigan, USA.
