Contents

Introduction

This supplementary information provides further information about data selection and analytical steps taken in the analysis of SPDs. All scripts and data are available online via https://github.com/andybrown81/NZ_Population or via DOI 10.5281/zenodo.2250314.

Data

Acquisition

Analyses were carried out using a collection of radiocarbon dates sourced from the following locations:

- New Zealand Radiocarbon database (www.waikato.ac.nz/nzcd/)
- Published literature
- Grey literature, particularly unpublished cultural resource management reports provided by Heritage New Zealand Pouhere Taonga

All available radiocarbon dates from secure archaeological contexts (geological/palynological dates were omitted) were compiled in a database. The database includes contextual information

such as site name (Site Name), a unique site identifying number from the New Zealand Archaeology Associations site recording scheme (Site ID), information concerning the context from within the site that that material was sampled (Assemblage) and the analytical region each date has been placed into (Region).

The second set of information relates specifically to the date. This includes the unique laboratory number (Lab ID), the uncalibrated date (C14 Age) and error (C14SD), the δ13C (Delta 13) and information about the material (Material) and species (Material ID).

Finally the source of the above information (Source) is provided alongside the spatial information for each site in both the projected coordinate system (NZTM) and Latitude and Longitude.

Selection for analysis

In the first instance we applied a conservative selection criterion in terms of the materials used, selecting wood and charcoal dates and excluding shell and bone. The materials used were selected for two reasons. First the use of shell and bone introduces another level of inference in the form of reservoir corrections that we felt were better avoided when applying a novel methodology. Secondly, wood and charcoal dates were also selected because these materials are most likely to derive from activities that were consistently practiced across the archaeological sequence (e.g. fires for heating or cooking). Thus, by excluding shell and bone, we hope to minimise bias caused by temporal variation in cultural behaviour, particularly changing economic activities, which may mimic demographic shifts by inflating site-to-population ratios. For example, suppose a group of people entered a new environment and began to exploit a relatively easily won resource (e.g. a shellfish bed). Over time we may expect a decline in foraging efficiency that may lead to a change in the use of that resource (e.g. Allen 2012) or a shift toward more profitable practices – perhaps gardening. Thus, in such a case the underlying chance of obtaining shell dates will increase according to the intensity of resource exploitation, and not necessarily population. While this exact scenario may not always occur, the complexity associated with changing subsistence practices introduces a level of unknown bias into this type of analysis, which we would prefer to avoid.

In the next step, charcoal and wood dates were then put through to a hygiene protocol (e.g. Anderson 1991) in order to remove any potentially spurious dates. However, it should be noted

that, as our goal is not to provide increased chronological resolution of a specific event (e.g. colonisation) but to assess a large number of dates, our hygiene process was simplified compared to other criterions (e.g. Anderson 1991). Specifically, we only required dates to:

- Have secure archaeological context; however, unlike Anderson (1991) single dates from sites were retained for the purposes of this research.
- Derive from species considered reliable for dating and deemed to have minimal in-built age (McFadgen *et al.* 1994: Table S1). In cases where species identification is limited to 'short-lived species' we referred to the original report/publication where information regarding the dates were published. Where clear mention was made of a material selection process being followed to gain a radiocarbon sample appropriate for dating we included that date, if this was not present it was rejected.
- It should be noted that for some AMS dates the University of Waikato Lab has not reported the δ13C, although the fractionation has been accounted for in the CRA. In our judgement these dates are acceptable and are retained for our analysis.

Table S 1. The life expectancy of plant species commonly used for dating in New Zealand (from McFadgen *et al.* 1994)

Short (<100 years)	Medium (100-300 years)	Long (>300 years)
Aristotelia serrata	*Ackama rosifolia*	*Agathis australis*
Brachyglottis sp.	*Alectryon excelsus*	*Dacrydium cupressinum*
Carmichaelia sp.	*Beilschmiedia sp.*	*Halocarpus kirkii*
Carpodetus serratus	*Cordaline australis*	*Lagarostrobus colensoi*
Cassini asp.	*Corynocarpus laevigatus*	*Laurelia novaezelandia*
Coprosma sp.	*Discaria toumatou*	*Libocedrus bidwillii*
Coriaria sp.	*Dysoxylum spectibile*	*Metrosideros sp.*
Corokia macrocarpa	*Hoheria sp.*	*Nothofagus sp.*
Geniostoma rupestre	*Knightia excelsa*	*Phyllocladus sp.*
Hebe sp.	*Kunzea ericoides*	*Podocarpus totara*
Hedycarya arborea	*Myrsine divaricata*	*Prumnopitys spicatus*
Leptospermum scoparium	*Myoporum laetum*	*Vitex lucens*
Leucopogon fusciculatus	*Nestigis sp.*	
Lophomyrtus obcordata	*Olearia sp.*	
Macropiper excelsus	*Pseudopanax sp.*	
Melicytus ramiflorus	*Paratropus microphylla*	
Melicytus sp.	*Pittosporum eugenoides*	
Myrsine australis	*Pittisoprum tenufolium*	
Myrsine sp.	*Plagianthus sp.*	
Olearia rani	*Sophora microphylla*	
Psuedopanax arboreus	*Sophora sp.*	
Psuedopanax crassifolius	*Weinmannia sp.*	
Psuedowintera sp.		

Method

The methods used in this paper were based on quantitative analysis of summed probability distributions (SPDs) developed by Shennan *et al.* (2013), the non-parametric extension devised by Crema *et al.* (2016) and the spatial permutation test developed by Crema *et al.* (2017). These sources can be consulted for further information on the method; the basic workflow for each method is outlined below.

1. Developing SPDs

The first step in all analyses is the creation of SPDs. This process follows the following steps.

- Radiocarbon dates are 'binned' according to their archaeological context (Step A - Figure S2). When multiple dates are present from a single site they are aggregated on the basis of archaeological context and when their distance in ^{14}C age time was less than 100 years (the exact procedure consist of carrying out a hierarchical cluster analysis using the complete linkage method and then using a cut-off value of 100 years to separate the observations). Thus, were we to have a site with two occupation deposits (Layer 2 and Layer 4) our analysis would return at least two bins based on context alone. Were Layer 2 to contain three dates (A – 1400, B – 1420 and C – 1645), dates A and B would be aggregated but C would be placed in its own bin due to the distance in ^{14}C age time between it and A and B being greater than 100 years. The selection of 100 years for binning is arbitrary; however, bin sensitivity analysis (Figure S1) shows this choice has no impact on the accuracy of results.
- Dates are calibrated using the southern hemisphere 13 calibration curve (Hogg *et al.* 2013) using the *rcarbon* package (Bevan and Crema 2017), part of the R statistical environment (R Core Team 2014). Multiple dates within a bin are calibrated and summed 'inside' the bin and subsequently divided by the number of dates so that each

archaeological context contributes a single date distribution to the overall SPD (Step B - Figure S2).

- The pooled mean probabilities from the bins are summed to produce an empirically based SPD for each region (Step C - Figure S2).

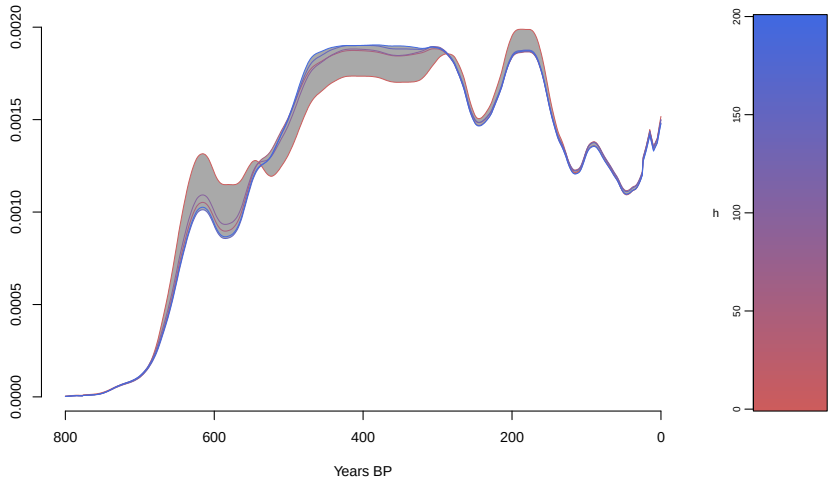

Figure S 1 – Bin sensitivity analysis showing the arbitrary choice of a 100 year cut-off has no impact on results (i.e. all bin sizes fit within the simulated envelope).

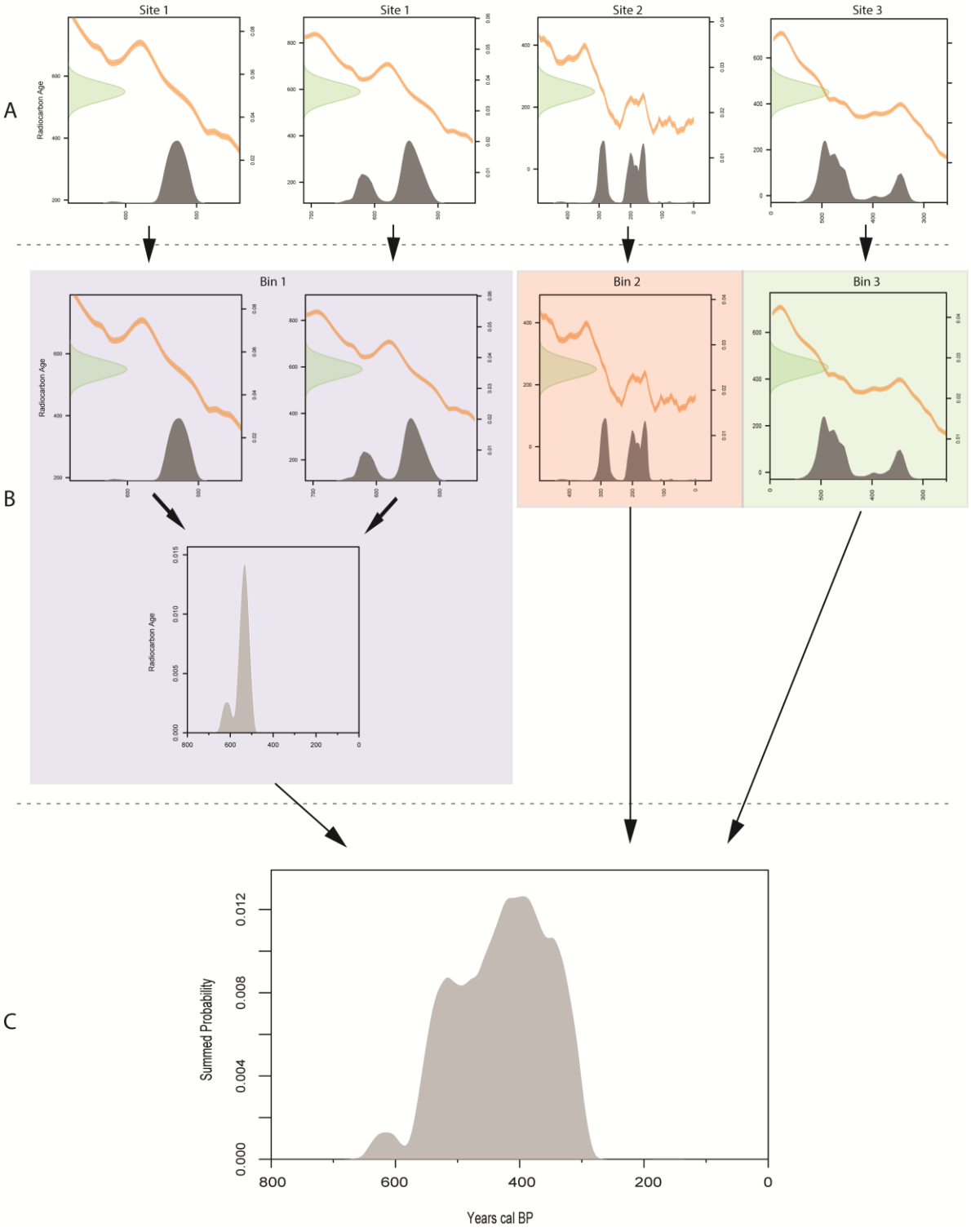

Figure S 2 – Basic workflow involved in creating summed probability distributions (SPDs)

2. *Model Testing*

The model testing procedure compares the observed SPDs with null models derived from known distributions of dates. The purpose is to test specific hypotheses about the pattern of population growth, in this paper we use a fitted logistic model of growth. The steps are as follows:

- Following the steps outlined above we created SPDs for New Zealand and three sub-regions – northern, central and southern (Step A - Figure S3).
- Each observed SPD has a logistic model fitted (Step B - Figure S3).
- A radiocarbon model is then simulated; the probability of each date being sampled is based on an adjusted fitted null model that takes into account the non-linearity of the calibration curve. Error ranges associated with each date are produced by randomly sampling with replacement from the pool of errors present in the observed data. The dataset is made up of the same number of bins that contribute to each regional SPD.
- The simulated uncalibrated dates are then calibrated and theoretical SPDRD are generated.
- This approach was repeated 5,000 times and a local Z-score produced to remove the effects of short term wiggles and other trends in the data. Using the simulated data, a 95% upper and lower envelope is computed (Step C - Figure S3).
- Observed SPDs are then compared to the simulation envelopes. Portions of the observed regional SPD that fall outside the envelope are said to be statistically significant local deviations from the null model (red and blue areas in Step D - Figure S3). Based on the methods outlined by Timpson *et al.* (2014) a significance value is then calculated by calculating the area outside the 95% confidence range for both the observed data and each simulated SPD. The proportion of simulations which have a summary statistic as or more extreme than the observed data provides the global *p*-value for each region.
- In the example below (Figure S3 – Step D) the plot of the left exhibits significant positive deviations (red area) and negative deviations (blue areas) from the logistic model. The plot on the right has no significant deviations from the model suggesting the fitted model is a reasonable approximation of the population curve in that example.

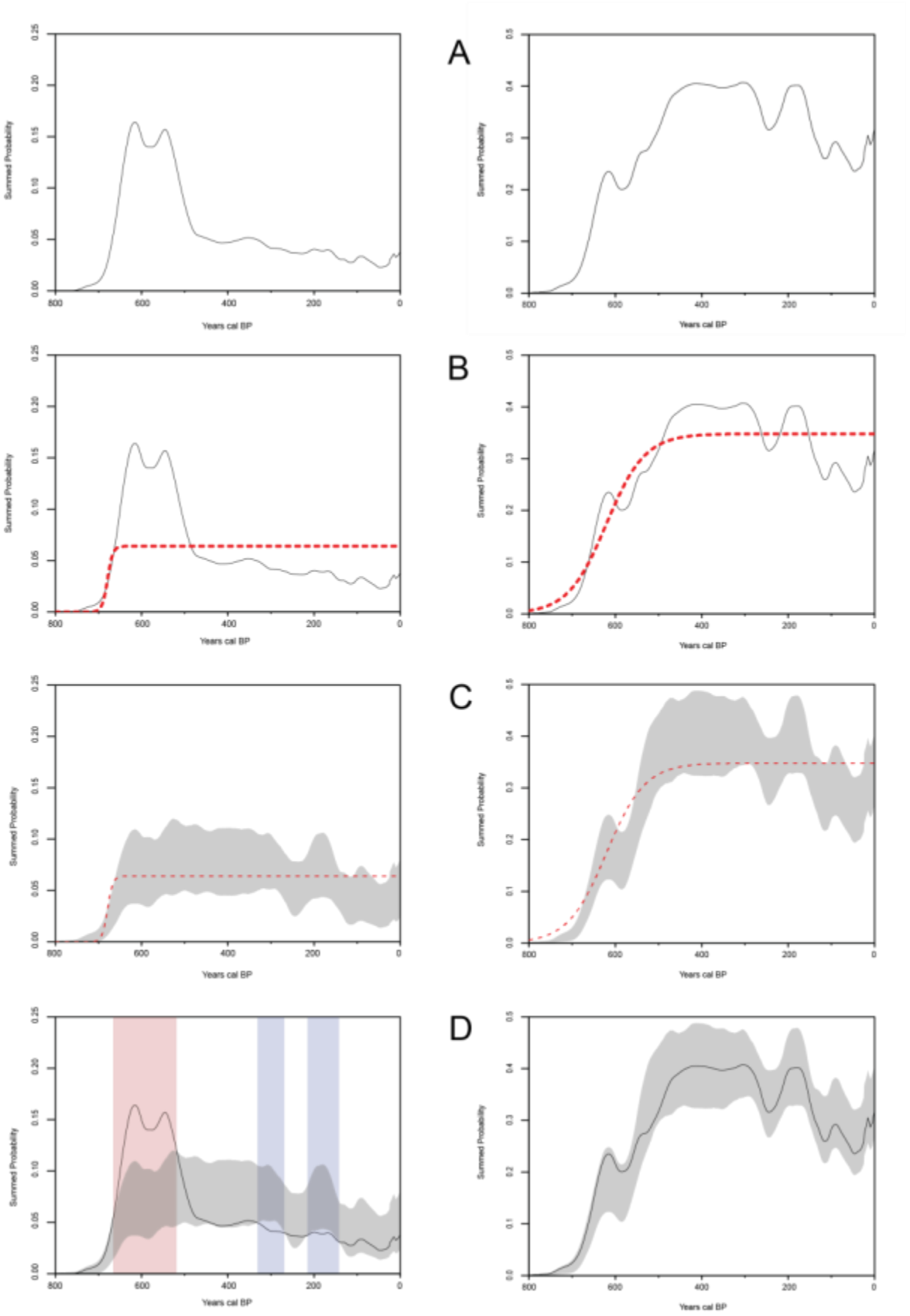

Figure S 3 – The basic steps involved in the model testing procedure outlined above.

3. *Regional Comparison (Permutation testing)*

This research is also concerned with empirically testing the variation between regions in New Zealand. To achieve this each region is compared to a null model based on the SPD of radiocarbon dates from across New Zealand.

- Regional SPDs are produced following the methods outlined above (Figure S4).
- To develop the null model the regional assignment of each bin is dropped to create a nationwide dataset. Uncalibrated dates are then randomly sampled from this dataset; the number of dates drawn is the same as the number of bins which contribute to each regional SPD. An SPD is generated from the randomly sampled dates (Step A - Figure S4).
- The previous step is repeated 5,000 times for each region and a local Z-score produced to remove the effects of short term wiggles and other trends in the data.
- Using the simulated data, a 95% upper and lower confidence envelope is computed (Step B - Figure S4). Observed SPDs from each region are then overlaid onto the confidence interval envelopes. Portions of the observed regional SPD that fall outside the confidence envelope are said to be statistically significant local deviations from the null model. Based on the methods outlined by Timpson *et al.* (2014) a significance value is then calculated by calculating the area outside the 95% confidence range for both the observed data and each simulated SPD. The proportion of simulations which have a summary statistic as or more extreme than the observed data provides the global *p*-value for each region.
- In the below example the observed southern SPD significantly exceeds the null model representing the general growth trends across New Zealand (red area, Step C – Figure S4) before declining below the null model (blue area, Step C – Figure S4).

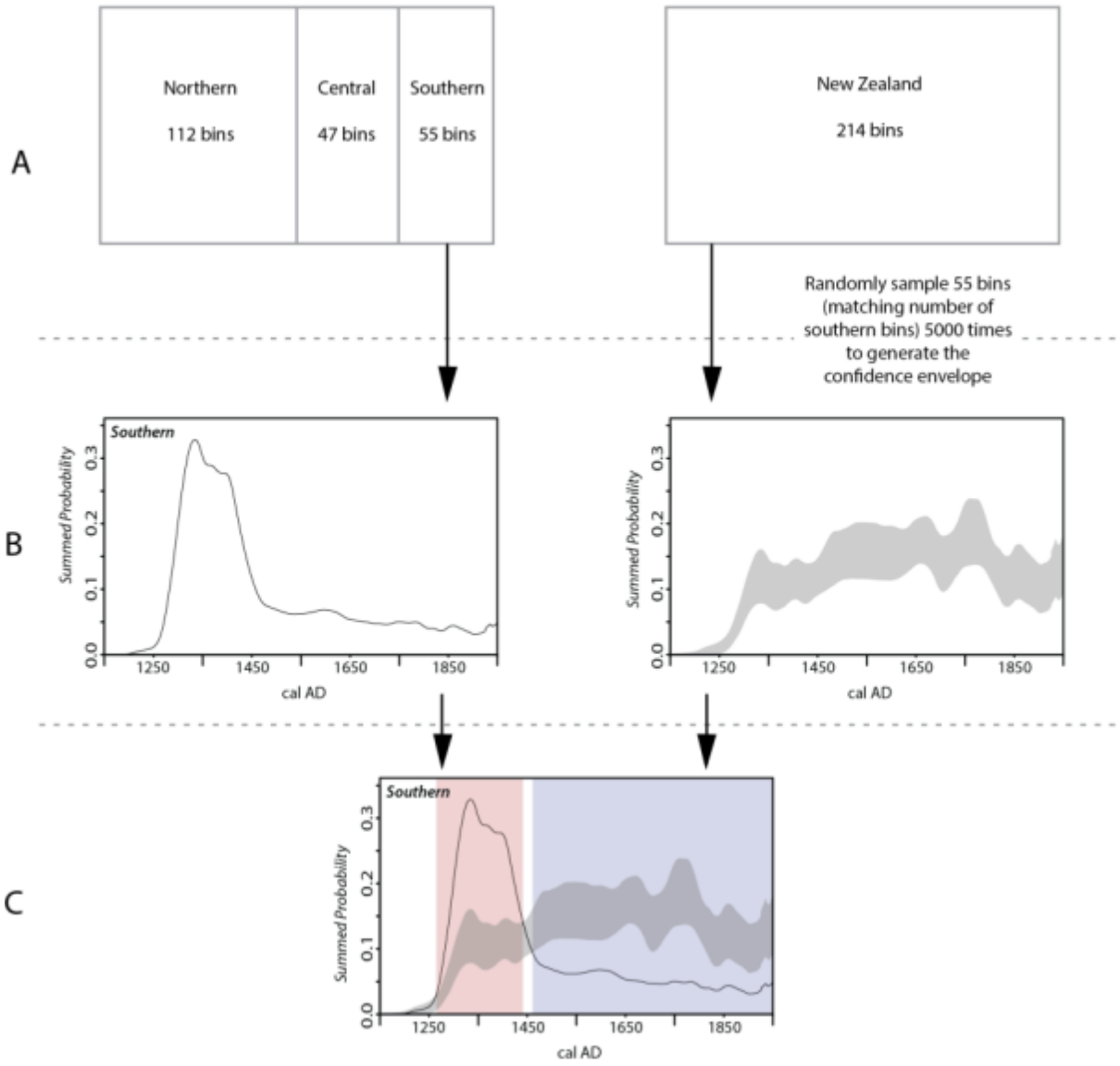

Figure S 4 – The steps involved in the permutation test procedure outlined above.

4. *Spatial Permutation Test*

The spatial permutation is an extension on the permutation test outlined above, with the crucial difference being that the spatial permutation tests allows for the assessment of variation without the imposition of *a priori* regions of analysis. The steps involved in the spatial permutation test are outlined in details by Crema *et al.* (2017) and can be summarised as follows:

- Site SPDs are created following the first three steps outlined in the 'developing SPDs section' above. The fourth step, combining local SPDs based on regions is not carried out.
- Local SPDs are created by weighting the contribution of sites based on their distance to focal site using a Gaussian distance decay function with a bandwidth of 100km with sites nearer to the focal contributing more to the local SPD.
- Sensitivity analysis of the spatial bandwidth used to create local SPD (Figure S5) around focal sites reveals a generally consistent pattern across all bandwidths, suggesting our conclusions can be regarded with a high level of confidence.
- Define the time slices to be analysed, the current research uses one-hundred year slices between AD 1200 and 1800.
- Weighted SPDs for each location and temporal slices are calculated using the methods outlined above.
- Calculate the geometric growth rate at the transitions between time slices. This provides the observed pattern of growth across New Zealand.
- The permutation test is carried out to test the significance of growth. This is done by randomly shuffling the bins from each site to new locations and calculating the weighted SPDs within each time slice. This is simulation process is carried out 5000 times.
- Hot and cold spots (areas of significance) are defined as areas where the local growth exceeds the growth observed in the simulations.
- Following the methods discussed in Crema *et al.* (2017), two measures of significance are produced in the course of the spatial permutation test. *P*-values are measures of significance between observed local growth and simulated growth rates. However, because of our use of a multiple testing approach, there is a potential for compounding false positive results (i.e. some local SPDs will be higher or lower than the theoretical expectation by chance alone). We therefore compute the more robust *q*-value by adjusting *p*-values to account for false discovery rate. Thus, in our results, a *p*-value of 0.05 suggests that 5 *per cent* of all tests will result in false positives, a q-value means that only 5 *per cent* of results with a value of 0.05 or below are false positives.

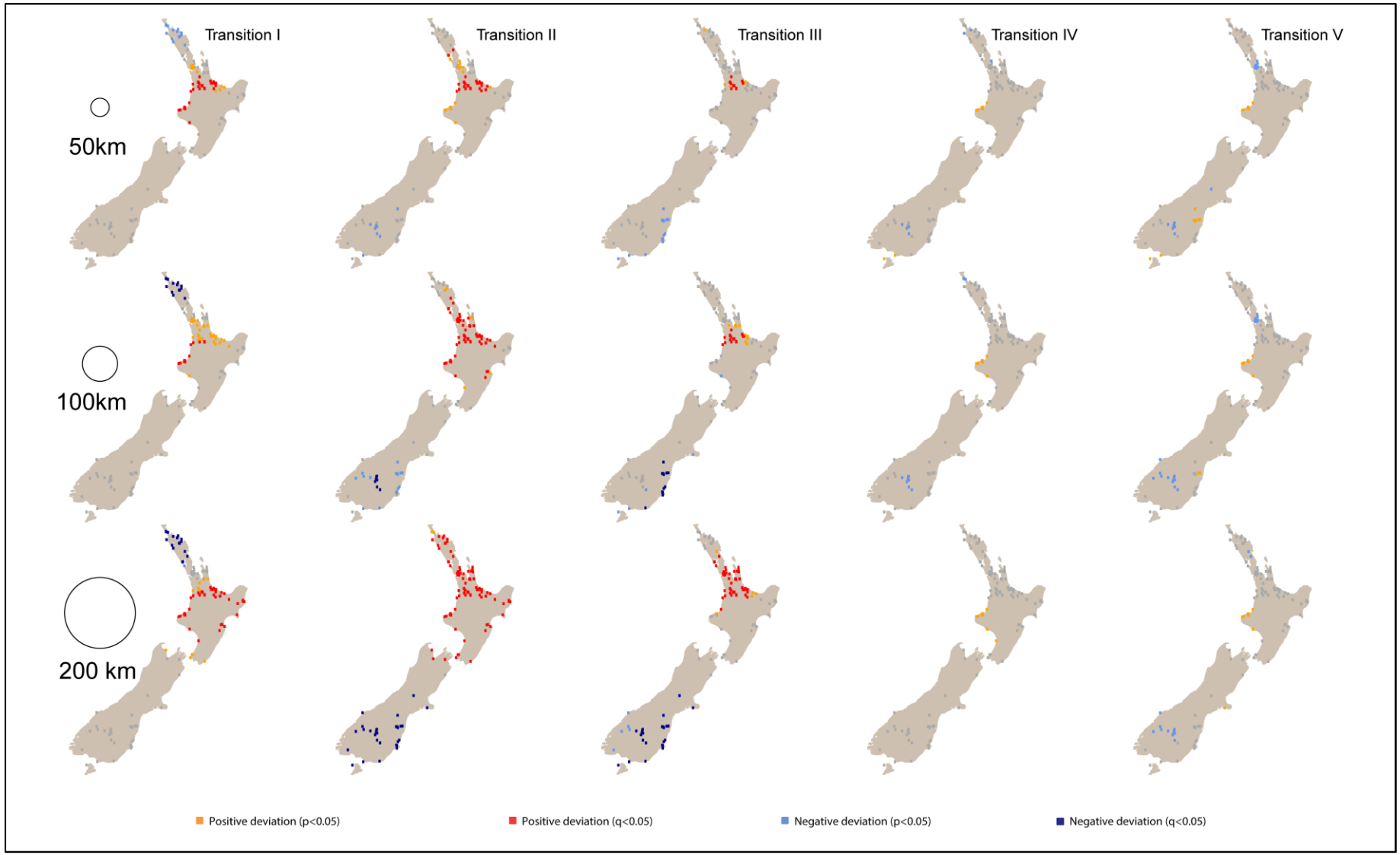

Figure S5 – Sensitivity analysis of the spatial bandwidth used for construction of local SPDRD around a focal site.

Discussion of Method

The methods used are robust to many potential biases in archaeological data. The papers from which these methods were drawn may be consulted for general discussion of methods, here we focus on concerns that may be raised from our own data.

The influence of standard errors

The chronometric hygiene approach introduced to New Zealand archaeology by Anderson (1991) included the focus on material type, context and the removal of large errors that may introduce 'noise' which blurs temporal signal. In these analyses we are concerned with material type and context, but have not culled dates based on wide standard errors. Instead the uncertainty introduced by errors is integrated. Specifically, the Monte Carlo simulation we employed to generate the envelope of expected summed probability values utilises the observed distribution of standard errors through re-sampling. This effectively generates larger envelopes with larger standard errors, making the rejection of the null hypothesis harder under those circumstances. In other words, although counterintuitive, larger errors actually lead to a more conservative analytical outcome and decrease the likelihood of making false claims.

Spatially heterogeneous sampling

There are clear regional differences in sampling intensity (e.g. northern region has nearly double the bins of the other regions) across New Zealand. However, our analysis overcomes this bias by comparing the shape of SPDRDs and not date density. More specifically, the spatial permutation method employed in this paper consists of comparing the shape of the 'local' SPDs (more specifically the sequence of geometric growth rates between abutting temporal blocks) against an envelope generated by randomly permuting the 'location' of the samples. This means that regions with lower sampling intensity will be characterised by a wider range of SPD shapes in the permuted sets, and hence wider simulation envelopes. In contrast regions with higher sampling intensity will be characterised by a narrower envelope (Figure S6). The specific null hypothesis being tested in this case is that growth rates are stationary over space. If all regions experienced the same growth trajectory then effectively the spatial marker of the sample wouldn't matter and we would observe a similar trajectory, although regions with a smaller number of dates could deviate more wildly due to higher sampler error. The method overcomes this issue by adaptively changing the simulation envelope based on the local density of samples.

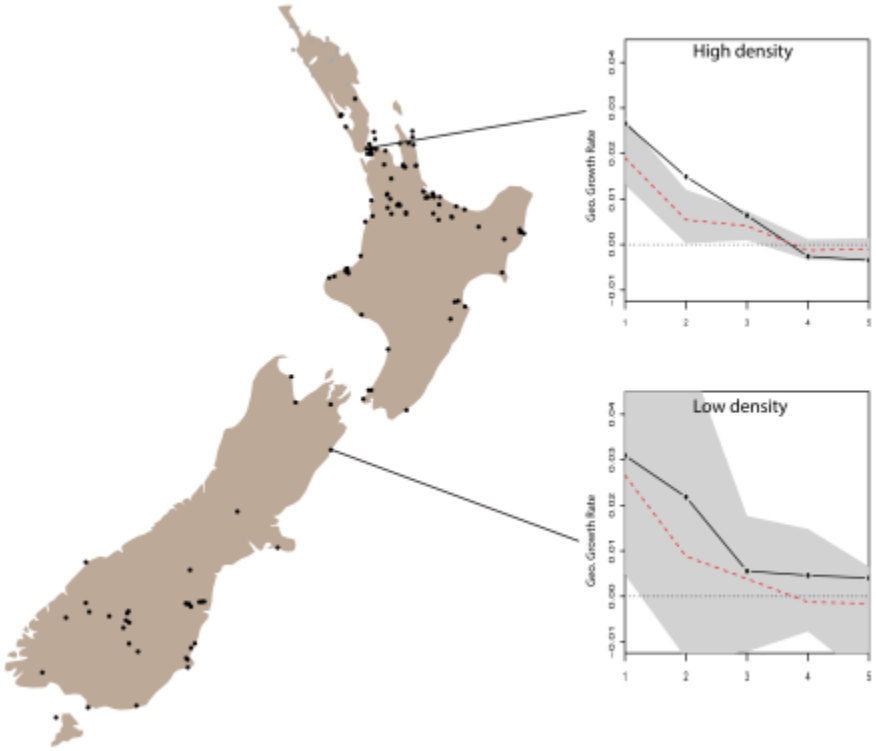

Figure S6: A graphical representation (similar to Figure 7 in our manuscript) of how our method deals with spatial heterogeneity in sampling. In areas of high date density the simulation envelope is narrower, while in areas of lower density it is much wider. Thus, the possibility of large deviation due to small sample size is mitigated by a higher threshold for significance.

Despite these methodological actions assessing temporal heterogeneity remains complex. However, we believe at the national scale that sampling of archaeological sites can be regarded as semi-random. This is because a large amount of the radiocarbon dates in our database arise from cultural resource management archaeology, which does not bias particular time periods. Rather, investigations are largely carried out in coastal areas, where early, mid and late sites are all concentrated. One area where temporal bias may occur is southern New Zealand, where in a large amount of research has focussed on the early and the very late periods of the sequence. However, a strength of the archaeology of the southern region is the large number of systematic surveys undertaken as part of infrastructure works (e.g. Mason et al. 1976) and, more recently, as a result of coastal erosion (Jacomb et al. 2010). The latter survey was conducted along the south coast of southern New Zealand and was followed up with an extensive dating programme, which Jacomb et al. (2010) suggest showed results consistent with a two-phase pattern of occupation consisting of early and very late in prehistoric phases. Thus, even when early sites are not

specifically the target of investigation; the data appears to fall mostly within this period suggesting the pattern is probably 'real'.

The influence of calibration

Difficulties imposed by 'wiggles' in the calibration curve (McFadgen *et al.* 1994) are overcome in this analysis through the comparison of observed data with theoretic models that have passed through the same calibration process. This process includes the effects of calibration in the model simulation envelopes, allowing us to determine if SPDRD fluctuations are likely to be genuine (i.e. are outside the envelope) or simply a result of idiosyncrasies in the calibration curve and/or sampling error (i.e. inside the envelope).

Despite these efforts, it remains possible that our results are influenced by other factors, such as temporal variation in site-to-population ratio. For example, a change in settlement size from large villages to smaller dispersed camps increases the number of residential features that may be dated. This may in turn inflate the SPDRD giving the appearance of higher population density although in fact it remains unchanged. In New Zealand the nature and abundance of resources had an impact on the aggregation and dispersal of Maori population (Allen 2012); however, the underlying pattern of settlement shows remarkable spatio-temporal consistency (Walter *et al.* 2006), reducing the likelihood of systematic bias in our results. Nevertheless, the identification of potential biases, together with the integration of other dating materials (e.g. marine shell), are key directions for the development of SPDRD analysis in New Zealand.

References

Allen, M. 2012. Molluscan foraging efficiency and patterns of mobility amongst foraging agriculturalists: a case study from northern New Zealand. *Journal of Archaeological Science,* 39(2): 295-307.

Anderson, A. 1991. The chronology of colonisation in New Zealand. *Antiquity* 65: 767-95.

Bevan, A. & E. Crema. 2017. Rcarbon: methods for calibrating radiocarbon dates.

Crema, E., J. Habu, K. Kobayashi and M. Madella. 2016. Summed Probability Distribution of ^{14}C Dates Suggests Regional Divergences in the Population Dynamics of the Jomon Period in Eastern Japan. *PLoS One* 11(4): e0154809.

Crema, E., Bevan, A. and S. Shennan. 2017. Spatio-temporal approaches to archaeological radiocarbon dates. *Journal of Archaeological Science* 87: 1-9.

Hogg, A., Q. Hua, P. Blackwell, M. Niu, C. Buck, T. Guilderson, T. Heaton, J. Palmer, P. Reimer, R. Reimer, C. Turney & S. Zimmerman. 2013. SHCal13 Southern Hemisphere Calibration, 0–50,000 Years cal BP. *Radiocarbon*, 55(4): 1889-1903.

McFadgen, B., F. Knox & T. Cole. 1994. Radiocarbon curve variations and their implications for the interpretation of New Zealand prehistory. *Radiocarbon* 36: 221-36.

R Core Team. 2017. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Shennan, S., S. Downey, A. Timpson, K. Edinborough, S. Colledge, T. Kerig, K. Manning & M. Thomas. 2013. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications* 4: 2486 doi: 10.1038/ncomms3486.

Timpson, A., S. Colledge, E. Crema, K. Edinborough, T. Kerig, K. Manning, M. Thomas & S. Shennan. 2014. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. *Journal of Archaeological Science* 52: 549-557.

Walter, R., I.W.G. Smith & C. Jacomb. 2006. Sedentism, subsistence and socio-political organisation in prehistoric New Zealand. *World Archaeology* 38(2): 274-90.

Williams, A. 2012. The use of summed radiocarbon probability distributions in archaeology: a review of methods. *Journal of Archaeological Science,* 39: 578–589.