

Estudios paleodemográficos basados en conjuntos de dataciones radiométricas. Una revisión crítica¹⁹²

DANIEL A. CONTRERAS¹ y JOHN MEADOWS^{2,3}

¹ DEPARTMENT OF ANTHROPOLOGY, UNIVERSITY OF MARYLAND, COLLEGE PARK, MARYLAND 20742, UNITED STATES (DANIELALEXANDERCONTRERAS@GMAIL.COM)

² ZENTRUM FÜR BALTISCHE UND SKANDINAVISCHES ARCHÄOLOGIE, STIFTUNG SCHLESWIG-HOLSTEINISCHE LANDESMUSEEN, SCHLOSS GOTTORF, D-24837 SCHLESWIG, GERMANY

³ CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL, LEIBNIZ-LABOR FÜR ALTERSBESTIMMUNG UND ISOTOPENFORSCHUNG, MAX-EYTH-STR. 11-13, D-24118 KIEL, GERMANY

21.1. Introducción¹⁹³

En un artículo de 1987, que ha sido citado casi exactamente el doble en su tercera década ($n = 154$ según Google Académico) que en los 20 primeros años después de su publicación ($n = 64$ según Google Académico), John Rick utilizó las dataciones radiocarbónicas del periodo precerámico andino central para argumentar que las fechas ¹⁴C podrían ser útiles para el análisis arqueológico a múltiples escalas (Rick 1987). Es decir, Rick argumentó que, junto al papel tradicional de las fechas absolutas para establecer marcos cronológicos para yacimientos y regiones arqueológicas las dataciones ¹⁴C también podrían ser utilizadas en los análisis de los patrones demográficos espaciotemporales a escala general. Su argumentación central fue que, a pesar de diferentes factores de confusión, las fechas radiocarbónicas podrían servir como un indicador de población, suponiendo que la producción de material fechable fuese aproximadamente proporcional al tamaño de la población.

El reciente aumento en el número de citas de dicho artículo (además de lo señalado arriba, Google Académico indica 24 citas en 2018) refleja la confluencia del interés continuado en la demografía prehistórica, la creciente disponibilidad de bases de datos con un número relativamente elevado de fechas radiocarbónicas, y la entusiasta acogida (aunque a menudo implícita) del argumento subyacente donde se vincula la producción de material fechable a la población.

Esta investigación se basa en la premisa de que los conjuntos de fechas ¹⁴C son indicadores válidos para estimar el tamaño de poblaciones humanas. Véase al respecto

¹⁹² Partes de este texto han sido adaptadas y actualizadas del artículo: Contreras, D.A., y Meadows, J. 2014. "Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach, *Journal of archaeological Sciences* 52: 591: 608.

¹⁹³ Traducción del original inglés por S. Pardo-Gordó y J.A. Barceló..

lo que se afirma en los capítulos 19 y 20 de este libro. Se argumenta que la “Suma de distribuciones de probabilidad [*sumatorios de probabilidad*]”¹⁹⁴ constituye una herramienta adecuada para resumir estos conjuntos. En general, este planteamiento ha sido impulsado por la conveniencia de las series demográfico-temporales como una variable explicativa en los análisis del cambio cultural (Shennan 2009) y por la resolución cronológica deficiente de otros indicadores demográficos potenciales y, especialmente, la relativa facilidad de la minería de datos con relación al radiocarbono (cf. Chaput y Gajewski 2015) en comparación con la obtención de datos e interpretación de variables más complejas y/o cualitativas, como el número de estructuras, las cantidades de artefactos o los patrones de asentamiento.

Como indica el recuento de citas en relación con el trabajo de Rick de 1987, esta literatura ha crecido tan rápidamente que ya no resulta práctico revisarla en detalle y con una perspectiva crítica. En lugar de intentarlo, en este capítulo tratamos que el lector tome conciencia de la amplia literatura existente y proporcionamos algunas preguntas críticas orientativas con las que consideramos que estos estudios deben ser evaluados.

21.2. La premisa y los problemas: dataciones como datos

Las estimaciones paleodemográficas como las presentadas en el capítulo 20 de este libro se enfrentan a dos desafíos básicos:

- a) si las fechas ¹⁴C pueden utilizarse como un indicador válido del tamaño de la población, y
- b) cómo pueden interpretarse y resumirse estos conjuntos de datos radiométricos, aceptando que el primer reto sea factible.

A continuación, revisamos estos dos problemas antes de pasar a una discusión más específica sobre el uso de las distribuciones de probabilidad acumulada de las fechas radiocarbónicas calibradas (suma de probabilidades), que se han convertido en la herramienta más popular para abordar el segundo desafío.

21.2.1. ¿Cómo pueden utilizarse las fechas ¹⁴C como indicador demográfico?

Con la finalidad de considerar un conjunto de fechas radiocarbónicas como un indicador demográfico válido, uno debe: a) aceptar la premisa básica de que la producción de material fechable es proporcional al tamaño de la población y b) estar convencido de que los problemas tafonómicos, sesgo de investigación y muestreo se han abordado de tal manera que cualquier patrón recuperable refleje con precisión la estructura de la población original (irrecuperable) de material fechable. Si bien muchos trabajos se han centrado en gran medida en los factores que potencialmente pueden afectar la fiabilidad de la estimación paleodemográfica, lo cierto es que la premisa básica no se acepta universal-

¹⁹⁴ Donde las fechas radiocarbónicas calibradas se combinan agregando la altura de cada distribución calibrada al calendario y normalizando la altura resultante de la función de la densidad de probabilidad acumulada.

mente (cf. Attenbrow y Hiscock 2015; Mökkönen 2014; Torfing 2015a). Las críticas se basan principalmente en si la relación entre la población y el material fechable es constante a lo largo del tiempo (en lugar de cambiar como lo hacen los patrones de comportamiento), así como cuestiones acerca de la perduración y recuperación de diferentes conjuntos de materiales arqueológicos.

21.2.1.1. La producción de material fechable

La premisa fundamental de cualquier aproximación de “dataciones en tanto que datos” reside en que la cantidad de material orgánico (potencialmente fechable por el método radiométrico) producida en un momento dado en el tiempo es proporcional al tamaño de la población. Desde el principio, esto se reconoció como una simplificación excesiva, y Rick incluyó la advertencia (que no fue siempre aceptada por investigadores posteriores) de que la cantidad de material fechable se entendía mejor como una medida relativa para la comparación entre poblaciones/regiones tecnológicamente y ambientalmente similares (Rick 1987: 57). Recientemente Freeman y colegas (2017) se han centrado en dicha cuestión, argumentando que la cantidad de material fechable se entiende mejor como una medida de consumo que de población. Esto, según los autores, introduce la necesidad de una argumentación intermedia que relacione el consumo energético con la población en cualquier contexto cultural y/o ambiental dado.

En cualquier caso, la investigación se ha centrado en el reto más próximo de tratar de estimar cuáles *eran* esas cantidades de material fechable. En relación con esto, tres cuestiones se resaltan: a) si el material orgánico fechable ha perdurado de manera significativa hasta el momento de la observación arqueológica; b) si este ha sido recuperado de manera fiable y objetiva en el momento de la excavación arqueológica; y c) si la muestra resultante de material fechable es suficiente y representativa.

21.2.1.2. Perduración y recuperación del material fechable

Estas cuestiones consisten en sesgos potenciales de tipo tafonómico y de investigación y de tamaños de muestra.

21.2.1.2.1. Tafonomía

Cualquier enfoque cuya lógica resida en la comparación diacrónica (en el caso de “dataciones como datos” como aproximación al tamaño de la población en el pasado, de la cantidad de material fechable producido en diferentes momentos) debe enfrentarse al problema de la tafonomía (en este caso, la perduración desigual de posibles muestras de ^{14}C). Esto puede, o no, presentar un reto significativo para la estimación de la población el pasado. La “corrección” para la destrucción a lo largo del tiempo, utilizando una curva de decaimiento exponencial simple (es decir, asumiendo una ratio constante de destrucción de los sitios arqueológicos; cf. Surovell y Brantingham 2007; Peros et al. 2010), puede proporcionar una visión más realista de la abundancia relativa del material fechable en diferentes periodos, y la corrección incluso puede no ser necesaria si el periodo analizado es relativamente breve en comparación con su edad absoluta.

Asimismo, la comparación dentro y entre conjuntos de datos que representan diferentes tipos de elementos arqueológicos (por ejemplo, asentamientos, cementerios, monumentos) han sido sugeridos y explorados como un medio para diferenciar entre los patrones “reales” y tafonómicos con relación a la abundancia de material fechable (e.g., Collard et al. 2010; Downey et al. 2014; Hinz et al. 2012; Riede 2009), o al menos como herramienta para evaluar la consistencia entre diferentes indicadores (por ejemplo Crema et al. 2016; Palmisano et al. 2017).

Sin embargo, a escala local, la tafonomía puede producir patrones falsos; donde el colágeno óseo se conserva deficientemente, por ejemplo. Por lo demás, este es un indicador radiocarbónico perfecto que podría rastrear la popularidad del rito funerario de cremación, en lugar de la población en su totalidad. Se supone que, en grandes conjuntos de datos, en nuestra opinión de forma irrealista, los filtros tafonómicos no discriminarán entre regiones y tipología de yacimientos, y que la cantidad de material fechable perdurable por tanto debe reflejar las tendencias subyacentes en la producción de material fechable (que refleja el número de personas vivas que produjeron tal material). Los modelos más complejos, que intentan evaluar los cambios paleoambientales, pueden proporcionar una visión más realista de la perduración del material fechable, pero existe un gran *corpus* de trabajos basados en el supuesto de que (a igualdad de condiciones) los patrones tafonómicos pueden tener un efecto neutral sobre el conjunto radiocarbónico resultante, o al menos, un efecto que puede simplemente corregirse. El método de corrección desarrollado por Surovell y colegas (Surovell et al. 2009; Bluhm y Surovell 2018) utiliza una base de datos de fechas ^{14}C geológicas asociadas a depósitos volcánicos (Bryson et al. 2006) como una medida de la distribución de la frecuencia de los sedimentos terrestres en varias edades y, suponiendo que la producción mundial de sedimento volcánico haya estado distribuida aleatoriamente a lo largo del Cuaternario, compara la distribución empírica con la distribución esperada, produciendo una estimación global del impacto de los factores tafonómicos en la perduración de los sedimentos terrestres de diferente cronología. Aunque los autores señalan que se trata de una aproximación global simplista y sugieren que el mejor enfoque sería desarrollar correcciones locales para cualquier estudio, sin embargo, su corrección global ha sido implementada por muchos investigadores e investigadoras (por ej., Downey et al. 2016; Edinborough et al. 2017; Kelly et al. 2013; Porčić et al. 2016; Williams 2013), simplemente por estar disponible. No obstante, esta corrección global puede o no puede ser apropiada en cualquier caso dado, y está claro que los factores tafonómicos tienen el potencial de sesgar los resultados (cf. Ballenger y Mabry 2011, donde se presenta un caso de estudio en que otros elementos superan a la producción como un factor determinante de la abundancia de material fechable y Davies et al. 2015, quienes muestran a través de modelos simulados que la tafonomía del paisaje puede producir aparentemente un registro ^{14}C complejo). Además, teniendo en cuenta algunos de los posibles agentes tafonómicos (por ej., el cambio del nivel del mar, la deposición y erosión aluvial) existen razones para pensar que los sesgos tafonómicos deberían ser heterogéneos tanto en el espacio como en el tiempo. Incluso, en una única región de estudio, una corrección tafonómica específica (que ningún estudio, que sepamos, ha intentado desarrollar) es probable que incluya áreas con diferentes historias paisajísticas, lo que resulta en una distribución espaciotemporal sobrerrepresentada o infrarrepresentada y, en consecuencia, una excesiva o baja representación de fechas radiocarbónicas y estimación de la población.

21.2.1.2.2. Intensidad de la investigación

Resulta tan desafiante el tema de la influencia de la intensidad diferencial de la investigación en la cantidad disponible de fechas para un intervalo temporal o región geográfica, que ya fue reconocido en los inicios del uso de la cantidad de fechas como estimación paleodemográfica (Rick 1987: 57-58). Un riesgo clave es que la disponibilidad presupuestaria puede dirigir la investigación: además de las preferencias eclécticas de investigación a nivel local, el número de muestras de ^{14}C realizadas en cualquier región puede reflejar la mejor suerte económica de esa región en los últimos 30 años, en lugar de la cantidad de población en la prehistoria. Incluso dentro de regiones de prosperidad económica comparable, las percepciones en torno a la importancia relativa de diferentes fenómenos arqueológicos o periodos y la relativa utilidad del carbono 14 y otros métodos de datación implican que los recursos pueden ser dirigidos de manera desigual hacia la datación de diferentes etapas. Otros factores de complicación son que los investigadores que recopilan las fechas radiométricas publicadas en la literatura académica pueden desconocer conjuntos de datos más grandes y quizás conjuntos de datos menos selectivos generados por la arqueología comercial, tal y como Combré y Robinson (2014) han observado recientemente, y que los resultados pueden estar estructurados por protocolos regionales (destacable, por ejemplo, el sobresaliente caso de Wyoming en relación a los datos de la base de datos CARD, cf. capítulo 22 de este libro) (ver Chaput et al. 2015: fig 1; Crema et al. 2017:2).

Las recientes técnicas estadísticas que utilizan la suma de distribuciones para resumir los conjuntos radiométricos intentan "corregir" la distribución agregada resultante con relación a la intensidad de investigación diferencial, ya sea agrupando las medias calibradas de yacimientos individuales o fases del yacimiento (e.g., Shennan and Edinborough 2007; Buchanan et al. 2008; Tallavaara et al. 2010), y sumando las fechas calibradas para sitios individuales o fases del yacimiento antes de sumar las distribuciones (e.g., Collard et al. 2010; Crema et al. 2016; Hinz et al. 2012; Shennan et al. 2013), o combinando fechas de yacimientos (e.g., Balsera et al. 2015) o áreas (e.g., Goldberg et al. 2016) antes del sumatorio de los intervalos de confianza probabilísticos de las dataciones seleccionadas. Chaput y colegas (2015) utilizan las distribuciones espaciales del conjunto en su totalidad como medida de la distribución espacial de la investigación, controlando (argumentan) la variabilidad en la intensidad de la muestra en el espacio.

Todas estas técnicas tienen como objetivo abordar la sobrerrepresentación de datos procedentes de las excavaciones bien financiadas, que producen docenas de fechas radiocarbónicas, pero van en contra de la suposición fundamental de que las poblaciones más grandes producirían más material fechable. La combinación da igual peso a cada sitio o fase del yacimiento, combinando así asentamientos de tamaño desigual: grandes y pequeños. Es decir, se espera que las poblaciones de diferentes dimensiones separadas por más de una cantidad mínima de tiempo produzcan diferentes cantidades de material fechable, pero las poblaciones de tamaño variable separadas en el espacio no lo son. La combinación de esta manera deja sin resolver la cuestión de cuándo la cantidad de fechas de un yacimiento, área o periodo cronológico en particular representa una anomalía en la atención prestada por la investigación en esa región/sitio/periodo, y cuándo esta hace referencia a una concentración de población. Kent Flannery describió el riesgo de no encontrar la metrópolis de Teotihuacan en el Valle de México debido a una estrategia de muestreo de superficie demasiado rígida (Flannery 1976: Cap. 5). Existe un riesgo análogo por estudios temporales: la combinación de varias fechas para

minimizar el sesgo resultante de unas cuantas investigaciones bien financiadas puede conducir a los analistas a ignorar unos sitios los cuales tienes muchos fechados *específicamente porque* son sitios grandes con poblaciones grandes.

También Crombé y Robinson (2014) discuten cómo los cambios en el patrón de asentamiento pueden dar lugar a fluctuaciones falseadas en la suma de las distribuciones con la intención de resumir los conjuntos de fechas radiométricas. Al igual que en el caso del sesgo tafonómico, este es un problema que la innovación metodológica en las aproximaciones de fechas como datos han preferido obviar (como ha sido señalado por los críticos de dicho enfoque al evaluar las respuestas a sus críticas [e.g., Hiscock y Attenbrow 2016; Torfing 2015b]). También como en la tafonomía, es especialmente peligroso en el sentido de que espaciotemporalmente es desigual, afectando a diferentes subconjuntos de grandes *corpus* radiocarbónicos de manera diferente a medida que los sesgos investigadores cambian tanto en el tiempo como en el espacio.

21.2.1.2.3. Tamaño de la muestra

Cualquier conjunto de datos radiométricos dado representa una muestra única, e inevitablemente pequeña, de una población desconocida de material fechable producida durante el momento cronológico investigado. La probabilidad de que esta muestra se aproxime a la población de la que se extrajo se relaciona directamente con el tamaño de esta, y en los casos de largos intervalos de tiempo más particularmente con la densidad de datos (es decir, el número de fechas ^{14}C por intervalo de tiempo dado). La población que se está muestreando es la cantidad total de material fechable originalmente producido; asumiendo una relación directa (y constante) entre ese material y el tamaño de la población humana, la complejidad demográfica a lo largo del tiempo puede considerarse como heterogeneidad poblacional, es decir, donde esta es más grande se requiere una muestra más amplia para caracterizar la población con precisión. En términos prácticos, eso significa que para que los patrones robustos (que podrían ser atribuidos a cambios poblacionales) se puedan discernir en los conjuntos de ^{14}C , las densidades de datos deben ser lo suficientemente altas como para revelar cambios poblacionales en la duración y magnitud postuladas. En general, si bien es posible detectar el simple caso de que las poblaciones tienden al incremento rápido con la adopción inicial de la agricultura, o aumentan con el tiempo, la demografía prehistórica con este nivel grosero ya es accesible a través de otros indicadores (es decir, tamaño de asentamientos y frecuencia). La detección de cambios de población más complejos, a corto plazo o sutiles debe superar los problemas de dispersión estadística inherentes a los datos radiométricos, incluso suponiendo que se hayan abordado los sesgos de intensidad de la investigación y los aspectos tafonómicos. Además, los vacíos en las secuencias ^{14}C deben ser una consecuencia esperada del muestreo y no deben indicar necesariamente un *hiatus* de ocupación (Rhode et al. 2014).

21.2.2. Resumen e interpretación de los conjuntos de fechas radiocarbónicas

Los problemas derivados de la estimación paleodemográfica a partir de dataciones ^{14}C acumuladas, como los discutidos anteriormente, han sido reconocidos durante décadas,

pero se han obviado más que resuelto. Con la excepción del trabajo de Surovell y otros sobre aproximaciones simples con relación a la corrección tafonómica (Surovell et al. 2009; Surovell y Brantingham 2007; Bluhm y Surovell 2018), la innovación metodológica (por ejemplo, Bevan et al. 2017; Bronk Ramsey 2017; Crema et al. 2017; Edinborough et al. 2017; Shennan et al. 2013; Timpson et al., 2014) se ha centrado en cómo recuperar patrones de conjuntos de ^{14}C y no en el grado en que esos conjuntos radiométricos reflejan la población original de material fechable generado en el pasado y depositado sucesivamente. Estos problemas relativamente más manejables de resumen e interpretación de fechas radiocarbónicas no son en absoluto sencillos, como pone de manifiesto la discusión en el capítulo 19 de este libro.

¿Cómo se pueden resumir las fechas ^{14}C , que incorporan tanto la incertidumbre de medición como la de calibración, para poder leerlas como un indicador de población? Es decir, ¿cómo se pueden ver las frecuencias de las fechas en el tiempo cuando cada una de estas se compone no de una medición puntual (un año), sino de una distribución desigual de su probabilidad en un lapso de tiempo?

El enfoque inicial de Rick no abordó los problemas introducidos por la calibración, pero sí que abordó los problemas de incertidumbre de la medición al seleccionar un ancho de intervalo suficientemente grande para abarcar la desviación estándar de la mayoría de las dataciones. Sin embargo, la simple combinación de fechas ^{14}C se complica por el hecho de que esas dataciones son mediciones que consisten en aproximaciones probabilísticas de años calendáricos, que incorporan tanto la incertidumbre de medición normalmente distribuida como los efectos de calibración irregularmente distribuidos. Un conjunto de fechas radiométricas no es un listado de años calendáricos, sino más bien una colección de densidades irregulares de probabilidades. La necesidad de tener en cuenta los efectos de la calibración en la producción de una distribución irregular de probabilidad para cada datación conllevó finalmente a un cambio general en la combinación de dataciones (o el valor medio de ellas) a la suma de ^{14}C , en que las fechas calibradas son combinadas añadiendo la altura de cada distribución calibrada a cada año del calendario, y normalizando la altura de las densidades de probabilidad acumulada resultante.

Aunque se siguen explorando algunas alternativas, por ejemplo, el *binning* (Weitzel y Codding 2016) y los rangos resumidos (Drake et al. 2016), abordar las distribuciones irregulares de probabilidad de las dataciones calibradas a partir del uso de la suma de las distribuciones de probabilidad se ha convertido en el método dominante para resumir los conjuntos de ^{14}C , a pesar de las diferentes críticas metodológicas y teóricas (e.g., Attenbrow y Hiscock 2015; Bamforth y Grund 2012; Bayliss et al. 2007; Chiverrell et al. 2011; Contreras y Meadows 2014; Culleton 2008; Mökkönen 2014; Torfing 2015a). Es probable que esto se deba en gran parte a la facilidad con la que puede calcularse, junto con la incapacidad de las críticas para sugerir una alternativa más viable. La reciente discusión de Bronk Ramsey sobre varios métodos para resumir las fechas radiocarbónicas (Bronk Ramsey 2017: 1810–1813) concluye que la estimación de la densidad kernel (KDE, siglas en inglés) proporciona la herramienta más prometedoras y la más efectiva para separar la señal (frecuencia de la datación) del ruido (principalmente efectos de la curva de calibración y muestreo). Sin embargo, debido a que el método KDE aún no ha sido adoptado por la comunidad científica y ya que existe una gran cantidad de publicaciones recientes que utilizan la suma de probabilidades, consideramos este enfoque en detalle.

21.2.2.1. El método de consenso: la suma de distribuciones de probabilidad de fechas calibradas.

En una calibración radiocarbónica individual, la máxima probabilidad –punto más alto de la curva de la función de densidad– en cualquier año calendárico corresponde a la probabilidad de que esta sea la fecha verdadera del material datado. Las calibraciones individuales de ^{14}C se pueden combinar agregando la altura de cada punto de la curva del intervalo calibrado en cada fecha del calendario y normalizar la altura de la función de la densidad de probabilidad acumulada resultante, que debe contener todas las muestras involucradas (figura 21.1). La suma de probabilidades puede ser fácilmente calculada utilizando diferente software de calibración: CalPal (Weninger et al. 2007), Calib (Stuvier y Reimer 1986) u OxCal (Bronk Ramsey 2009), o paquetes del software estadístico R como BChron (Parnell 2015).

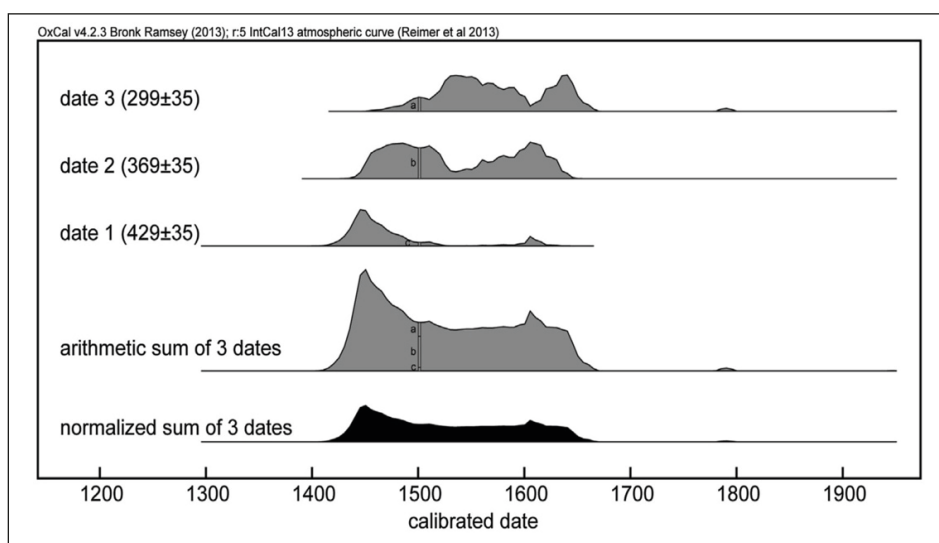


Figura 21.1. Suma de distribuciones normalizadas (negro), obtenidas con OxCal v4.2.3 (Bronk Ramsey, 2009a) agregando las alturas de las distribuciones calibradas para las dataciones 1-3 en cada fecha del calendario y normalizando la altura de la distribución de la suma aritmética en 1. Se marca una única fecha del calendario arbitrariamente para ilustrar la suma de las tres probabilidades distintas.

En publicaciones basadas en el “enfoque de la suma de probabilidades” (e.g., Buchanan et al. 2008; Collard et al. 2010; Crema et al. 2016; Gamble et al. 2005; Gayo et al. 2015; Gkiasta et al. 2002; Marsh 2015; Riede 2009; Shennan et al. 2013, Shennan y Edinborough 2007; Williams 2013; se incluyen más ejemplos en la tabla 21.1), los picos y valles en una distribución de probabilidades sumada se consideran proporcionales al número de muestras ^{14}C de la edad del calendario correspondiente (ver Williams 2012 para una revisión metodológica). En efecto, la suma de probabilidades se utiliza como un indicador para la distribución de frecuencia de fechas¹⁹⁵.

¹⁹⁵ Las sumas de probabilidades pueden proporcionar una aproximación razonable en relación con los cambios en la frecuencia temporal de la muestra fechada cuando un modelo bayesiano incluye una densidad suficientemente alta de datos simulados (Bronk Ramsey 2001).

La suma de las distribuciones de las dataciones no calibradas resulta, por tanto, engañosa, puesto que la relación entre el calendario y la edad ^{14}C no es del todo lineal: si las muestras de ^{14}C se distribuyen uniformemente a lo largo de la escala calendárica, habrá muchos más resultados ^{14}C correspondientes a la calibración de “mesetas” que a secciones más inclinadas de la curva de calibración (Reimer et al. 2013). A pesar de la sugerencia de Hinz et al. (2012), según el cual la curva de calibración de ^{14}C no produce picos y valles falsos en una suma de distribuciones con un número suficiente de dataciones calibradas, se puede demostrar fácilmente que las fechas radiométricas simuladas cuyos valores están espaciados uniformemente (por ejemplo, intervalos de 10 años) proporcionarán una suma de distribuciones cuyos picos/valles se corresponderán directamente con las ondulaciones de la curva de calibración (e.g., Chiverrell et al. 2011; Prates et al. 2013). El mismo patrón emerge cuando las fechas calendáricas de dataciones simuladas se generan aleatoriamente a partir de una distribución uniforme (e.g., Armit et al. 2013; Bleicher 2013).

La creación de la suma de probabilidades a partir de la función de densidad de probabilidades calibradas irregulares en lugar de la función de Gauss, que solo considera el error de medición, produce por lógica una suma de distribuciones más heterogénea (Bronk Ramsey 2017). A menos que un investigador pueda compensar tales irregularidades, lo que (pensamos) es imposible sin conocer la distribución subyacente de las dataciones de la muestra (el mismo patrón que un investigador espera reconstruir con la suma de probabilidades), la calibración crea un reto extra. Debido a que la variabilidad en la curva de calibración a lo largo del tiempo, las distorsiones de la distribución de la suma de probabilidades como consecuencia de la calibración dependerán del intervalo de tiempo considerado.

Los intentos de abordar este problema han incluido restar una suma de distribuciones “uniforme” generada artificialmente a la distribución del sumatorio de probabilidades observadas en el registro arqueológico radiocarbónico (e.g., Johnstone et al. 2006), o dividir una distribución por otra (Hoffmann et al. 2008), para separar de la calibración derivada los picos y valles de la señal “residual”, pero esto solo puede funcionar si la frecuencia temporal de las muestras de ^{14}C también fuera constante. De manera similar, restar a la suma de distribuciones generada artificialmente mediante el muestreo de una distribución exponencial de una distribución de suma de probabilidades ^{14}C solo podría corregir las fluctuaciones de calibración si las dataciones reales se distribuyeron exponencialmente. Esta cuestión presenta una gran actividad metodológica, donde los enfoques computacionales más sofisticados parecen tener un cierto éxito en la reducción del impacto de las fluctuaciones de calibración, principalmente comprobando las distribuciones observadas con modelos nulos (por ejemplo, Crema et al. 2016; Edinborough et al. 2017; Kerr and McCormick 2014; Shennan et al. 2013; Timpson et al. 2014). Bronk-Ramsey (2017) argumenta que estos esfuerzos metodológicos están mal dirigidos, lo que sugiere que las sumas de ^{14}C son inherentemente poco adecuadas para separar la señal de la población del ruido de calibración y que los enfoques de densidad por kernel (KDE) proporcionan mejores herramientas para resumir los conjuntos de dataciones.

En el ejemplo poco apreciado de Bronk Ramsey (2001, figura 3), la distribución de la suma es la suma de las estimaciones de densidad posterior de la muestra radiométrica simulada, no de calibraciones simples. El hecho de que los límites de fase estén exactamente en los puntos correctos (dónde cambia la frecuencia de las fechas) ayuda a producir una suma de distribuciones que se asemeja mucho a la distribución original de las dataciones de la muestra, que están espaciadas uniformemente dentro de cada fase.

Tabla 21.1. Comparación de densidad de datos ¹⁴C en una selección de publicaciones que usan series de ¹⁴C como estimadores de población

Study	Number of dates	Approximate period (total calendar years)	Approximate spatial scale (km ²)	Approximate data density (dates/year) [Effective density]	Approximate overall data density (dates/km ²) [Effective density]
Armit et al., 2013	1554	1200 cal BC - cal AD 400 (1600)	Ireland (84420)	1.0	0.02
Balsera et al. 2015	4402 (2448 after pooling)	8000 - 3000 BP (5000)	Iberia (582000)	0.88 [0.49]	0.008 [0.004]
Bevan et al., 2017	30516 (16506 bins)	8000 cal BC - 1500 cal AD (9500)	Britain and Ireland (subset into four regions) (293750)	3.21 [1.73]	0.1 [0.06]
Buchanan et al., 2008	1509 (628 dates or pooled means)	15 - 9 ka cal BP (6000)	N. America excluding Mexico (18065460)	0.25 [0.1]	0.00008 [0.00003]
Chaput et al., 2015	29609	13000 - 500 BP (12500)	N. America excluding Mexico (18065460)	2.37	0.002
Collard et al., 2010	4246/2129 (sum of 1762 sums)	8000 - 4000 cal BC (4000)	Britain (209330)	1.1/0.5 [0.4]	0.02 [0.01]
Contreras and Meadows, 2014	200, 1000, 2000 simulated	cal AD 1000 - 1700 / 1000 - 1800 (700 / 800)	Europe / Mexico	0.29, 1.43, 2.86 / 0.25, 1.25, 2.50	
Crema et al., 2016	1433 (87, 128, and 186 bins)	7000 - 3000 BP (4000)	portions of eastern Japan (125520)	0.36 [0.02 - 0.05]	0.01
Crombé and Robinson, 2014	571	9000 - 2000 cal BC (7000)	Belgium (30530)	0.08	0.02
Drake et al. 2016	1165 (1007 after pooling)	8300 - 5000 BP (3300)	Iberia (subset in various ways) (582000)	0.35 [0.31]	0.002 [0.002]
Edinborough et al., 2017	336 (212 bins) / 187 (127 bins)	10000 - 0 BP (10000)	Prince Rupert Harbor, B.C., Canada (300)	0.03 [0.02] / 0.02 [0.01]	1.12
Gamble et al., 2005	2255, split into subsets of 28-500	"25 - 11 ka calibrated radiocarbon years" (14,000)	western Europe (1405230)	0.15 [0.002 - 0.04]	0.002
García Puchol et al., 2017	834	8000 - 5500 BP (2500)	Iberia (subset in various ways) (582000)	0.33	0.002

Aunque algunos autores reconocen que las sumas de probabilidades son, en el mejor de los casos, indicadores imperfectos, otros pueden dar la impresión a los lectores de que las características principales de la suma de probabilidades reflejan las fluctuaciones de la población de forma bastante precisa, tanto en términos de escala como de tiempo. Por ejemplo, Collard et al. (2010) y Kelly et al. (2013) utilizan las pendientes de picos de la suma de probabilidades para calcular las tasas de crecimiento de la población, mientras que Shennan et al. (2013: 4) declaran que *“en prácticamente todas las regiones examinadas aquí, hay fluctuaciones demográficas significativas y en la mayoría hay indicaciones en ciertos puntos de disminución de la población del orden del 30-60%”*. Endinborough y colegas (2017) sostienen que las tendencias demográficas del orden de 100 a 200 años son detectables.

La presunta relación entre la altura de la curva de la suma de probabilidades y el tamaño de la población humana correspondiente ha sido criticada por razones metodológicas y teóricas (véase, Attenbrow y Hiscock 2015; Bamforth y Grund 2012; Bayliss et al. 2007; Contreras y Meadows 2014; Culleton 2008; Torfing 2015a), ya que otros factores además del tamaño de la población pueden crear picos y caídas en la curva de la suma de probabilidades (analizadas en detalle en el apartado 21.2.2.2). Sin embargo, Shennan (2013: 305) sostiene que *“[el] punto clave es que, aunque una sola fecha pueda tener un amplio rango calibrado, la acumulación de las distribuciones de probabilidad de un gran número de fechas produce un alto grado de resolución cronológica haciendo posible rastrear las fluctuaciones demográficas con un detalle considerable”*. El atractivo por los indicadores paleodemográficos es tal que las publicaciones donde se utiliza una aproximación basada en la suma de probabilidades continúan apareciendo a un ritmo constante, como lo indica la trayectoria de citas para el documento de Rick de 1987 descrito en la introducción del presente capítulo (los ejemplos más recientes incluyen varios periodos de tiempo y abarcan África, Australia, Norteamérica, América del Sur, Asia oriental y diferentes regiones europeas (ver en lugar las referencias en el capítulo 20 de este libro, así como Armit et al. 2013; Chaput et al. 2015; Crema et al. 2016; Drake et al. 2016; García Puchol et al. 2017; Gayo et al. 2015 ; Hinz et al. 2012; Manning y Timpson 2014; Mulrooney 2013; Oh et al. 2017; Porčić et al. 2016; Wang et al. 2014; Williams 2013); véase la tabla 21.1 para obtener más información).

21.2.2.2. Desafíos en la interpretación de la suma de distribuciones de probabilidad de fechas calibradas.

En un trabajo previo (Contreras y Meadows 2014) utilizamos datos simulados para examinar los efectos del muestreo aleatorio, la densidad de datos, la incertidumbre de la medición y la curva de calibración en las sumas de ^{14}C . En dicho estudio examinamos si era posible recuperar, utilizando la suma de probabilidades de fechas simuladas que abarcan dos eventos demográficos conocidos (la Muerte Negra [o *Peste Negra*] en Europa y el colapso catastrófico posterior a la conquista del centro de México), poblaciones de las que se extrajeron las fechas simuladas. Como el enfoque de la suma de probabilidades sigue siendo abrumadoramente el método más común para resumir los conjuntos radiométricos, los problemas que planteamos siguen siendo pertinentes. Los retos discutidos en dicho trabajo, y que resumimos aquí, son específicos de la suma de probabilidades, pero también requerirán consideraciones en otros métodos. Incluso si se pudiese abordar las cuestiones tafonómicas y de sesgo de investigación, los problemas

relacionados con el muestreo, la densidad de datos y la incertidumbre de la medición seguirán siendo la lucha para cualquier método de resumen de un conjunto radiométrico.

21.2.2.2.1. Muestreo aleatorio

Incluso sin los picos y caídas en las sumas de probabilidades que se derivan de la calibración de las fechas ^{14}C (en otros términos, utilizando la función *C_Simulate* de OxCal para generar una dispersión de medición realista, en lugar de *R_Simulate*; ver el capítulo 14 de este libro), en nuestro trabajo de simulación encontramos que con 200 muestras radiométricas durante 700 años (una densidad de dataciones mucho mejor que la mayoría de los ejemplos publicados, ver la tabla 21.1) los efectos del muestreo son sorprendentes (Contreras y Meadows 2014: figura 6). No solo la salida de estas curvas a partir de la distribución de la población de las que derivan es evidente; la variabilidad entre las muestras también es destacable: las fluctuaciones más destacadas en cada curva no son visibles en la mayoría de las demás. Una suma de probabilidad de radiocarbono correspondería solo a una de estas curvas, y aunque las submuestras podrían compararse entre sí, la distribución de la población subyacente sigue siendo fundamentalmente desconocida. Como resultado, es difícil, si no imposible, determinar cuáles de estas fluctuaciones efectivamente existen, y si representan cambios demográficos reales. La reducción de la variabilidad entre muestras fue evidente con el incremento de la densidad de datos (Contreras y Meadows 2014, figura 7), pero prevalece el caso en que el muestreo puede tener efectos notables en la estructura de las curvas de las sumas, y los caprichos del muestreo solo pueden ser suficientes para introducir o difuminar patrones que puedan ser tomados como indicadores demográficos. Este es el caso tanto en términos prácticos como teóricos.

En términos prácticos, esto introduce el problema de que en la mayoría de supuestos el patrón resultante de la suma de ^{14}C es resultado de otros factores más que de fluctuaciones demográficas en el pasado. Como discutimos a continuación, el muestreo aleatorio puede producir tanto falsos negativos (fallos en la detección de fluctuaciones poblacionales reales) como falsos positivos (características en la suma de probabilidades que pueden confundirse con los resultados de cambio demográfico pero que en realidad se relacionan con otros factores).

En términos teóricos, como cualquier conjunto de datos radiométricos comprende una sola muestra (de muchas teóricamente posibles) el muestreo introduce un grado de incertidumbre en cualquier afirmación de la relación dada entre un conjunto de ^{14}C y la población de material fechable de la cual se extrajo. La población original (de material fechable) de la que estas muestras fueron extraídas permanece esencialmente desconocida y, por lo tanto, hay cierta incertidumbre inherente a cualquier vinculación entre la cantidad acumulada de datos radiométricos y el tamaño de la población de material fechable originalmente producido (ver Drennan 2009: 93:95). Esta imprecisión es independiente a cualquier relación putativa entre la producción de material fechable y la demografía.

21.2.2.2.2. Densidad de datos

Las muestras más grandes se acercan más a la población original de la que provienen, principio fundamental de la estadística. Sin embargo, los enfoques prescriptivos para

el tamaño de la muestra necesario (por ejemplo, Hinz et al. 2012; Timpson et al. 2014; Williams 2012, citados como justificaciones por Wang et al. 2014; Whitehouse et al. 2014; entre otros) deben tener en cuenta el lapso de tiempo y la magnitud y duración de las fluctuaciones o tendencias que el estudio busca detectar. Las afirmaciones sobre los números necesarios y/o suficientes de dataciones ^{14}C deben ser aseveraciones sobre la densidad de datos, en lugar de número de muestras.

Hemos explorado los efectos de la densidad de datos con dos ejemplos simulados (véase Contreras y Meadows 2014: figuras 8 y 9). En ausencia de otros factores de confusión (es decir, la propia calibración ^{14}C), cuanto mayor es la muestra (es decir, mayor densidad temporal de datos) mejor resulta el ajuste de las curvas de la suma de probabilidades de datación a la curva de fluctuación de la población a lo largo del tiempo. Sin embargo, incluso en el caso de una muestra con una densidad muy elevada (2000 dataciones para un lapso temporal de 700 años, 2/3 fecha por año), la fluctuación demográfica que supuso la Muerte Negra en Europa está lejos de resultar observable en la curva agregada. No obstante, Edinborough y colegas (2017) argumentan que, con métodos estadísticos de filtrado de datos mejorados, los efectos demográficos de la Muerte Negra son, de hecho, detectables, al menos en densidades de 1000 fechas en 700 años. Con menor fiabilidad, si la densidad de la muestra es de 500 fechas en 700 años. En su estudio utilizan fechas ~ 250 [marinas] y ~ 140 [terrestres] durante un lapso temporal de ~ 3500 años. Nuestras simulaciones oscilaron entre una densidad promedio de 0,25/año (dicho de otro modo, una fecha cada 4 años) a 2,86/año; la mayoría de los ejemplos publicados utilizan conjuntos de datos arqueológicos entre 0,25 - 3,2 fechas por año, aunque los datos a menudo se agrupan o dividen geográficamente, lo que da densidades de datos reales de 0,1 fechas por año, o incluso menos (consúltese la tabla 21.1).

A medida que los investigadores han comenzado a lidiar, cada vez más, con la distribución diferencial de las poblaciones en el espacio y tiempo (e.g., Chaput et al. 2015; Crema et al. 2017; García Puchol et al. 2017; Manning y Timpson 2014), la densidad espacial de las muestras ^{14}C en regiones de estudio de tamaño muy variable también se convierte en un problema. En análisis donde la escala alcanza lo continental, las densidades de fechas en el espacio pueden ser bastante bajas, incluso sin considerar las densidades específicas para cualquier lapso temporal dado. Para los estudios que resumimos en la tabla 21.1, por ejemplo, las densidades espaciales varían desde 0,00003 a 1,9 fechas por km^2 con una mediana de 0,003 dataciones por km^2 es decir, una fecha por 332 km^2 , aunque, por supuesto, esas densidades subsumen distribuciones espacialmente desiguales. Incluso aquellos estudios que consideran la dimensión espacial, sin embargo, aún no han resuelto la probable heterogeneidad espacial provocada, por diferencias tafonómicas y el sesgo de investigación. Chaput et al. (2015) han intentado tratar la intensidad espacial variable de la investigación (ver discusión en Crema et al. 2017: 1 y 2).

21.2.2.2.3. Precisión de medida

La precisión de las dataciones ^{14}C que sirve como base para la suma de probabilidades es también un parámetro vital. La base general para nuestro estudio fue el mejor escenario realista con errores en la desviación de las dataciones en un rango de 20 a 40 años, que es la desviación típica en las dataciones radiométricas más recientes (Holoceno medio o final). Si, como es probable que incorporen conjuntos de datos más antiguos, los errores se simulan entre un rango de 20 a 100 años, las sumas de probabilidades resultantes

son más suavizadas y la probabilidad de documentar fluctuaciones es menor (Contreras y Meadows 2014: figura 10; ver también Bronk Ramsey 2017; Culleton 2008). La inclusión de datos legados, una característica de casi todos los ejemplos publicados, aumenta la densidad de dataciones, pero desdibuja cualquier particularidad¹⁹⁶.

Una complicación adicional con respecto a la vinculación de la suma de probabilidades de ¹⁴C con las poblaciones del pasado es la posibilidad de algún desajuste entre la edad de la muestra y la fecha del evento analizado. Problemas como las compensaciones de la edad de la muestra de madera, la duración de una ocupación, etc., reducen la efectividad de la precisión de la medida, pero no con un carácter simétrico (ver discusión en Bronk Ramsey 2009b). Incluso cuando las bases de datos se someten a una limpieza cronométrica rigurosa (Spriggs 1989), es probable que con frecuencia nos encontremos con una sobrerrepresentación, lo que de nuevo se relaciona con el problema endémico del uso de datos heredados, procedentes de la bibliografía históricas o de bases de datos. Esto puede contribuir a la disipación de señales (por ej., introduciendo una dispersión de fechas en un evento singular, dando la apariencia de una duración larga y/o haciendo que la señal sea aún más difícil de detectar en la suma de probabilidades), o la atribución de una fecha precisa pero inadecuada para un evento en particular; dicho de otro modo, en el caso de que el evento esté representado solo por dataciones radiométricas que no coincidan correctamente con el evento. En la medida en que la correspondencia entre eventos particulares y material fechado puede ser fundamental para la interpretación de los sumatorios de probabilidad (por ejemplo, si la investigación intenta relacionar una fluctuación de la población con un evento externo como un cambio ambiental destacable [véase Bevan et al. 2017; Buchannan et al. 2008; Kelly et al. 2013; Manning y Timpson 2014; Riede 2008; Roberts et al. 2017; Shennan et al. 2013; Wang et al. 2014; Williams et al. 2015]), la erosión de la precisión por cualquiera de estos factores presenta un obstáculo que debe ser gestionado.

21.2.2.2.4. Reconociendo patrones y separando patrones de ruido

La interacción de estos parámetros puede introducir varios problemas, incluso si ignoramos los diversos factores que pueden atenuar o complicar la vinculación entre la población y la producción/perduración/análisis del material fechable. Problemas específicos como aquellos que hemos ilustrado con datos simulados (Contreras y Meadows 2014) y que describimos a continuación no son necesariamente inevitables, pero pueden ocurrir, parcialmente como el resultado del sesgo del muestreo, incluso en una alta densidad de dataciones. Su posible aparición, en el caso de muestras arqueológicas inevitablemente singulares y no replicables, remarca la dificultad de interpretación de la suma de probabilidades. Identificando la ocurrencia de estos problemas, en ausencia de una curva demográfica con la que se puedan comparar los sumatorios, sigue siendo un reto importante.

El problema fundamental de todo es el de reconocimiento, es decir, ¿qué constituye un suceso demográfico y qué no? La falta de una definición consensuada sobre la du-

¹⁹⁶ Los algoritmos de suavizado (por ejemplo, la media continua [Rolling mean]) también se han aplicado a la suma de probabilidad en algunos estudios (por ejemplo, Hinz et al. 2012; Kelly et al. 2013), aunque existe un claro riesgo de borrar las señales reales, así como el ruido de fondo, y a pesar de la inexistencia de cualquier justificación estadística específica.

ración y magnitud de cada fluctuación demográfica “significativa” o de un movimiento “significativo” en la suma de probabilidades sigue siendo un obstáculo conceptual considerable. Además de la calidad y el carácter de nuestros datos, debemos preocuparnos por la magnitud de la fluctuación, la duración de la misma e incluso el momento en cual se produce. La distinción entre falsos/reales positivos y negativos, en la ausencia de una curva de referencia, se convierte en el mayor reto de los enfoques basados en el sumatorio de probabilidad. Una desviación de una distribución ideal o de una tendencia subyacente modelada (e.g., Shennan et al. 2013) es una posibilidad cuyo uso se ha explorado, pero estos enfoques aún se enfrentan a desafíos importantes. Estos incluyen los problemas metodológicos para distinguir la duración y la magnitud de las desviaciones, el problema interpretativo de determinar si el cambio de población o algún factor de confusión ha(n) producido la(s) fluctuación(es), así como el problema teórico de qué curva de referencia ideal utilizar. La comparación de varios modelos nulos (por ejemplo, curvas exponenciales, logísticas o uniformes, o aquellas procedentes de modelos demográficos derivados independientemente [cf. Balsera et al. 2015; Crema et al. 2016; Edinborough et al. 2017; Shennan et al. 2013; Timpson et al. 2014]) o la comparación de diferentes subconjuntos de datos (e.g., Crema et al. 2016; Drake et al. 2016) están siendo exploradas como herramientas para separar la señal dejada por las fluctuaciones demográficas del ruido producido por otras variaciones a lo largo de la escala temporal. Tales enfoques obviamente dependen de su poder explicativo en relación con el modelo nulo seleccionado. Crema y colegas (2017) abordan esta cuestión buscando las fluctuaciones locales en relación con las tendencias regionales. Con ese fin utilizan medias regionales como modelo nulo, haciendo eco del énfasis que ponía Rick (1987) en la comparación relativa.

El ruido de confusión puede consistir en:

- *Falsos positivos y negativos.* Un riesgo es que los picos o caídas que resultan de un muestreo aleatorio o de la curva de calibración puedan interpretarse como reflejo de los cambios en la población. Lo contrario también es un riesgo: la eliminación de patrones reales por el muestreo aleatorio, calibración o dispersión de la medición. En tales casos, la distribución de la suma de probabilidades no muestra el punto máximo o mínimo donde debería existir, según la población subyacente, o al menos, la desviación de la uniformidad no es más destacable que otras desviaciones que son claramente falsas.
- *Tiempo [timing].* Incluso en los casos en que un pico o caída se puede detectar con precisión y confianza, puede ser difícil establecer una fecha correcta o estimar su duración. Esto puede potencialmente crear falsos sincronismos con eventos de cronología precisa, o enmascarar simultaneidades reales.
- *Escala.* La amplitud de un pico o caída depende de muchos factores, y es difícil (pero sí posible) traducir esto en un cambio proporcional en la población, y establecer umbrales apropiados para su significación. Además, tenemos que definir modelos de lo que se considera una crisis demográfica, o un auge de población para poder especificar los picos que estamos buscando.
- *Efecto limítrofe.* Uno de los efectos de la medición de la incertidumbre y la calibración es extender las fechas radiométricas calibradas más allá del intervalo de tiempo desde el cual se tomaron las muestras. Estos efectos de límite son evidentes en las muestras ^{14}C simuladas de una población conocida, pero presenta un problema cuando se trata de una población arqueológica desconocida, que es en sí misma objeto de investigación, particularmente si los investigadores es-

tán interesados en el inicio o final de la ocupación de un sitio determinado (e.g., Rieth et al. 2011; Mulrooney 2013) y, por consecuencia, no pueden ignorar las colas de la distribución de los sumatorios de probabilidad. En otras situaciones, los autores han sumado deliberadamente las fechas de un periodo de tiempo más amplio que el de análisis, a fin de minimizar los efectos de borde (e.g., Edinborough et al. 2017; Porčić et al. 2016; Shennan et al. 2013).

21.3. Discusión

Las simulaciones de fluctuaciones demográficas históricamente comprobadas con repercusiones socio-políticas destacables que hemos explorado (Contreras y Meadows 2014) muestran que, incluso bajo condiciones ideales, es complicado diferenciar entre patrones de fluctuación demográfica reales y falsificados. También resulta difícil fechar con precisión fluctuaciones bruscas, incluso con densidades de datos mucho más elevadas que en la mayoría de los intentos usualmente publicados. Tanto los defensores como los críticos del enfoque “sumatorio” pueden esperar que los estudios de simulación produzcan criterios de referencia en relación con la densidad de datos y la duración y magnitud de las fluctuaciones de la población postulada. Dicho de otro modo, un medio simple para estimar si una fluctuación es de una duración y magnitud suficiente para sugerir que es reflejo de patrones de población real en lugar de ruido estadístico. Por desgracia, la abundancia de variables (por ej., tamaño de población, duración y magnitud del evento, magnitud de la desviación típica, rango de calibración de la datación) y la distribución irregular de la población que estamos intentando reconstruir (de interés precisamente porque son irregulares, es decir, picos o caídas de población) significa que cualquier regla de este tipo probablemente represente una ilusión, en lugar de una evaluación rigurosa. Adicionalmente, los juicios sobre la “realidad” de los patrones detectados en los sumatorios de datos radiométricos enfrentan el mismo desafío identificado por Cowgill (1977) con respecto a las pruebas estadísticas de significación en arqueología: su uso adecuado en arqueología requiere juicio y argumento en lugar de una aceptación o rechazo sin discusión.

Es posible que la suma de un gran número de muestras verdaderamente aleatorias, cuya frecuencia de producción, preservación y análisis sea proporcional al tamaño de la población (y no afectada por factores como la desigual preservación, variabilidad en la intensidad de investigación, etc.), genere sumatorios de probabilidad que se confundan con fluctuaciones en la población prehistórica. Nuestro análisis con datos artificiales (Contreras y Meadows 2014) se observó que se requiere una extrema precaución para evitar una falsa identificación. Esto nos llevó a argumentar que las aproximaciones de suma de probabilidades pueden tanto oscurecer como revelar, cambios rápidos en la población. No obstante, la sospecha de falsos positivos y eventos inoportunos podría sugerir relaciones engañosas entre la población y acontecimientos exógenos.

Algunos de los casos de estudio citados en la tabla 21.1 abarcan milenios y áreas geográficas del tamaño de continentes (e.g., Chaput et al. 2015; Goldberg et al. 2016; Peros et al. 2010; Wang et al. 2013; Williams et al. 2015), y otros están centrados en episodios más cortos y/o locales en los que se percibe que la población cambia rápidamente (e.g., Edinborough et al. 2017; Kelly et al. 2013; Shennan et al. 2013). Los intervalos de tiempo investigados varían desde 600 a 42 000 años y áreas desde 150 km² a 18 millones km². Aquellos que se centran en escalas de tiempo más cortas ofrecen los tipos de

resolución temporal necesarios para abordar numerosas preguntas de investigación arqueológica y antropológica. Como indica Shennan, *“dada la velocidad con la que operan los procesos demográficos... es probable que tengan una escala de tiempo característica de 100 años más que de 1000 años (...), así que la resolución temporal de nuestros métodos debe ser capaz de lidiar con esto. En la actualidad hay una gran variedad de métodos para detectar la existencia de fluctuaciones de población a la escala de resolución requerida, no menos importante el uso de los sumatorios de probabilidad radiocarbónica como un indicador demográfico”* (Shennan 2013: 302).

Sin embargo, las tendencias a largo plazo también han sido de interés, y en algunos de los mismos problemas emergen en los estudios de trayectorias demográficas en escalas de tiempo de muchos milenios. Estos casos también deben enfrentarse a los efectos de la dispersión de la medida, calibración, tafonomía, y el sesgo investigador; nuestro estudio de simulación (Contreras y Meadows 2014) sugiere que mientras los dos primeros pueden ser menos significativos en obscurecer las tendencias a largo plazo en los sumatorios de probabilidades, los efectos derivados de la preservación diferencial y la intensidad de la investigación pueden ser más destacables.

Las mejoras metodológicas se han centrado en el problema de separar la señal del ruido, con la notable excepción de la reciente revisión de Bronk Ramsey (2017), que argumenta que el KDE es mejor herramienta que las sumas de probabilidades para resumir los conjuntos de ^{14}C . Como Bronk Ramsey indica, sin embargo, el uso del KDE no puede obviar la necesidad de una argumentación cautelosa en torno a si los conjuntos radiométricos de hecho reflejan patrones de población. Lo mismo ocurre con los refinamientos metodológicos de las aproximaciones basadas en sumatorios (recientemente, e.g., Crema et al. 2016; 2017; Edinborough et al. 2017; Shennan et al. 2013; Timpson et al. 2014). Por ejemplo, comparaciones con modelos nulos hacen los mismos supuestos sobre la producción y perduración de los conjuntos radiométricos, y no abordan los problemas de los posibles efectos heterogéneos de los sesgos tafonómicos o de investigación. El énfasis de Crema y colegas (Crema et al. 2017; 2016) en comparar subconjuntos de datos ^{14}C entre sí parece el más prometedor en este sentido.

El enfoque de prueba formal de hipótesis estadísticas, desarrollado y refinado por Shennan y colegas (por ejemplo, Crema et al. 2016; Edinborough et al. 2017; Shennan et al. 2013; Timpson et al. 2014), se centra en la comprobación de hipótesis nulas que postulan que los patrones observados en la suma de intervalos de probabilidad podrían ser producidos por muestreo aleatorio en una población simple, representando, por ejemplo, una población constante o un crecimiento exponencial. El problema principal se convierte en identificar qué desviaciones de la curva de la hipótesis nula son estadísticamente significativas. Los resultados son afirmaciones sobre las fluctuaciones de la población del pasado (Shennan et al. 2013: tabla 1) que podrían ser comprobados utilizando aproximaciones simuladas, como la que se ha presentado anteriormente o la presentada por Edinborough et al. (2017). Si bien esto da como resultado una capacidad más robusta para identificar las fluctuaciones significativas en los sumatorios de probabilidad, obviamente depende fundamentalmente de la selección del modelo nulo usado en la contrastación estadística. Crema y colegas (2016; 2017) y Bevan y colegas (2017) han enfocado el problema de la selección de la hipótesis nula, centrándose en la desviación del subconjunto ^{14}C del partir del patrón del conjunto radiométrico, así como entre los desajustes entre subconjuntos. Esta aproximación puede ser igualmente aplicable a las curvas KDE que Bronk Ramsey (2017) defiende que resultan menos problemáticas que los sumatorios de probabilidad. Esto puede servir para

desviar el enfoque de cuestiones de identificación de la señal y el ruido y hacia los problemas subyacentes de la producción, preservación y recuperación del material fechable, y la heterogeneidad espaciotemporal de estos procesos.

Los diversos factores de confusión que hemos descrito aquí sugieren que, si bien en ocasiones el tamaño de un conjunto de dataciones ^{14}C puede reflejar el tamaño de una población pasada, no debemos convertirlo en un principio general siempre válido. No es una justificación suficiente citar otros casos de estudio cuyos autores han realizado similares afirmaciones. Lo que debemos comprobar, como se ha discutido en los apartados 21.2.1.2.1 y 21.2.1.2.2, es si un conjunto radiométrico refleja las variaciones potenciales de la población en el espacio y tiempo y cómo lo hace. En el caso de fluctuaciones a corto plazo o tendencias a largo plazo, recomendaríamos que cuando se utilicen conjuntos ^{14}C como estimadores del tamaño de una población antigua, incumbe a los investigadores argumentar: 1) qué relación existe entre la producción, en el pasado de material orgánico fechable, su preservación y cómo han podido alterar la supuesta relación con el tamaño original de la población; 2) que los datos radiométricos que se utilizan pueden ser considerados como una muestra aleatoria; 3) que, si los objetivos son sucesos y no tendencias, los sucesos que se intentan identificar tuvieron una duración temporal suficiente, en relación con las incertidumbres del promedio de medición, que será posible identificarlos, y que el tamaño de la muestra es lo suficientemente grande, y proporcional al intervalo temporal en estudio; y 4) que los patrones observados no son el resultado de la curva de calibración.

Como hemos señalado en anterioridad, el desarrollo metodológico se ha centrado en la tercera y cuarta de estas cuestiones. Además, las respuestas a las críticas (por ejemplo, Timpson et al. 2015; Williams y Ulm 2016) han recurrido a la defensa de que la investigación paleodemográfica en base a dataciones agregadas (*dates as data*) representa un comienzo necesario, y que en el curso natural del proceso científico los resultados serán mejorados. No obstante, el enfoque de la investigación en los conjuntos ^{14}C ha sido generalmente en las interpretaciones últimas (correlaciones de eventos demográficos con cambios climáticos, consecuencias demográficas en relación con los cambios en la subsistencia y/o complejidad sociopolítica, etc.) más que en las cuestiones fundamentales para establecer la idoneidad del enfoque. Separar la señal del ruido es, sin lugar a dudas, un problema metodológico vital (especialmente al sumar los sumatorios de probabilidad; puede ser menos importante en las aproximaciones KDE), pero determinar cuán sensibles son las interpretaciones a los problemas de calidad, densidad o cobertura de los datos subyacentes es igualmente importante, aunque menos abordado. El problema fundamental, que consideramos como un foco valioso para la investigación, sigue siendo comprender qué está causando la señal que expresa la curva agregada de la función de densidad de probabilidad.

A la luz de las suposiciones necesarias y los retos interpretativos que describimos aquí, parece claro que estos métodos deberían utilizarse con cautela. La probable variabilidad espaciotemporal de los sesgos tafonómicos y de investigación sugiere, creemos, que estos problemas deberán ser abordados en particular más que en abstracto. Dicho de otro modo, para cualquier conjunto de fechas radiométricas dado se debe articular una argumentación específica en cuanto a: 1) ¿Por qué los sesgos tafonómicos y de investigación no juegan un papel importante en la estructuración de los datos radiocarbónicos? 2) ¿Por qué estos sesgos deberían considerarse comparables en cualquier región o lapso de tiempo analizado? O bien, ¿cómo se contabiliza su variabilidad en el caso que no lo sean? 3) ¿Cuándo la densidad de datos de fechas ^{14}C es suficiente para identificar

los diferentes patrones demográficos que el estudio examina? Solo cuando esos argumentos se hayan formulado podemos pasar a la suficiencia del método para resumir los conjuntos radiométricos y la interpretación de los patrones discernidos.

Además, podríamos argumentar que la innovación KDE de Bronk Ramsey no resuelve la mayoría de estos problemas, por lo que lo que se necesita son propuestas para lidiar metodológicamente con las incertidumbres subyacentes, incluso si se multiplica la cantidad de trabajo necesario para poner en crecimiento las bases de datos radiocarbónicas para uso analítico (véase capítulo 22 de este libro. Cf. también Chaput y Gajewski 2015 para una revisión de las bases de datos disponibles). Su uso puede resultar engañoso, porque la calidad de datos subyacentes y los problemas de sensibilidad del modelo no suelen tenerse en consideración a la hora de seleccionar y anotar los datos. Eso no quiere decir que estas bases de datos no deben ser utilizadas; al contrario, su compilación y análisis es el tipo de investigación sintética de “*big data*” que se ha considerado como una contribución importante en la que la arqueología puede hacer preguntas de gran alcance sobre el pasado humano que también son relevantes para el mundo contemporáneo (Kintigh et al. al. 2014).

Además, la paleodemografía no tiene por qué ser el único objetivo analítico. Como señalan Attenbrow y Hiscock (2015: 34), los conjuntos radiocarbónicos pueden y deben usarse para investigar otras cuestiones más allá de las demográficas. Por ejemplo, podrían arrojar luz sobre el momento y la variación regional de los cambios en las prácticas de subsistencia, las colonizaciones/migraciones, o los procesos tafonómicos. Los factores de complicación para una reconstrucción demográfica amplia pueden ser temas de interés por sí mismos, y su investigación puede elucidar sobre la mejor forma de utilizar los conjuntos de ^{14}C subyacentes para los propósitos de la reconstrucción demográfica.

Agradecimientos

Gracias a Juan A. Barceló por su invitación para realizar este capítulo, y a la colaboración de Salvador Pardo-Gordó en la traducción del inglés. Las discusiones que fueron la base de nuestro artículo de 2014 en la revista *Journal of Archaeological Science*, y predecesor del presente trabajo, fueron posibles gracias a la beca Humboldt que el primer autor disfrutó en la Universidad de Kiel.

BIBLIOGRAFÍA

- Armit, Ian, Graeme T Swindles, y Katharina Becker (2013). From dates to demography in later prehistoric Ireland? Experimental approaches to the meta-analysis of large ^{14}C data-sets. *Journal of Archaeological Science* 40(1): 433–438.
- Attenbrow, Val, y Peter Hiscock (2015). Dates and demography: are radiometric dates a robust proxy for long-term prehistoric demographic change? *Archaeology in Oceania* 50(2): 29–35.
- Ballenger, Jesse A M, y Jonathan B Mabry (2011). Temporal frequency distributions of alluvium in the American Southwest: taphonomic, paleohydraulic, and demographic implications. *Journal of Archaeological Science* 38(6): 1314–1325.
- Balsera, Verónica, Pedro Díaz-del-Río, Antonio Gilman, Antonio Uriarte, y Juan M Vicent (2015). Approaching the demography of late prehistoric Iberia through summed ca-

- librated date probability distributions (7000 - 2000 cal BC). *Quaternary International* 386(C): 208–211.
- Bamforth, Douglas B, y Brigid Grund (2012). Radiocarbon calibration curves, summed probability distributions, and early Paleoindian population trends in North America. *Journal of Archaeological Science* 39(6): 1768–1774.
- Bayliss, Alex, Christopher Bronk Ramsey, Johannes van der Plicht, y Alasdair Whittle (2007). Bradshaw and Bayes: Towards a Timetable for the Neolithic. *Cambridge Archaeological Journal* 17(Supplement S1): 1–28.
- Bevan, Andrew, Sue Colledge, Dorian Fuller, Ralph Fyfe, Stephen Shennan, y Chris Stevens (2017). Holocene fluctuations in human population demonstrate repeated links to food production and climate. *Proceedings of the National Academy of Sciences*: 1–8.
- Bleicher, Niels (2013). Summed radiocarbon probability density functions cannot prove solar forcing of Central European lake-level changes. *The Holocene* 23(5): 755–765.
- Bluhm, L.E. & T.A. Surovell (2018). Validation of a global model of taphonomic bias using geologic radiocarbon ages *Quaternary Research*, 1–4.
- Bronk Ramsey, Christopher (2001). Development of the radiocarbon calibration program OxCal. *Radiocarbon* 43(2; PART A): 355–364.
- Bronk Ramsey, Christopher (2009a). Bayesian analysis of radiocarbon dates. *Radiocarbon* 51(1): 337–360.
- Bronk Ramsey, Christopher (2009b). Dealing with Outliers and Offsets in Radiocarbon Dating. *Radiocarbon* 51(3): 1023–1045.
- Bronk Ramsey, Christopher (2017). Methods for Summarizing Radiocarbon Datasets. *Radiocarbon* 59(6): 1809–1833.
- Bryson, RU, RA Bryson, y A Ruter (2006). A calibrated radiocarbon database of late Quaternary volcanic eruptions. *Earth Discussions* 1(2): 123–134.
- Buchanan, Briggs, Mark Collard, y Kevan Edinborough (2008). Paleoindian demography and the extraterrestrial impact hypothesis. *Proceedings of the National Academy of Sciences* 105(33): 11651–11654.
- Chaput, Michelle A, y Konrad Gajewski (2015). Radiocarbon dates as estimates of ancient human population size. *Anthropocene*: 1–10.
- Chaput, Michelle A, Björn Kriesche, Matthew Betts, Andrew Martindale, Rafal Kulik, Volker Schmidt, y Konrad Gajewski (2015). Spatiotemporal distribution of Holocene populations in North America. *Proceedings of the National Academy of Sciences* 112(39): 12127–12132.
- Chiverrell, Richard C, Varyl R Thorndycraft, y Thomas O Hoffmann (2011). Cumulative probability functions and their role in evaluating the chronology of geomorphological events during the Holocene. *Journal of Quaternary Science* 26(1): 76–85.
- Collard, Mark, Kevan Edinborough, Stephen Shennan, y Mark G Thomas (2010). Radiocarbon evidence indicates that migrants introduced farming to Britain. *Journal of Archaeological Science* 37(4): 866–870.
- Contreras, Daniel A., y John Meadows (2014). Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *Journal of Archaeological Science* 52: 591–608.
- Cowgill, George L (1977). The Trouble with Significance Tests and What We can do about It. *American Antiquity* 42(3): 350–368.
- Crema, Enrico R, Andrew Bevan, y Stephen Shennan (2017). Spatio-temporal approaches to archaeological radiocarbon dates. *Journal of Archaeological Science* 87: 1–9.
- Crema, Enrico R, Junko Habu, Kenichi Kobayashi, y Marco Madella (2016). Summed Probability Distribution of ¹⁴C Dates Suggests Regional Divergences in the Population Dynamics of the Jomon Period in Eastern Japan. *PLoS ONE* 11(4): e0154809.

- Crombé, Philippe, y Erick Robinson (2014). 14C Dates as demographic proxies in Neolithisation models of northwestern Europe: a critical assessment using Belgium and northeast France as a case-study. *Journal of Archaeological Science* 52: 558–566.
- Culleton, Brendan J (2008). Crude demographic proxy reveals nothing about Paleoindian population. *Proceedings of the National Academy of Sciences* 105(50): E111.
- Davies, Benjamin, Simon J Holdaway, y Patricia C Fanning (2015). Modelling the palimpsest: An exploratory agent-based model of surface archaeological deposit formation in a fluvial arid Australian landscape. *The Holocene* 26(3): 450–463.
- Downey, Sean S, Emmy Bocaege, Tim Kerig, Kevan Edinborough, y Stephen Shennan (2014). The Neolithic Demographic Transition in Europe: Correlation with Juvenility Index Supports Interpretation of the Summed Calibrated Radiocarbon Date Probability Distribution (SCDPD) as a Valid Demographic Proxy. *PLoS ONE* 9(8): e105730.
- Downey, Sean S, W Randall Haas Jr, y Stephen J Shennan (2016). European Neolithic societies showed early warning signals of population collapse. *Proceedings of the National Academy of Sciences* 113(35): 9751–9756.
- Drake, B Lee, Antonio Blanco-González, y Katina T Lillios (2016). Regional Demographic Dynamics in the Neolithic Transition in Iberia: Results from Summed Calibrated Date Analysis. *Journal of Archaeological Method and Theory*: 1–17.
- Drennan, Robert D (2009). *Statistics for Archaeologists: A Common Sense Approach*. Springer.
- Edinborough, Kevan, Marko Porčić, Andrew Martindale, Thomas Jay Brown, Kisha Supernant, y Kenneth M Ames (2017). Radiocarbon test for demographic events in written and oral history. *Proceedings of the National Academy of Sciences* 114(47): 12436–12441.
- Flannery, Kent V. ed. (1976). *The Early Mesoamerican Village*. New York: Academic Press.
- Freeman, Jacob, David A Byers, Erick Robinson, y Robert L Kelly (2017). Culture Process and the Interpretation of Radiocarbon Data. *Radiocarbon*: 1–15.
- Gamble, Clive, William Davies, Paul Pettitt, Lee Hazelwood, y Martin Richards (2005). The Archaeological and Genetic Foundations of the European Population during the Late Glacial: Implications for 'Agricultural Thinking.' *Cambridge Archaeological Journal* 15(02): 193–223.
- García Puchol, Oreto, Agustín Diez Castillo, y Salvador Pardo-Gordó (2017). New insights into the neolithisation process in southwest Europe according to spatial density analysis from calibrated radiocarbon dates. *Archaeological and Anthropological Sciences*: 1–14.
- Gayo, Eugenia M, Claudio Latorre, y Calogero M Santoro (2015). Timing of occupation and regional settlement patterns revealed by time-series analyses of an archaeological radiocarbon database for the South-Central Andes (16–25 S). *Quaternary International* 356: 4–14.
- Gkiasta, Marina, Thembi Russell, Stephen Shennan, y James Steele (2003). Neolithic transition in Europe: the radiocarbon record revisited. *Antiquity* 77(295): 45–62.
- Goldberg, Amy, Alexis M Mychajliw, y Elizabeth A Hadly (2016). Post-invasion demography of prehistoric humans in South America. *Nature* 532(7598): 232–235.
- Hinz, Martin, Ingo Feeser, Karl-Göran Sjögren, and Johannes Müller (2012). Demography y the intensity of cultural activities: an evaluation of Funnel Beaker Societies (4200–2800 cal BC). *Journal of Archaeological Science* 39(10): 3331–3340.
- Hiscock, Peter, y Val Attenbrow (2016). Comment: Dates and demography? The need for caution in using radiometric dates as a robust proxy for prehistoric population change. *Archaeology in Oceania* 0: 1–3.
- Hoffmann, Thomas, Andreas Lang, y Richard Dikau (2008). Holocene river activity: analysing 14C-dated fluvial and colluvial sediments from Germany. *Quaternary Science Reviews* 27(21–22): 2031–2040.
- Johnstone, Eric, Mark G Macklin, y John Lewin (2006). The development and application of a database of radiocarbon-dated Holocene fluvial deposits in Great Britain. *Catena* 66(1–2): 14–23.

- Kelly, Robert L, Todd A Surovell, Bryan N Shuman, y Geoffrey M Smith (2013). A continuous climatic impact on Holocene human population in the Rocky Mountains. *Proceedings of the National Academy of Sciences* 110(2): 443–447.
- Kerr, T R, and F McCormick (2014). Statistics, sunspots and settlement: influences on sum of probability curves. *Journal of Archaeological Science* 41(C): 493–501.
- Kintigh, Keith W., J H Altschul, M C Beaudry, Robert D Drennan, Ann P Kinzig, Timothy A Kohler, W Frederick Limp, H D G Maschner, William K Michener, Timothy R Paukettat, P Peregrine, J A Sabloff, T J Wilkinson, H T Wright, y M A Zeder (2014). Grand challenges for archaeology. *Proceedings of the National Academy of Sciences* 111(3): 879–880.
- Manning, Katie, y Adrian Timpson (2014). The demographic response to Holocene climate change in the Sahara. *Quaternary Science Reviews* 101(C): 28–35.
- Marsh, Erik J. (2015). The emergence of agropastoralism: Accelerated ecocultural change on the Andean altiplano, 3540–3120 cal BP. *Environmental Archaeology* 20(1): 13–29.
- Mökkönen, Teemu (2014). Archaeological radiocarbon dates as a population proxy: skeptical view. *Fennoscandia Archaeologica* 31: 125–134.
- Mulrooney, Mara A (2013). An island-wide assessment of the chronology of settlement and land use on Rapa Nui (Easter Island) based on radiocarbon data. *Journal of Archaeological Science* 40: 4377–4399.
- Oh, Yongje, Matthew Conte, Seungho Kang, Jangsuk Kim, y Jaehoon Hwang (2017). Population Fluctuation and the Adoption of Food Production in Prehistoric Korea: Using Radiocarbon Dates as a Proxy for Population Change. *Radiocarbon* 59(6): 1761–1770.
- Palmisano, Alessio, Andrew Bevan, y Stephen Shennan (2017). Comparing archaeological proxies for long-term population patterns: An example from central Italy. *Journal of Archaeological Science* 87: 59–72.
- Parnell, Andrew (2015). Bchron: Radiocarbon Dating, Age-Depth Modelling, Relative Sea Level Rate Estimation, and Non-Parametric Phase Modelling. *R Foundation for Statistical Computing*, Vienna, R package version 4.1. 2."
- Peros, Matthew C., Samuel E. Munoz, Konrad Gajewski, y André E. Viau (2010). Prehistoric demography of North America inferred from radiocarbon data. *Journal of Archaeological Science* 37(3): 656–664.
- Porčić, Marko, Tamara Blagojević, y Sofija Stefanović (2016). Demography of the Early Neolithic Population in Central Balkans: Population Dynamics Reconstruction Using Summed Radiocarbon Probability Distributions. *PLoS ONE* 11(8): e0160832.
- Prates, L, G Politis, y J Steele (2013). Radiocarbon chronology of the early human occupation of Argentina. *Quaternary International* 301: 104–122.
- Reimer, Paula J, Edouard Bard, Alex Bayliss, J Warren Beck, Paul G Blackwell, Christopher Bronk Ramsey, Caitlin E Buck, Hai Cheng, R Lawrence Edwards, y Michael Friedrich (2013). IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55(4): 1869–1887.
- Rhode, David, P Jeffrey Brantingham, Charles Perreault, y David B Madsen (2014). Mind the gaps: testing for hiatuses in regional radiocarbon date sequences. *Journal of Archaeological Science* 52: 567–577.
- Rick, John W. (1987). Dates as data: an examination of the Peruvian preceramic radiocarbon record. *American Antiquity* 52(1): 55–73.
- Riede, Felix (2008). The Laacher See-eruption (12,920 BP) and material culture change at the end of the Allerød in Northern Europe. *Journal of Archaeological Science* 35(3): 591–599.
- Riede, Felix (2009). Climate and Demography in Early Prehistory: Using Calibrated 14C Dates as Population Proxies. *Human Biology* 81(2–3): 309–337.

- Rieth, Timothy M, Terry L Hunt, Carl Lipo, y Janet M Wilmshurst (2011). The 13th century polynesian colonization of Hawai'i Island. *Journal of Archaeological Science* 38(10): 2740–2749.
- Roberts, Neil, Jessie Woodbridge, Andrew Bevan, Alessio Palmisano, Stephen Shennan, y Eleni Asouti (2017). Human responses and non-responses to climatic variations during the last Glacial-Interglacial transition in the eastern Mediterranean. *Quaternary Science Reviews*.
- Shennan, Stephen (2009). Evolutionary Demography and the Population History of the European Early Neolithic. *Human Biology* 81(2–3): 339–355.
- Shennan, Stephen (2013). Demographic Continuities and Discontinuities in Neolithic Europe: Evidence, Methods and Implications. *Journal of Archaeological Method and Theory* 20(2): 300–311.
- Shennan, Stephen, y Kevan Edinborough (2007). Prehistoric population history: from the Late Glacial to the Late Neolithic in Central and Northern Europe. *Journal of Archaeological Science* 34(8): 1339–1345.
- Shennan, Stephen, Adrian Timpson, Kevan Edinborough, Susan M Colledge, Tim Kerig, Katie Manning, Mark G Thomas, y Sean S Downey (2013). Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications* 4: 1–8.
- Spriggs, Matthew (1989). The dating of the Island Southeast Asian Neolithic: an attempt at chronometric hygiene and linguistic correlation. *Antiquity* 63(240): 587–613.
- Stuiver, M., y PJ Reimer (1986). *CALIB Radiocarbon Calibration Program*.
- Surovell, Todd A, y P Jeffrey Brantingham (2007). A note on the use of temporal frequency distributions in studies of prehistoric demography. *Journal of Archaeological Science* 34(11): 1868–1877.
- Surovell, Todd A, Judson Byrd Finley, Geoffrey M Smith, P Jeffrey Brantingham, y Robert L Kelly (2009). Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science* 36(8): 1715–1724.
- Tallavaara, Miikka, Petro Pesonen, y Markku Oinonen, (2010). Prehistoric population history in eastern Fennoscandia. *Journal of Archaeological Science* 37(2): 251–260.
- Timpson, Adrian, Sue Colledge, Enrico Crema, Kevan Edinborough, Tim Kerig, Katie Manning, Mark G Thomas, y Stephen 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(C): 549–557.
- Timpson, Adrian, Katie Manning, y Stephen Shennan (2015). Inferential mistakes in population proxies: A response to Torfing's "Neolithic population and summed probability distribution of 14C-dates." *Journal of Archaeological Science* 63: 199–202.
- Torfing, Tobias (2015a). Neolithic population and summed probability distribution of 14C-dates. *Journal of Archaeological Science* 63: 193–198.
- Torfing, Tobias (2015b). Layers of assumptions: A reply to Timpson, Manning, and Shennan. *Journal of Archaeological Science* 63: 203–205.
- Wang, Can, Houyuan Lu, Jianping Zhang, Zhaoyan Gu, y Keyang He (2014). Prehistoric demographic fluctuations in China inferred from radiocarbon data and their linkage with climate change over the past 50,000 years. *Quaternary Science Reviews* 98(C): 45–59.
- Weitzel, Elic M, y Brian F Codding (2016). Population growth as a driver of initial domestication in Eastern North America. *Royal Society Open Science* 3(8): 160319.
- Weninger, Bernhard, Olaf Jöris, y Uwe Danzeglocke (2007). *Cologne Radiocarbon Calibration & Palaeoclimate Research Package*.
- Whitehouse, Nicki J, Rick J Schulting, Meriel McClatchie, Phil Barratt, T Rowan McLaughlin, Amy Bogaard, Susan M Colledge, Rob Marchant, Joanne Gaffrey, y M Jane Bunting (2014). Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland. *Journal of Archaeological Science* 51: 181–205.

- Williams, Alan N (2013). A new population curve for prehistoric Australia. *Proceedings of the Royal Society B: Biological Sciences* 280(1761): 20130486–20130486.
- Williams, Alan N (2012). The use of summed radiocarbon probability distributions in archaeology: a review of methods. *Journal of Archaeological Science* 39(3): 578–589.
- Williams, Alan N, y Sean Ulm (2016). Radiometric dates are a robust proxy for long-term demographic change: A comment on Attenbrow and Hiscock (2015). *Archaeology in Oceania* 0: 1–3.
- Williams, Alan N, Peter Veth, Will Steffen, Sean Ulm, Chris S M Turney, Jessica M Reeves, Steven J Phipps, y Mike Smith (2015). A continental narrative: Human settlement patterns and Australian climate change over the last 35,000 years. *Quaternary Science Reviews* 123(C): 91–112.