THE UNIVERSITY OF HULL

Extending the flood record – assessing the uncertainty and viability of palaeoflood data

Being a Thesis submitted for the Degree of PhD

in the University of Hull

by

Jessica Laura Moloney MPhysGeog

December 2019

For my grandpa who is my inspiration

Abstract

This study assesses the uncertainty and viability of palaeoflood records in relation to the British database, which is a collection of radiocarbon dated geomorphological fluvial deposits used to infer the flood-frequency record during the Holocene. There are different forms of evidence used to interpret flood-frequency records and there are inherent uncertainties associated with both the data used to for analysis and the method of data analysis used. Previous studies, which have used summed probability distribution functions, have failed to show how sensitive the shape of the curve is to characteristics of the data used and to the radiocarbon calibration curve. This study firstly applies sensitivity analysis testing to the British database using the summed probability distribution methodology. This study also discusses the potential to apply a robust quality control protocol to the British database to verify the ^{14}C ages currently available in line with geochronology studies that apply ^{14}C dating. Sub-datasets of the British database were created based on the following criteria: number of samples per site, sample material, archaeological context and likely association to a flood event and analysed using summed probability distribution functions. Statistical indicators were used to show how similar the sub-datasets were to the unfiltered British database. This study identifies that statistically the most reliable results are generated when five or more samples from a single site location are analysed. Secondly, an alternative technique is used to analyse the data: Lomb-Scargle spectral analysis to test the data for cyclicities. Spectral analysis is used to identify cyclicities within the British database and the residual $\Delta^{14}C$ data to identify cyclicities between the two datasets to determine if any cycles present are probable or a result of the radiocarbon calibration process. The results from this

study impact researchers using summed probability distribution functions to interpret environmental and climatic data in any field.

Acknowledgements

Firstly, I would like to express my upmost gratitude to my primary supervisor Professor Tom Coulthard. Tom, you have provided me with a constant guiding light and without your patience and understanding I wouldn’t have (sanely) made it this far! I would also like to thank another of my supervisors from the University of Hull; Dr Mike Rogerson. Mike, your love and enthusiasm that you put into your work has inspired me since I first sat in one of your lectures many years ago. Your passion has inspired me to believe that I can achieve my dreams. My final supervisor, Professor Jim Freer from the University of Bristol, I thank you for your involvement especially during the start of my PhD when I was still finding my feet. Your comments on my upgrade report really helped me to understand the wider context of this project.

I am grateful to all of the staff within the Geography department at the University of Hull but I would like to thank a couple of people individually. I would like to personally thank Dr Chris Skinner for being the best project buddy I could have asked for and being a mentor to me from the day I started and you still are now! I have learnt a great deal from you and I thank you for answering my incessant little questions! I would also like to thank Dr Anna Bird, who helped me gain a better understanding of different methods of data analysis and providing constructive feedback after my first-year upgrade report. All of the other postgraduate students that I met along my journey have made my experience so much better and provided a supportive community. Thank you to Chloe Morris, Jazmin Scarlett, Laura-Beth Jordan, Leiping Ye and Rita Santos.

I would like to express my thanks to people that have helped me in specific chapters of my thesis. Thank you to Professor Richard Chiverrell for meeting me in the early stages of my PhD and helping me to understand Oxcal calibration software; thanks to Professor Christopher Bronk-Ramsey for responding to my query on the Oxcal user forum; and Dr Tim Horsefield for teaching me about spectral analysis and sharing your experience of using it.

My family have been supportive of me throughout all of my studies and have encouraged me when I lacked in confidence and reminded me that I am not alone when I felt like I was. Thank you to my mum and sisters. I am grateful to my nephews for always being on hand for cuddles when I needed them. I am eternally grate to my fiancé and step-son, Mike and Tom for their love and support always. Mike, thank you for staying by my side through my struggles and bringing me back to the light when I lost it. Tom, thank you for reminding me to take a step back and enjoy the little things in life that are just as important as the big things!

Further thanks go to my colleagues at Hull City Council in the flood risk team, who have supported me throughout my writing up period.

Finally, this PhD was possible thanks to the Natural Environment Research Council for funding my scholarship as part of the wider project: Flooding From Intense Rainfall. I was very lucky to meet and work with lots of great people on the FFIR project and we had many fun project meetings together. Thank you Professor Hannah Cloke OBE, Dr Rob Thompson, Dr David Flack and Dr Lee Hawkness-Smith.

Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

AEP	annual exceedance probability
AMS	accelerator mass spectrometry
BC/AD	before Christ / Anno Domini
Cal years BP	calibrated years before present (BP = 1950)
CalPal	Cologne program package calibration curve
CSD	calibration stochastic distortion
FAP	false alarm probability
GPC	gas proportional counting
IntCal	international radiocarbon calibration curve
L-S	Lomb-Scargle periodogram
LSC	liquid scintillation counting
MSE	mean squared error
NOA	North Atlantic oscillation
OSL	optically stimulated luminescence
PDFs	probability distribution functions
PSIs	palaeostage indicators
RPPs	relative probability plots
SA	spectral analysis
SMSE	summed mean squared error

SSD	sum of squares of deviation
SWDs	slack-water deposits
UK	United Kingdom
USA	United States of America

Nomenclature

^{14}C	radioactive isotope of carbon with 14 neutrons, also referred to as radiocarbon
^{12}C	isotope of carbon with 12 neutrons
^{13}C	isotope of carbon with 13 neutrons
^{137}C	radioactive isotope of caesium
^{210}Pb	radioactive isotope of lead
^{7}Be	radioactive isotope of beryllium with 7 neutrons
^{10}Be	radioactive isotope of beryllium with 10 neutrons
^{14}N	stable isotope of nitrogen with 14 neutrons
CO_2	Carbon Dioxide
$\Delta^{14}C$	delta radiocarbon (per mil)
Λ	wavelength (years)
±	plus and minus (years)

Chapter 1 Introduction

1.1 Flood risk

Flooding is one of the greatest natural hazards in the world; in the UK 2.44 million properties are at risk of flooding from rivers and the sea (Environment Agency, 2016). The natural courses of rivers have been manipulated for hundreds of years to suit the development of human civilisation and agriculture (Wheater, 2006). Floodplains and locations with a water view are favoured locations for development and agriculture because of the fertile soil composition and flat land. However, floodplain development has significantly restricted the natural regime of the physical environment and the proximity of property to rivers has led to flooding becoming a risk to human life and property. The single widest used method for estimating future flood frequencies and magnitudes – and thereby risk – is probabilistic. It ranks past flood events for a given river to establish a percentage chance that a flood of a given magnitude may happen (Apel *et al.,* 2006). The annual exceedance probability (AEP) is used to manage flood risk by designing flood defences to a specific standard, such as a flood wall designed to protect against a 0.5% AEP event (Environment Agency, 2019).

Such methods based on historical instrumental data can be effective but are limited as their predictions are based on data spanning the length of the instrumented record, thereby only encompassing the range of events occurring during that time. This inherently limits the variety of events captured and makes non-stationary events (e.g. due to climate change) especially hard to capture (Met Office, 2018; Wilby *et al.,* 2007; Wheater, 2006). Instrumental data and their associated predictions can be extended by using records of past flood events from other sources. Such records may, for example, be stored in the geological/ sedimentological records. However, the methods of using such longer term sedimentological records generate a series of issues itself – that this thesis aims to explore.

1.2 Fluvial records

There are many different forms of fluvial records that are used to interpret flooding. These range from gauged records that may extend back up to 100 years, historical or anecdotal records that may stretch back further, and palaeoflood records – where sedimentological evidence of past flooding can extend the record back thousands of years.

1.2.1 Gauged records

On average, gauge records provide data for the last 35 years but there is some data available reaching back to the 19th century (Macdonald, 2006). There are around 1,500 gauging stations in UK rivers that are installed at fixed points of rivers where they record the river discharge (National River Flow Archive, 2018). There are many different types of river gauges, for example electromagnetic stations and weirs, an example of a river level gauge is shown in Figure 1.1. In the UK, river gauge records are predominantly used for water resource management (e.g. water supply), warning and managing flood risk, and for research (National River Flow Archive, 2018).

Figure 1.1 River level gauge station on the River Brant (Croft, 2006).

Gauged records can provide the most accurate high temporal resolution data on river flows. Older gauged records may only record daily means or maximums but advances in technology, including telemetry and data-logging, allow the widespread gathering of 15 minute (or smaller time-steps) resolution data. Gauged records are also very useful in corroborating with other records because they provide systematic data to support other forms of data, such as qualitative newspaper articles (Benito *et al.,* 2004). Therefore, gauged records can be used alone as evidence of floods as well as alongside other fluvial records. Empirical relationships derived from river gauge records can also be used to estimate rainfall patterns in areas that are not gauged, for example in Africa (Wilk *et al.,* 2006) and to develop hydrological models worldwide (Li *et al.,* 2015; Wilk *et al.,* 2006; Yang *et al.,* 2004). However, due to the short time length of gauged and systematic records it is not possible to gain a long-term perspective of flood frequency and magnitude. Furthermore, river gauge records can become submerged during high flows and so are unable to provide accurate readings during high magnitude flood events. Therefore, other forms of fluvial data are needed to provide long term records.

1.2.2 Historical records

Historical records are often recorded in the form of personal accounts, such as diaries, journals or newspapers as well as high-water marks on buildings (Macdonald and Black, 2009). Other historical records include aerial photography and rainfall records (National River Flow Archive, 2018). Historical records are common along the largest rivers in the world, for example the historical records of flooding on the River Nile dates back 5000 years (Knox, 2000). Long term historical records are capable of providing information for hundreds of years in the UK and have been used to reconstruct short term localised flood records, for examples see Archer and Fowler (2018) and Archer et al. (2016). Furthermore, by combining historical data along with for example, gauged records – a greater understanding of the veracity of both records can be established. However, whilst historical records offer the opportunity to extend gauged records but also *only* record the largest event, are limited temporally and spatially and also have the issue of reliability. For example, personal accounts of flooding can be exaggerated (Benito *et al.,* 2004) or land may have been subject to elevation changes so high-water marks on buildings or bridges may not be accurate.

1.2.3 Palaeoflood data

Palaeoflood hydrology is the study of fluvial sediment deposits to infer the occurrence of past floods (Baker, 2006; 2008). This can be carried out using a variety of methods encompassing geochronology, geomorphology, statistics, physical modelling and numerical hydraulic modelling (Baker, 2006; 2008). The study of palaeoflood hydrology has increased globally over the last three decades to extend the historical and gauged flood records to gain a long-term perspective on flooding. It is suggested that understanding the past frequency of floods during natural climatic and environmental changes will help us better understand future changes whilst also considering anthropogenic input (Jones *et al.*, 2010).

Palaeoflood data can be inferred from slack-water deposits (SWDs), which are fine grained sediments that are entrained during high flows and deposited in areas of low flow, for example see Figure 1.2 (Baker, 2008). The benefit of SWDs is that they are indicative of a flood event, rather than non-flood fluvial activity, and they also have a higher chance of being preserved over time (Baker, 2008) unless a larger flood occurs and reworks the original deposit. SWDs are common in America (Baker, 2008; Kochel and Baker, 1988) and Spain (Thorndycraft *et al.,* 2005), but not in the UK because there are very few bed-rock gorges, which are favourable environment for SWDs to form (Jones *et al.,* 2010).

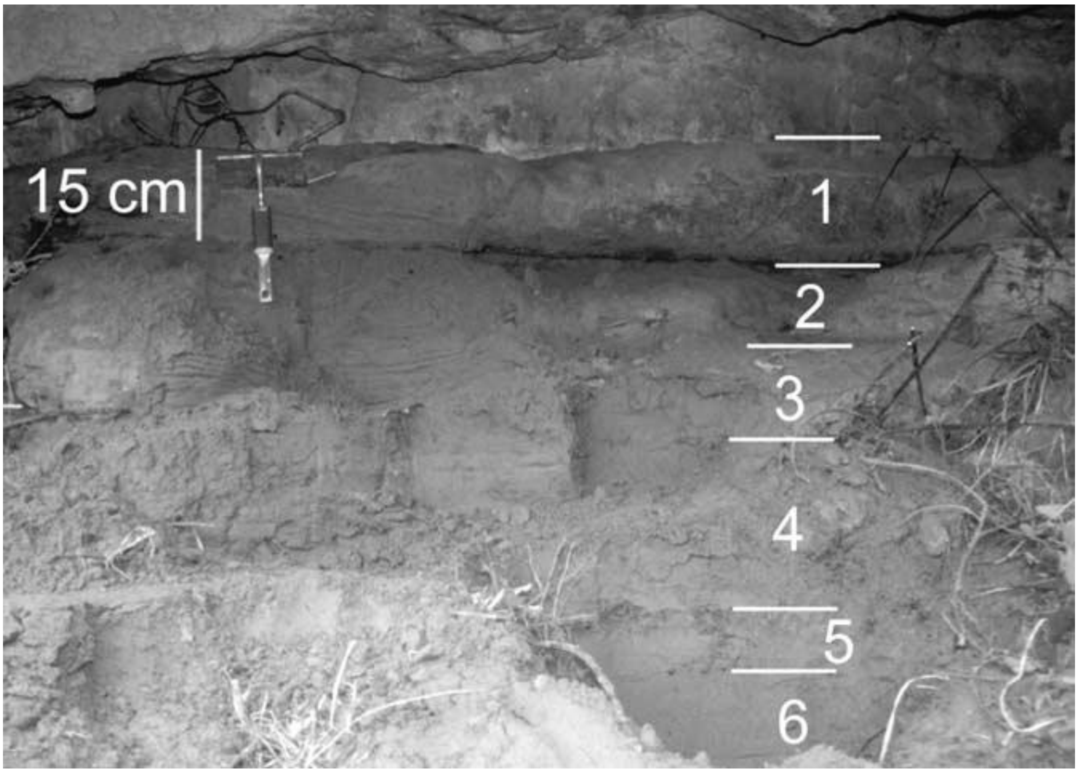


Figure 1.2 Photograph of a slack-water deposit from the Llobregat River showing consecutive flood deposits (Benito et al., 2004).

Other forms of palaeoflood data include geomorphological evidence, such as organic material like wood or plant remains, that are buried in fluvial depositional environments, for example in peat deposits or along river terraces. Geomorphological data is common in the UK (Macklin *et al.,* 2012) and Poland (Michczyńska *et al.,* 2007). Each type of palaeoflood data is explained further in 2.1.2.1.

Both SWDs and geomorphological data of fluvial activity can be dated using geochronology techniques, such as radiocarbon dating, to provide a measurement which is then inferred to represent an age (e.g. Macklin *et al.,* 2010).

1.3 Geochronology

Geochronology refers to the dating of sediment deposits or rock formations to establish a chronology of when they are formed/ deposited. For fluvial deposits, radiocarbon dating is a commonly used method of geochronology that can provide age estimations of the time of formation for organic material (thereby containing carbon: e.g. wood, bone, peat) up to 50,000 years old (Wood, 2015).

There are three different isotopes of naturally occurring carbon on Earth: carbon-12 (^{12}C), carbon-13 (^{13}C) and carbon-14 (^{14}C; Zare, 2012). ^{12}C and ^{13}C are stable isotopes and ^{12}C is the most abundant isotope of carbon on Earth (Zare, 2012). ^{14}C, also referred to as radiocarbon, is the only radioactive isotope of carbon and is rare, but the presence of ^{14}C within organic material (living and dead) can be used as a chronological tool by comparing the amount of ^{14}C in a sample to past atmospheric levels of radiocarbon (Stuvier *et al.,* 1998). A radiocarbon measurement is made from organic material using 1 of 3 techniques: gas proportional counting (GPC), liquid scintillation counting (LSC) or accelerator mass spectrometer (AMS). Each technique measures radiocarbon differently, for example, accelerator mass spectrometry (AMS) directly counts the number of ^{14}C atoms, whereas gas proportional counting measures the decay of carbon isotopes (Wood, 2015; Bronk Ramsey, 2008). Each method of radiocarbon dating has different levels of accuracy and precision of measurements.

Radiocarbon measurements are reported as radiocarbon determinations with an associated laboratory uncertainty value, for example 2,500 ± 50 years. The associated laboratory uncertainty represents the range of potential ages that a sample could be if multiple ^{14}C measurements of the same sample were taken, and so it is assumed that the true age of a sample lies within the uncertainty range of potential ages (Scott *et al.,* 2007). The reported uncertainty can be calculated using counting statistics and reflects both measurement and systematic error; the measurement error can be reduced by repeating measurements but systematic error can only be reduced if the source of error can be identified. Measurement error is lower in radiocarbon measurements made using AMS and higher in measurements made using GPC and LSC because there is greater precision when directly measuring ^{14}C atoms, because smaller samples with smaller amounts of ^{14}C can be measured whereas larger samples are required when using GPC or LSC (Wood, 2015).

Radiocarbon measurements must be calibrated because the amount of radiocarbon in the atmosphere varies over time (Reimer *et al*., 2013), and to convert the measurements from an age into a date on a calendric timescale (Blackwell and Buck, 2008). Calibrated radiocarbon dates are presented as *n* calibrated years before present (cal years BP), where BP represents the year 1950 AD, so radiocarbon dates are *n* years before 1950 (Van der Plicht and Hogg, 2006). For example, a radiocarbon age of 2,500 ± 50 years is calibrated into a range of potential dates between 2,744 – 2381 cal years BP, which is 794 – 431 years BC (Bronk Ramsey, 2009). Radiocarbon dating relies on measurements of the concentration of ^{14}C in the atmosphere over time rather than the actual rate of ^{14}C production, which is poorly understood (Wood, 2008). However, it is understood that the production of ^{14}C is effected by cosmic ray fluxes, which change the amount of cosmic rays reaching the Earth's magnetic field and hence leading to the production of ^{14}C in the atmosphere (Reimer *et al*., 2013). The production and concentration of ^{14}C is also influenced by the carbon cycle (Wood, 2015). Carbon reservoirs, which are areas where carbon is stored, can lead to the production of ^{14}C *in situ* as well as in the atmosphere (Wood *et* *al*., 2015), and affect changes in the global carbon cycle. The global carbon cycle is affected by major climatic events, such as changes in ocean circulation or transition from glacial to post-glacial periods (Siegenthaler *et al*., 1980). Therefore, the concentration of ^{14}C in the atmosphere is not stable and changes over time so calibration of radiocarbon measurements is important.

Radiocarbon calibration is carried out using a calibration curve, which is a curve constructed of known ages from dendrochronological records and their associated ^{14}C measurements (Reimer *et al.,* 2013, 09, 04). However, the calibration curve is not smooth, and peaks and plateaus in the curve can affect the calibrated date interpretation of radiocarbon measurements (Williams, 2012).

Other methods of geochronology can also be applied to date fluvial deposits, such as optically stimulated luminescence (OSL) dating (see Wallinga, 2002 for a full review of OSL techniques). The sample material needed to carry out OSL dating is quartz or feldspar grains, which are abundant in fluvial environments (Murray and Wintle, 2003). OSL dating provides estimated ages that grains were last exposed to sunlight by measuring the amount of radiation trapped within grains since burial to provide an environmental dose rate (Wallinga, 2002). Although the abundance of quartz grains in fluvial environments is beneficial for this method of dating, fluvial sediment is constantly being eroded, transported and deposited; so there is uncertainty associated with the measurement and interpretation of the luminescence signal because if grains have been re-exposed to sunlight after initial burial (also referred to as impartial bleaching) this can affect the luminescence signal (Rittenour, 2008).

1.4 The British database

The British database is an extensive database of radiocarbon dated fluvial deposits from the UK that is used to reconstruct the frequency record of past floods over the last 12,000 years (Macklin *et al.,* 2012, 2010). Importantly, the database only provides evidence of the preservation of organic material that may or may not be associated to a flood event.

The British database started with 346 samples (Macklin and Lewin, 2003) and now consists of 844 individual fluvial deposits (Jones *et al.,* 2015). The main application of the record has been to identify periods of increased fluvial activity as opposed to individual flood events (Macklin *et al.,* 2010). As the number of dated samples within the British database has increased, the data has been categorised and filtered, most notably on the position of deposits in relation to rapid stratigraphic changes. Deposits that occur *terminus post quem* are referred to as 'change after' dates and these are considered to be the most reliable indicators of flooding and preferentially used. 'Change after' dates are preferred because floods are often characterised by the deposition of different, often coarser, material compared to the material above and below it (Jones *et al.,* 2015; Macklin *et al.*, 2010).

Another big advancement of flood frequency studies is the technical development of data analysis techniques. The analysis of the British database is referred to as meta-analysis and Jones et al. (2015) defines this as *"the use of a systematic review procedure and common set of statistical techniques to combine the results of several studies"*. Data analysis methods have developed from initially using histograms (Macklin and Lewin, 1993), to summed probability distribution functions (PDFs) (Johnstone *et al.,* 2006) to most recently relative probability plots (Jones *et al.,* 2015; Macklin *et al.,* 2012; Macklin *et al.,* 2010); descriptions of these methodologies are provided in 2.3. Meta-analysis of the British database has been used to correlate flood frequency with different records of past climate changes (Macklin *et al.,* 2012; Macklin *et al.,* 2010).

Some aspects of data analysis techniques used have been criticised, for example, the application of summed PDFs to represent unrelated groups of radiocarbon dates (Williams, 2012; Chiverrell *et al.*, 2011, Michczyński and Michczyńska, 2006; Michczyńska and Pazdur, 2004), because the shape of probability curves could be influenced by a number of factors. Although the British database has undergone significant methodological developments in recent years (Jones *et al.,* 2015), different methodological approaches have not been tested rigorously for sources of uncertainty. This is important – in order to have confidence in the British database and hence in extending the palaeoflood record. Testing for uncertainty and viability will enable a best practice to be established for future studies. It is widely accepted that fluvial records will not be able to provide continuous long-term contexts, such as lakes or ocean records, but geomorphological samples are recorders of events; not proxies (Baker, 2016). So, it is essential that the data that is available is used to its maximum potential.

1.5 Uncertainty and viability of palaeoflood data

Two key areas of uncertainty in palaeoflood hydrology have been introduced in this study so far: the method of radiocarbon dating and calibration, and the use of summed PDFs to represent groups of radiocarbon dates. Uncertainty is caused by error and there is guidance for reducing error and uncertainty in climatic and environmental data analysis (Mudelsee, 2010), but there is currently not a standard of guidelines for best practice to be followed in palaeoflood hydrology studies.

Radiocarbon dating has been widely applied to a range of disciplines including archaeology (Williams, 2012), palaeoflood hydrology (Macklin *et al.,* 2012), and palaeoenvironmental studies (Michczyńska *et al.,* 2007). However, the development in radiocarbon techniques and applications has led to questions arising about their reliability. For example, on the reporting of radiocarbon dates from different sources that are used to analyse multiple dates (Wood, 2015), the effect of the calibration curve on conventional radiocarbon dates (Williams, 2012), the interpretation of calibrated age ranges in relation to flood frequency analysis (Chiverrell *et al.,* 2011a, b), and the number of radiocarbon dates needed to generate statistically reliable results (Michczyńska and Pazdur, 2004). The limitations associated with radiocarbon dating have a universal impact on all radiocarbon users but is especially important when there is a lack of other data sources to aid data analysis and interpretations, such as in palaeoflood hydrology and the British database. Each of these errors leads to uncertainty in the methods used to analyse palaeoflood data; and then there is also the question of how viable is the palaeoflood data being analysed?

Viability refers to whether a dated flood unit is considered to be representative of a flood event(s) or effectively ‘viable’, which affects the interpretation of palaeoflood data. But this is at present a subjective decision and the level of contextual information (e.g. what type of deposit) can be variable or non-existent. For example, when multiple dates are radiocarbon dated and calibrated, sometimes the only information retained about the original sample is its age (Thorndycraft *et al.,* 2011) thereby excluding the context (e.g. depositional setting). Macklin et al. (2010) attempted to overcome this viability issue by the analysis of ‘change after’ meta-data. However, filtering data thus significantly reduces the number of samples included in analysis. This poses the question: are there current methods of data analysis available so that the existing meta-data in the British database could be studied in a more holistic way? Exploring such a question is essential because each radiocarbon dated sample has a range of information attached to it as well as its age and should be utilised (Thorndycraft *et al.*, 2011). Exploring this could improve understanding of the data used to infer past floods and hence the likelihood that a dated deposit *is* associated to a flood event (or not) and hence improving the viability of palaeoflood records.

Currently, a common standard of practise does not exist in palaeoflood hydrology studies that rely on geomorphological data in relation to data selection, radiocarbon dating, data analysis and data interpretation; and it is clear that there is a need for one as the pressure to understand the relationship between flooding and climate change increases. Assessing the uncertainty and viability of the British database is key to be able to determine if the palaeoflood data represents flood events or not.

1.6 Research aim

The aim of this thesis is to provide a robust and formal investigation into the sensitivity and usefulness of the data and analysis methodologies used to analyse and interpret the British database. This will also provide useful insights in to how other palaeo-reconstructions based on large groupings of radiocarbon dated material should be viewed as well as providing a best practice for future research.

This aim will be focused through the following research questions:

1. How sensitive is the British database to the characteristics of large datasets, for example the type of sample material or the number of samples per single site?
2. How well are sub-datasets of the British database represented by frequency distributions of groups of radiocarbon dates?
3. Does the palaeoflood record show a response to past climatic changes?

Each results chapter of this thesis aims to address these research questions.

1.7 Thesis overview

This thesis is written as eight chapters. Following the introduction, a literature review (chapter 2) is provided to set the narrative for this research. Chapter 3 will provide a general methodology and more detailed methodologies will be given in each results chapter. Chapters 4, 5 and 6 present results associating to research questions 1-3 above. A synthesis is given in chapter 7. Taking into account the findings from this study, recommendations for future research are provided. The Reference list and Appendix follow thereafter.

Chapter 2 Literature review

2.1 Palaeoflood hydrology

This chapter provides a critical review of existing literature related to palaeoflood hydrology, geochronology and data analysis of fluvial deposits to identify gaps in knowledge that set the focus of this study.

2.1.1 Palaeoflood hydrology

Palaeoflood hydrology is the study of the magnitude and frequency of recent, past or ancient floods using an interdisciplinary approach involving geomorphology, sedimentology, hydrology, statistics, and modelling (Baker, 2008; Benito and Thorndycraft, 2005; Kochel and Baker, 1982). The term palaeoflood hydrology was first identified and defined by Kochel and Baker (1982) where it was highlighted that long-term records of past fluvial activity (that can be defined as a record that exceeds the length of historical datasets) are essential to calculate return periods. Kochel and Baker (1982) applied palaeoflood hydrology to SWDs in South-western Texas, USA to produce a palaeoflood record of extreme flooding events for the last 10,000 years.

Since the first formal application of palaeoflood hydrology, there have been many examples of palaeoflood studies from around the world that use a combination of often multidisciplinary methods. For example, palaeostage indicators (PSIs), such as SWDs, are the most applied approach to identify past floods because the development of SWDs are energy based (Baker, 2008) and therefore can be hydrologically modelled (Benito *et al.,* 2004). This approach is common in palaeoflood studies in America and Spain, where bed rock gorges are present to enable to formation of SWDs (Thorndycraft *et al.,* 2011; Thorndycraft and Benito, 2005; Kochel and Baker, 1982).

In depositional environments that do not support the development of SWDs, such as in the UK, Poland and New Zealand, different methods are used (Macklin *et al.,* 2012, 2010; Michczyńska *et al.,* 2007). Other PSIs, such as large boulders, have been used to determine the length of time the boulder has been in its current location using lichenometry dating (Macklin *et al.*, 1992). More recently, the most common approach to infer palaeofloods in British and other European studies is to radiocarbon date organic material, such as plant remains or wood, from fluvial depositional environments (Macklin *et al.,* 2012; Lewin *et al.,* 2005, Michczyńska *et al.*, 2007) were the date of the material is used to reflect the date of a palaeoflood (Thorndycraft and Benito, 2006; Dotterweich, 2008; Macklin *et al.,* 2005). More examples of studies are provided in 2.1.2 and further examples are given by Baker (2008).

In recent years, there has been an increase in demand for research that links climate change to fluvial environments and particularly flooding (Saint-Laurent, 2004). This has led to an increased need for knowledge of past flood events, which can be passed on to the non-scientific community to better understand flood risk management. However, the approaches to analysing and interpreting palaeoflood data are highly dependent upon the data available.

2.1.2 Evidence used in palaeoflood studies

Palaeoflood hydrology is a multi-disciplinary field using many different types of data and many different techniques to interpret these data (Baker, 2008; Benito and Thorndycraft, 2005; Benito *et al.,* 2004). The figure below (Figure 2.1) shows the different kinds of fluvial deposits, which are used to infer palaeoflood data that can form in a range of fluvial depositional environments.

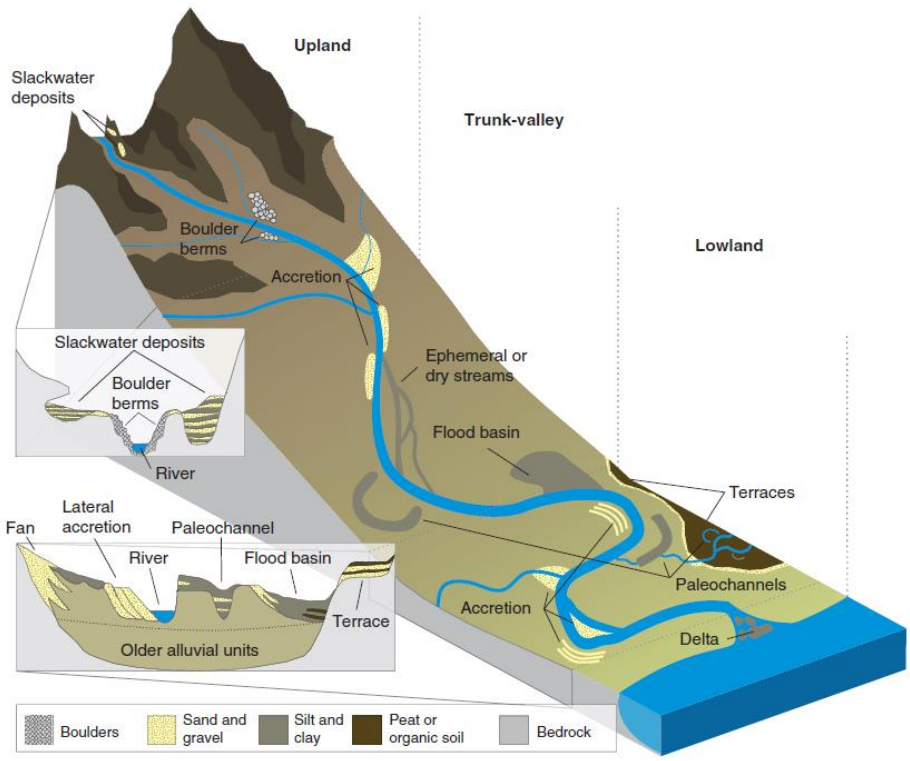


Figure 2.1 Schematic diagram of palaeoflood data across different depositional environments (Wilhelm et al., 2018).

Firstly, this review discusses PSIs, which include SWDs, and then the different forms of sedimentological evidence are introduced, including flood plain sediments and palaeochannel fills.

2.1.2.1 Palaeostage indicators

PSIs are geomorphological data within the physical environment that contain evidence of extreme events (Benito and Díez-Herrero, 2015). PSIs include SWDs, scars on trees (as shown in Figure 2.2) and large boulders.

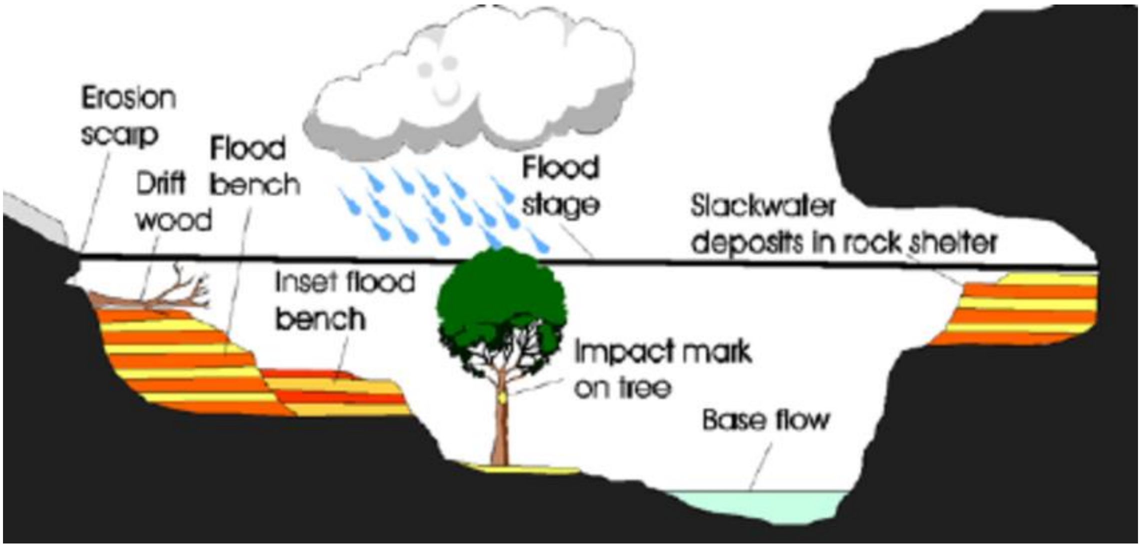


Figure 2.2 A conceptual drawing showing the different types of palaeostage indicators and flood deposits (Benito and Thorndycraft, 2005).

SWD are fine-grained sediments (silt and fine sand) that are entrained from riverbanks and deposited in areas of reduced flow velocity (Baker, 2008; Saint-Laurent, 2004; Benito *et al.,* 2004). They are predominantly found in bedrock canyons and gorges, which are common in south-western USA and central Australia (Baker *et al.,* 2009). Organic material, which includes material that has radiocarbon within it, comes in different forms, such as in pieces of wood, charcoal, or dispersed in soil organic material. Different types of organic material are defined in 2.1.4.2. If suitable organic material is present within SWDs they can be dated using radiocarbon dating techniques to provide an estimated age of the sample (Baker *et al.,* 2009). SWDs are considered a good source of palaeoflood data (Benito and Thorndycraft, 2005) because they are capable of providing information on frequency and magnitude of flooding as well as the areal extent (Patton *et al.,* 1982). Palaeodischarges can be calculated using cross-section profiles at locations where SWDs are present to model characteristics of palaeofloods (Baker, 2008).

There are issues with the preservation of PSIs because it is possible for subsequent events to erode evidence of past events, which would result in an incomplete and incorrect interpretation of the data. In addition, there are many different methodologies to calculate palaeodischarges (Baker, 2008) that may result in a lack of consistency between studies.

Other types of PSIs that can be used to indicate increased fluvial activity include boulder bars and berms. These are large boulders that have been transported and deposited during high magnitude flood events (Macklin *et al.,* 1992). They are most common in upland areas and can be used as an indicator of the frequency and magnitude of past floods because they signify higher than normal stream flow (Milan, 2012). Large boulders support the establishment and growth of lichen, which can be used to provide an estimation of their age. Lichen grows in dry environments and as it ages it grows, linearly expanding their diameter. By calibrating the rate of diameter growth to local known ages (e.g. lichen diameters on gravestones) researchers can estimate how long the boulder has been at its current location (Wohl, 1992). There are uncertainties associated with calibrating the rate of growth because there are many different statistical techniques available to generate lichenometric growth curves and no standardised approach for interpreting lichen data (Bradwell, 2009).

2.1.2.2 Sedimentology, Geomorphology and the depositional environment

Fluvial sediments are considered sensitive to environmental and climatic changes on a regional and global scale (Macklin and Lewin, 2003). Therefore, changes to the magnitude and/or the frequency of floods, can change the ability of a river to erode and deposit material (Lewin *et al.,* 2005). For example, typically more erosion will occur during periods with a high frequency of floods; equally, deposition will be greater during periods of stability and low frequency of floods (Lewin *et al.,* 2005). Therefore, by studying the sedimentology and geomorphology of rivers and their depositional environments we can infer past environmental and climate changes. However, it is not easy to distinguish autogenic sediment processes from climatic changes so an important challenge for palaeoflood hydrologists is interpreting the available data (Lewin *et al.,* 2005). Autogenic processes refer to internal processes that occur within localised fluvial systems that rework and deposit material without the influence of environmental change (Van de Wiel and Coulthard, 2010; Lewin *et al.,* 2005).

Figure 2.3 shows examples of the contexts that datable materials can be found based on palaeoflood data studied in the UK. Understanding the context of a sample could help to distinguish between autogenic processes and climatic changes.

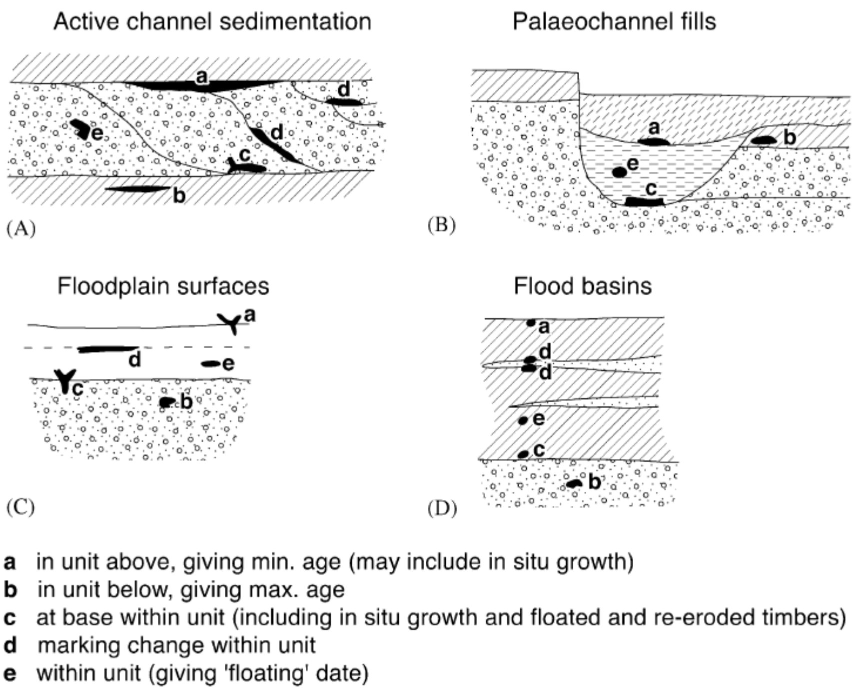


Figure 2.3 The sedimentary context for datable organic material in alluvial units (Lewin et al., 2005).

Figure 2.3 identifies the location of deposits based on criteria in the context of four different depositional environments: active channel sedimentation, palaeochannel fills, flood plain surfaces and flood basins (Lewin *et al.,* 2005). In the British database, deposits have been assigned a depositional environment based on the location of the sample in relation to surrounding stratigraphy (Macklin *et al.,* 2012). An exploration of each depositional environment is given below.

Organic material found in fluvial depositional environments are used to interpret potential ages of past flood events. The most common dating technique used in fluvial studies, particularly for Holocene studies, is radiocarbon dating (Baker, 2008). Radiocarbon dating of organic material provides an age of the sample, which can be interpreted based on the sample's relative location (see Figure 2.3). This is discussed in greater detail later in this chapter.

Based on the British database, most datable organic material is found in the middle of sedimentary units, which are interpreted to not represent flood events, but rather periods of stability that has enabled a large sedimentary unit to develop through continuous deposition of material. Material that is found directly above or below sedimentary boundaries is interpreted to have been transported and deposited there during a flood event to have caused the sedimentological change in deposition. These materials could therefore provide minimum or maximum ages of a palaeoflood (Lewin *et al.,* 2005). However, the complex nature of fluvial environments also affects the reliability of the palaeoflood record because they are constantly changing and sediment as well as organic material is reworked.

To take this into account, recent palaeoflood studies have recorded the alluvial ensemble of the depositional environment that a deposit was taken from (Lewin *et al.,* 2005). Ensembles, which have previously been termed 'genetic floodplains’ and ‘alluvial architecture’, refer to fans and cones, upland gullies and streams, braided systems and active/ inactive meandering, and anastomosing systems (Lewin *et al.,* 2005). Alluvial ensembles therefore represent the complexity of autogenic sedimentation processes within a flood plain, which are useful for determining which Holocene alluviation model to use on the data available (Lewin *et al.,* 2005).

Channel bed sediments can be analysed for palaeoflood evidence when incision of the river results in an exposed surface. Depending on the level of incision, the lithology of a sedimentary unit can be clearly shown. Organic material may also be present and easily accessible to date the sequence (Baker, 2008). Channel bed sediments are vulnerable to change by many environmental processes, including weathering and increased discharge, which may erode sediment, and flood events, due to the exposure of channel bed sediments. For example, in Staindale, North Yorkshire Moors, the incision and lateral migration of a meander exposed a former channel bed. This allowed the extraction of driftwood to date a palaeoflood whilst also considering the stratigraphy of the sedimentary unit (Richards, 1981).

Palaeochannels are parts of river channel that have become disconnected from the main channel, for example through channel avulsion or the formation of an oxbow lake (Knighton, 1998). Palaeochannels can accumulate material from flood deposits, for example when a channel spills over its banks into a palaeochannel and deposits material; as well as by organic accumulation (Lewin *et al.,* 2005). This type of depositional environment is considered highly favourable in palaeoflood studies because they preserve organic material well (Lewin *et al.,* 2005). Palaeochannels can also be used to reconstruct flow conditions using the physical parameters of the palaeochannel and using an environmental proxy, such as caddisfly assemblages, to assess the flow conditions based on knowledge of species preferences (Greenwood *et al.,* 2006). However, the structure of sediment within a palaeochannel can be subject to change. As mentioned previously, palaeoflood data can be erased by subsequent events but in addition bioturbation can also mix up fluvial deposits. Bioturbation can greatly alter the vertical sequencing, which alters the interpretation of fluvial deposits (Lewin *et al.,* 2005). Therefore, the vertical stratigraphy of palaeochannel fill units may not represent the correct chronological order of events.

Floodplain sediments are found in the fluvial floodplain, which are typically flat, lowland areas that are characterised by layers of coarse material in between layers of fine grained sediment deposits (Lewin *et al.,* 2005). Palaeofloods are interpreted by the presence of coarse layers (Lewin *et al.,* 2005; Brazier *et al.,* 1988). If organic material is present at a lithological change, the material can be radiocarbon dated to infer an age of the change, which is interpreted to reflect a palaeoflood (Lewin *et al.,* 2005). It is possible to gain evidence from floodplain sediments through coring, trenches, cut banks and other exposures to observe if abrupt lithological changes are present. These methods can be costly and can also lead to preferential sampling which ultimately influences the interpretation of palaeoflood data, because sections that do not show a clear presence of datable organic material or abrupt lithological changes are often overlooked (Chiverrell *et al.,* 2011). A large number of samples within the British database are from floodplain sediments (Macklin *et al*., 2012). It is important to consider that the presence of organic material does not always correspond to specific palaeoflood events (represented by coarse sediment layers), because organics can be deposited from aeolian activity or human activity as well as overbank flow, for example see Tayler and Lewin (1997). Furthermore, the radiocarbon measurement of organic material (e.g. wood) does not necessarily represent the date of deposition because the material may have been re-worked, for example through bioturbation. There are a small number of samples within the British database that are associated with archaeological studies through information provided in the original studies that the radiocarbon dates were taken from (Macklin *et al.,* 2012); but other than this clear identification of context, it is a great challenge to interpret the context of every sample.

Flood basins are depressions that are formed between a river channel and valley sides and are often caused by the development of levees (Lewin *et al.,* 2005). In Britain, such depressions are characterised by wetland environments because of moisture accumulation and hence organic deposition, which forms flood basin sediments (Lewin *et al.,* 2005). It is often hard to distinguish between flood basin and flood plain sediments when the deposit is dominated by fine sediment (Lewin *et al.,* 2005). Flood basin sediments are common in upland river systems of Highland Scotland because the reworking of glacial deposits, such as moraines, lead to the formation of depressions that flood basin sediments accumulate in (Lewin *et al.,* 2005). Flood basin deposits have been identified and dated for the Milfield Basin in Northumbria, Northeast England. Furthermore, the accumulation of flood basin sediments can be altered by mass movements, for example in the Borve Valley, Barra, Outer Hebrides of Scotland, where age-depth anomalies are attributed to mass movement as opposed to flooding (Ashmore *et al.,* 2000).

2.1.2.3 Lake deposits

Lakes deposits are another terrestrial archive of palaeoflood data that are used to interpret the frequency and magnitude of past flood events (Wilhelm *et al*., 2018). Lakes deposits are made up of layers of sediment that have been deposited on the lake bed over time (Chiverrell *et al*., 2019; Schillereff *et al*., 2014), and the thicknesses and material compositions of these layers may be used to interpret past environmental and climatic conditions (Chiverrell *et al*., 2019; Schillereff *et al*., 2019; Schillereff *et al*., 2014). The interpretation of lake deposits as palaeoflood records is based on the assumption that transportation of material deposited on lake beds is affected by river discharge (Schillereff *et al*., 2014). Unlike PSIs (as discussed in 2.1.1.2), which provide fragmentary records that are often not contemporaneous, lake records can provide continuous records over different timescales (Wilhelm *et al*., 2018; Schillereff *et al*., 2014). A three-stage approach is generally applied to palaeoflood studies of lake deposits: firstly, sediment coring to extract deposited layers of sediment, secondly, analytical methods to calculate a chronology of flood frequency, and thirdly, comparison with other environmental proxies to calculate palaeoflood magnitudes to make robust interpretations (Schillereff *et al*., 2014).

Sediments that form lake deposits can originate from contributing catchments that are eroded and transported during rainfall events to the lake (see Figure 2.4 below), and are typically made up of fine-grained clay and silt, diatoms and organic matter (Schillereff *et al*., 2014). Extreme flood events, which are interpreted from layers of flood deposits, can be identified in two main ways from lake sediments. There are different techniques currently available and widely used to interpret palaeoflood data from lake deposits; and palaeoflood data is generally used to identify flood-rich or flood-poor periods on millennial-timescales, or to identify individual flood events (Schillereff *et al*., 2014).

There are direct and indirect methods of analysing lake deposits. Direct methods include the analysis of particle sizes within a sample and visual inspection of sediment cores. The particle size of sediments is typically larger in flood deposits (Schillereff *et al*., 2014) and is used as an indicator for river energy and discharge to estimate flood magnitude (Wilhelm *et al*., 2018). Particle size can be analysed using micro-scale X-ray fluorescence (μXRF) core scanning (Schillereff *et al*., 2014). Additionally, changes in the colour of different layers can identify a change in sediment supply, for example sediment that is high in organic content is often darker in colour and so can be identified as a different layer to material above and below it (Wilhelm *et al*., 2018). The content of organic matter is dependent upon the surrounding catchment area that provides the source of material to the lake and can include detrital plant material and humic substances (Schillereff *et al*., 2014). Indirect methods include analysis of bulk geochemical composition of lake deposits and can provide a proxy for erosion by measuring the relative proportion of stable and unstable elements (Schillereff *et al*., 2014), such as lead, zinc and copper (Chiverrell *et al*., 2019). When flood deposits have a low organic content, they may not be visible and so require laboratory analysis techniques, for example geochemical proxies such as pollen or isotopic analysis.

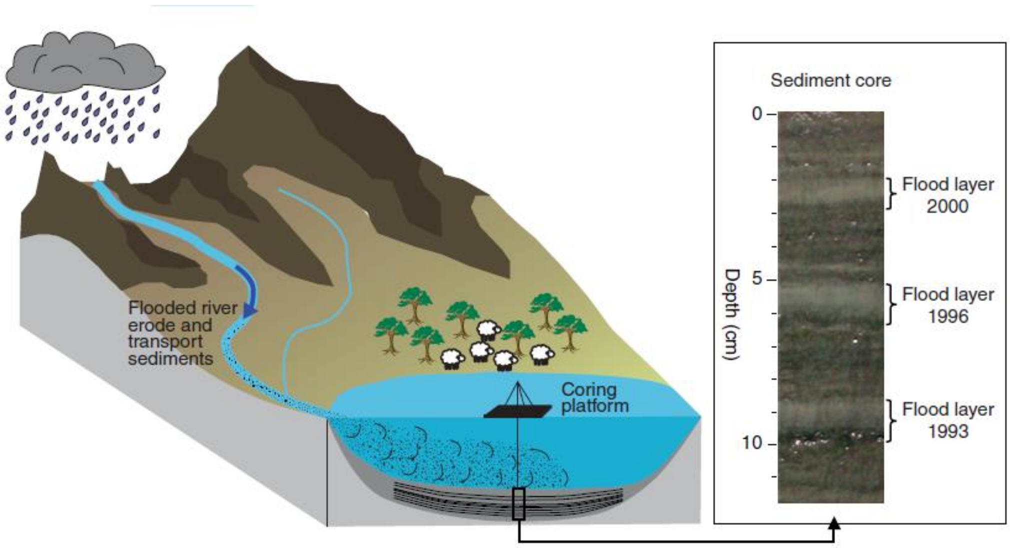


Figure 2.4 Schematic drawing showing the development of palaeoflood deposits in lakes and a photograph of deposited flood layers (Wilhelm et al., 2018).

In lake basins that have a good coupling relationship with its catchment, there is a greater chance that the deposition of material on the lake bed will be characterised by laminations, where different layers (varves) can be linked to annual layers of deposition (Chiverrell *et al*., 2019). Therefore, lake basins where a single dominant sediment source input can be identified are preferred sites to take sediment cores because the laminations or layers of sediments will have less interference from other sediment sources (Schillereff *et al*., 2019). In-lake processes are important to consider when interpreting the depositional mechanisms of lake deposits (Schillereff *et al*., 2014). For example, the sediment accumulation rate can be calculated to estimate the amount of sediment being delivered to a lake and the deposition of sediment is affected by the distance of the dominant source of material (Schillereff *et al*., 2014). A small-magnitude flood event may be evident in sediment cores taken from a location close to the sediment source but not at the opposite side of the lake, whereas a high-magnitude flood event may be evident in sediment cores across the lake bed (Schillereff *et al*., 2014).

The preservation of lake deposits depends on lake basin morphology, hydrology, sediment regime and sediment remobilisation (Schillereff *et al*., 2014). Similarly to sedimentary deposits in fluvial depositional environments discussed in 2.1.2.2, lake deposits are also prone to bioturbation. Furthermore, one flood event could deplete the source of sediment supply so that a proceeding flood event could be recorded with a smaller layer of coarse material, which may not accurately reflect the real magnitude of events (Schillereff *et al*., 2014). Schillereff et al. (2014) provide a conceptual model that can be used to determine advantageous and disadvantageous conditions for lake deposit formation and preservation.

After the collection of reliable lake deposits, chronologies are estimated to create Bayesian age-depth models and a widely applied methodology is given in Blaauw and Christen (2011). Similarly to the interpretation of other types of sedimentological palaeoflood data, researchers are faced with the challenge of identifying between human and climate forcing causal mechanisms (Schillereff *et al*., 2019). To aid this challenge, Schillereff et al. (2014) developed a protocol to aid the identification and interpretation of palaeoflood data from lake deposits (see Figure 2.5).

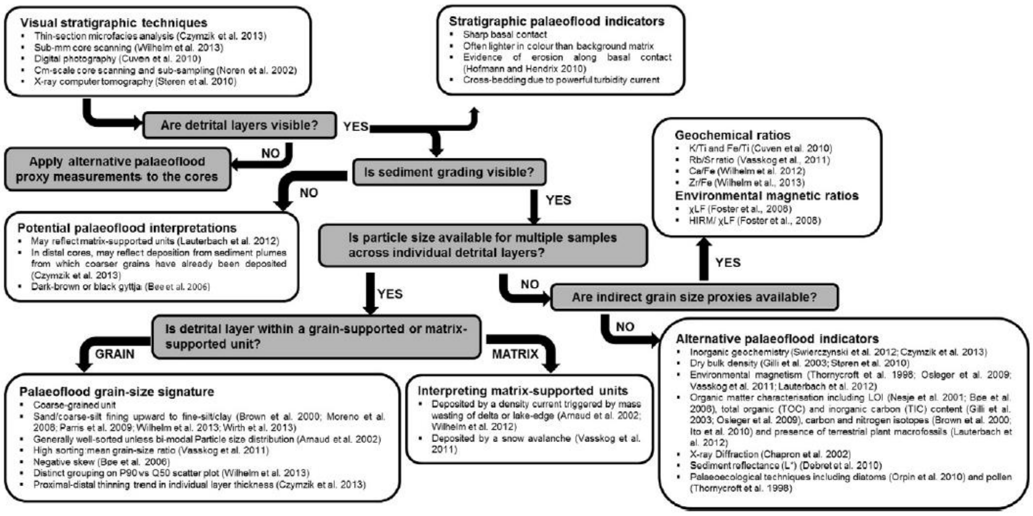


Figure 2.5 A protocol to aid interpreting palaeoflood data from lake records (Schillereff et al., 2014).

Precise chronologies of lake deposits are essential to estimate past flood frequency and to calculate accurate return periods/ AEPs (Schillereff *et al*., 2014). High precision chronologies can be made from varved deposits where seasonal-scale interpretations can be made, with most lake deposits able to provide decadal- to centennial-scale chronologies (Wilhelm *et al*., 2018). In recent lake deposits it can be possible to identify individual flood events but in most cases, periods of increased/ decreased deposition of flood deposits are identified rather than individual flood events (Wilhelm *et al*., 2018). Different methods of geochronology are applied to create age-depth profiles of lake deposits and in most studies a combination of methods are used (Chiverrell *et al*., 2019; Schillereff *et al*., 2014). Palaeoflood data from lakes have been interpreted as far back as 4500 years from Lake Atnsjøen, eastern Norway using ^{14}C dating (Nesje *et al*., 2001), and up to 700 years in the Iberian Peninsula using varve counting and geochemical analysis (Corella *et al*., 2014). The longest record from UK lake deposits extends back 1500 years from cores taken from Brotherswater, northwest England, using radiometric dating, ^{210}Pb, ^{226}Ra, ^{137}Cs, ^{241}Am (Schillereff *et al*., 2019).

Schillereff et al. (2014) highlight the challenges that are associated with extracting, analysing and interpreting palaeoflood data from lake records to create catchment-scale flood frequency and magnitude records. It is summarised that although lake deposits offer the longest and most continuous terrestrial source of palaeoflood data, care must be taken to correctly interpret the mechanism behind the formation of lake deposits.

2.1.3 The British database

Many of the different types of sedimentological evidence discussed in 2.1.2 are present within British fluvial catchments and feature in the British database (Jones *et al.,* 2015; Macklin *et al.,* 2012). The next part of the literature review focuses on the application and interpretation of the British database.

2.1.4.3 Interpretation of the British database

Jones et al. (2015) presents an overview and comparison of the periods of flooding identified as the data within the British database has increased and the methodologies have developed (see Table 2.1 below).

Table 2.1 Periods of flooding identified by interpreting the British database (Jones et al., 2015).

Episodes of major Holocene riverine flooding identified from the PDC of ^{14}C dated fluvial units in British rivers (Macklin et al., 2005) (506 dates; 263 'change dates')	Episodes of major flooding in the UK, based on the 2010 analysis of ^{14}C-dated Holocene fluvial units Macklin et al., 2010a) (776 dates; 236 'change after' dates)	Episodes of major flooding in the UK, based on the latest analysis of ^{14}C-dated Holocene fluvial units (this study) (844 dates; 252 'change after' dates)
11160	11800–11100	**11800-11100**
	10700–10400	**10700-10400**
	9400–9100	**9300-9100**
	7800–7700	**7800-7700**
	7600–7300	**7600-7300**
6820	6900–6500	**6900-6500**
5730	6300–6200, 6000–5700	**6300-5700**
5540		
	5300–5100	**5300-5100**
4840, 4520	4900–4500, 4300–4200	**4900-4200**
3540	3600–3400	**3600-3400**
	2900–2800	**2900-2800**
2730		
2550		
2280	2300–2000	**2300-2000**
1950		
1650		
1290	1500–1000	**1500-1000**
860, 660, 570	900–500	**900-500**
	300–0	**300-0**

It is evident from Table 2.1 that the interpretation of the British database has developed over time with the addition of new data. In the early studies individual dates were used to represent periods of flooding (Macklin *et al.,* 2005) but has now developed to being represented by a range of ages (Jones *et al.,* 2015; Macklin *et al.,* 2010). The development of the methodology used to interpret palaeoflood data is based on a probabilistic approach, which is discussed further in 2.3. However, despite the developments in analytical methods, the data may still be affected by errors associated with radiocarbon dating. Chiverrell et al. (2011a, 2011b) argue that it is not possible using the data in the British database to identify individual flood events, and the probabilistic approach applied may not be suitable because of the large amount of uncertainty associated with using radiocarbon dated palaeoflood data (Chiverrell *et al.,* 2011a, 2011b).

Previous interpretations of the British database, as summarised by Jones et al. (2015), have identified centennial and millennial-scale occurrences of flooding (Wilhelm *et al.,* 2018; Macklin *et al*., 2010). Macklin et al., (2010) argue that this identification allows fluvial data to be used alongside climate proxies, such as pollen records and North Atlantic drift ice, to determine the cause and effect of climate and environmental change on British rivers. However, it is essential to compare the British database with other proxies, such as pollen records, in order to determine the scale of the change shown in the record (Lewin *et al.,* 2005) and to avoid vague interpretations (Knighton, 1998). This is important because a key reason for studying flood frequency records is to gain a longer perspective of the frequency and magnitude of floods within the parameters of natural climate change so that future risk can be prepared for (Jones *et al.,* 2010).

2.1.4 Limitations of palaeoflood data

The interpretations that can be made from the British database are limited in comparison with other forms of palaeoflood data, such as SWDs and PSIs. For example, it is possible to calculate palaeodischarges from PSIs, such as SWDs (Baker, 2008). However, it is not possible to calculate palaeodischarges when interpreting deposited organic material because the flood extend, including height, is not always available using additional sources of data. For instance, high water markers on buildings, but even if multiple sources of data are available, there is no guarantee that each source of data represents the same event. PSIs provide clear evidence of an extreme event, interpreted as a palaeoflood, whereas the deposition of organic material may occur in both extreme and non-extreme events. Therefore, it is difficult to identify individual events and the magnitude of palaeofloods using the British database (Jones *et al*., 2010).

Baker (1973) determined that using geomorphological evidence to support the occurrence of palaeofloods was unreliable because it was not possible to date floods accurately enough for flood frequency analysis. There are further limitations associated with the methods of geochronology and data analysis applied to palaeoflood data (Williams, 2012; Chiverrell *et al.*, 2011) and these are discussed in 2.2 and 2.3. To address these issues, previous studies have applied meta-analysis techniques to interpret large numbers of radiocarbon dated palaeoflood data (Jones *et al.*, 2015, 2010; Macklin *et al.,* 2012; Michczyńska *et al.,* 2007). Furthermore, previous studies of the British database have analysed data based on a sample's stratigraphic location (Jones *et al.,* 2015; Macklin *et al.,* 2012, 2010). To increase the reliability and confidence of interpreting palaeoflood data (Lewin *et al*., 2005), river 'change dates' were suggested (Macklin *et al.*, 2010). 'Change dates' represent radiocarbon dated material that is found on the boundary of sedimentary unit change. Materials termed 'change after' dates signify changes in the sedimentary deposition giving the minimum age of a flood event (Macklin *et al.*, 2010). This method of data analysis could cause preferential sampling, which could produce clusters of radiocarbon ages that are not representative of the sediment record.

Sample bias could be caused, particularly in bulk organic material, when organisms are exposed to younger sources or carbon, which causes younger radiocarbon ages, or exposure to older sources of carbon, for example hard wood, which could cause older radiocarbon ages (Chiverrell *et al*., 2011). The collective analysis of radiocarbon dates significantly reduces the opportunity to identify sample bias when only the radiocarbon ages are studied and not any supporting data, such as sample material or whether ages are contemporaneous or not.

Other methods of data selection have also been explored, such as analysing data by depositional environment and precipitation region (Macklin *et al.,* 2010) but the analysis of 'change after' dates is the most favourable because it is more likely to practically represent a flood event (Jones *et al.,* 2015). However, the information captured when applying meta-analysis is limited; in fluvial studies the information typically captured is the radiocarbon determination (Thorndycraft *et al.,* 2011), which is interpreted to represent an age of a fluvial event. Other information about palaeoflood data, such as sample material, is lost (Chiverrell *et al.,* 2011; Pettitt *et al.,* 2003) and it is unknown what effect that other data characteristics have on the interpretation of fluvial deposits. Data selection offers the opportunity to filter large datasets and refine the interpretations made – though this filtering is in itself subjective. But, given the increase in the number of samples within the British database over recent years (Jones *et al.,* 2015), there is an opportunity to explore the effect of data selection/filtering.

2.1.4.1 Analysis of multiple dates from single sites

The British database is a compilation of single dates from multiple sites and multiple dates from individual sites. It is untested how site-specific data analysis could influence the interpretation of palaeoflood data, in particular where there are multiple radiocarbon dates available for individual sites that have not been studied in isolation from the British database.

2.1.4.2 Sample material

When palaeoflood records are interpreted from radiocarbon dated organic material, different material may indicate different depositional environments; for example peat accumulation in a palaeochannel or wood debris in a floodplain deposit. Previous research (Macklin *et al.,* 2010) categorised the British database up into sub-datasets based on depositional environment and precipitation region, but this has not been carried out for sample material even though this information is available for most samples.

Meta-analysis is widely applied to large numbers of radiocarbon dates (Jones *et al.,* 2015) and it is also applied without considering the effect that different sample material could have on the analysis and interpretation of data. For example, in soils there is a continuous input of carbon so this causes radiocarbon determinations of organic material within sections of soil formation to be younger than the true age of soils, which would cause the age of an associated event to be overestimated (Wang *et al.,* 1995). The development of soil is a continuous and on-going process and exchange with atmospheric carbon differs depending on climate and soil depth (Wang *et al.,* 1995). Within the British database, there are different identifications of organic material that include organic soil, buried soil, peat, humified peat, amorphous peat, palaeosol, organic mud and fine particulate organic sediment (Macklin *et al.,* 2012) that are analysed collectively, and their usefulness is determined by the classification of 'change after' dates only (Jones *et al.,* 2015; Macklin *et al.,* 2010). The method of sample material identification within the British database was based on data available from the original source (Macklin *et al.,* 2012). The collective analysis of bulk material is based on the assumption that all organic material was alive and died at the same time (Bronk Ramsey, 2008), which may not be a true reflection of reality. Furthermore, radiocarbon dating provides an estimate of the age of the incorporation of radiocarbon into an organism, which occurs at the start of an organism's life, rather than at the time of death (Bronk Ramsey, 2008). The time of the life of an organism, which would affect the radiocarbon concentration in an organism, and death, which would affect the length of time a deposit had been buried for, could affect the measurement and interpretation of bulk material. Therefore, radiocarbon determinations of bulk soil organic carbon are questionable and should be considered when included in analysis.

The type of sample material chosen often depends on the perspective of research, for example, archaeological studies will focus on radiocarbon dating charcoal and wood, to identify periods of human occupation (Haesarts *et al.,* 2010), and palaeoenvironmental studies favour the analysis of peat, for example to identify cold and dry periods during the Holocene (Weijian *et al.,* 2002). However, in the analysis of the British database, archaeological samples are excluded, which often includes charcoal, bone and wood sample material (Macklin *et al.,* 2012). In previous studies, the suitability of a sample has been based on its stratigraphic location rather than the sample material itself (Jones *et al.,* 2015). Plant remains are interpreted in the same way as other individual sample materials in palaeoflood studies; if the sample material is located at the base of a stratigraphic boundary change, 'change after' date, then the sample is included in analysis (Jones *et al.,* 2015; Macklin *et al.,* 2012).

The influence of sample material on the interpretation of fluvial deposits has received little attention even though it is considered an important data characteristic by geochronologists because the type of sample material determines the methods of radiocarbon pre-treatment and its subsequent level of accuracy (Rodríguez-Rey *et al.,* 2015; Brock *et al.,* 2010). Considering that palaeoflood records are reliant on the availability and interpretation of organic material, care should be given to the analysis of sample material by incorporating best practise methods from other studies and applying them to palaeoflood studies.

2.1.4.3 Sample material context in relation to archaeological context

Sample material context refers to the link between a fluvial deposit and archaeological activity, as material deposited within fluvial depositional environments may also include archaeological material, for example bone and charcoal. Given that radiocarbon dating and the use of summed probability distribution functions, described later, were initially applied to archaeological studies, it is unsurprising that the British fluvial database includes sample material that is associated with archaeology.

Macklin and Lewin (2003) previously determined that archaeological material should be excluded from analysis because of the likelihood that the material was deposited as a result of human activity rather than fluvial activity. Excluding archaeological material is therefore a subjective decision so there is an opportunity here to provide the first statistical analysis to show the influence of archaeological material on the shape of the summed probability distribution function generated using the British database.

2.1.4.4 Sample material association to a flood event

Association refers to the likelihood that a sample is associated with a flood event based on its stratigraphic location. Association has been used previously in quality control of radiocarbon dated material and dates that are interpreted not to be associated are excluded from analysis (Rodriguez-Rey *et al.,* 2015). In previous studies, 'change after' dates have been assumed to be associated with fluvial activity because the sample material was situated at the base of a stratigraphic change to provide a *terminus post quem* (Macklin *et al.,* 2010); indicating an environmental change in deposition. This study tests this assumption statistically and determines the influence 'change after' dates have on the interpretation of the British database.

2.2 Geochronology

There are several different methods available to date organic material in fluvial deposits. For recent sediments (< 100 years) dating techniques such as ^{210}Pb or ^{137}C can be applied (Aalto and Nittrouer, 2012), or dendrochronology to date the age of scarred trees on floodplains, which can then be correlated with local meteorological records (Ballesteros-Cánovas *et al.,* 2015; Zielonka *et al.,* 2008). For older sediments (> 100 years) other dating techniques are available, such as radiocarbon dating of organic material (Bronk Ramsey, 2008) and optically stimulated luminescence (OSL) to date quartz grains based on when a sample was last exposed to sunlight (Wallinga, 2002). Table 2.2 highlights geochronological techniques that are currently available and their associated age ranges, benefits and limitations (after Aalto and Nittrouer, 2012).

Table 2.2 A summary of the geochronological techniques available for providing age estimates of sedimentary deposits (after Aalto and Nittrouer, 2012).

Dating technique	**Age range (years)**	**Benefits**	**Limitations**
Fallout radionuclides			
^{210}Pb	1 - 110	Low cost and quick to measure with alpha spectrometry	Poor gamma emitter, chemistry for alpha
^{137}Cs	1959/63 peaks	Useful for recently deposited sediments to identify a known age	Signal grows weaker over time
^{7}Be	Under 0.6	Resolves fast events	Not a long age range
^{10}Be	1,000 – 5,000,000	Dates ancient sediments	Expensive to process, interpretations are based on difficult assumptions
Material in deposits			
^{14}C	100 – 50,000	Easy to collect samples, low cost and fast with accelerator mass spectrometry, lots of ^{14}C laboratories available	Expensive if more than one sample is used for dating, problems with influence of the calibration curve
Optically simulated luminescence	300 – 100,000	Provides age of material deposition	Expensive, problems with resetting if sample is moved after initial deposition
^{10}Be in situ	1,000 – 5,000,000	Provides age of sample exposure, only when shielding is known	Expensive, problems with inherited age and shielding

In relation to palaeoflood hydrology studies, radiocarbon dating and OSL dating are widely applied because advances in technology and accessibility of data have increased the practicality of using these techniques (Baker, 2008). The understanding of fluvial environments has also increased our ability to interpret dated fluvial deposits (Jones *et al.,* 2015; Lewin *et al.,* 2005).

2.2.1 Radiocarbon dating

2.2.1.1 Overview of radiocarbon dating

Radiocarbon dating is a method of measuring the radiocarbon content in organic material and comparing it with past levels of radiocarbon in the atmosphere to calculate an estimated age of sample material (Stuvier *et al.*, 1998). Carbon is an element that is present in all living organisms and all living organisms need it to survive (Blackwell and Buck, 2008). Radiocarbon is produced in the Earth's upper atmosphere by interaction between cosmic rays and nitrogen atoms (Bronk Ramsey, 2008). Radiocarbon can also be produced at the ground level (*in situ*), where the interactions of atoms are more complex, and the production rate is up to two magnitudes lower than production rates in the upper atmosphere (Bronk Ramsey, 2008). Despite the complexities of radiocarbon that is produced *in situ,* the radiocarbon content is still measurable in the same way as radiocarbon produced in the upper atmosphere is as long as there has not been contact with other sources of carbon (Bronk Ramsey, 2008). After production, radiocarbon enters the lower atmosphere and is dispersed in marine and terrestrial environments (Bronk Ramsey, 2008). Radiocarbon in marine environments is beyond the scope of this study so is not discussed. The presence of radiocarbon in terrestrial environments is initiated by photosynthesis of green plants (see Figure 2.4; Bronk Ramsey, 2009).

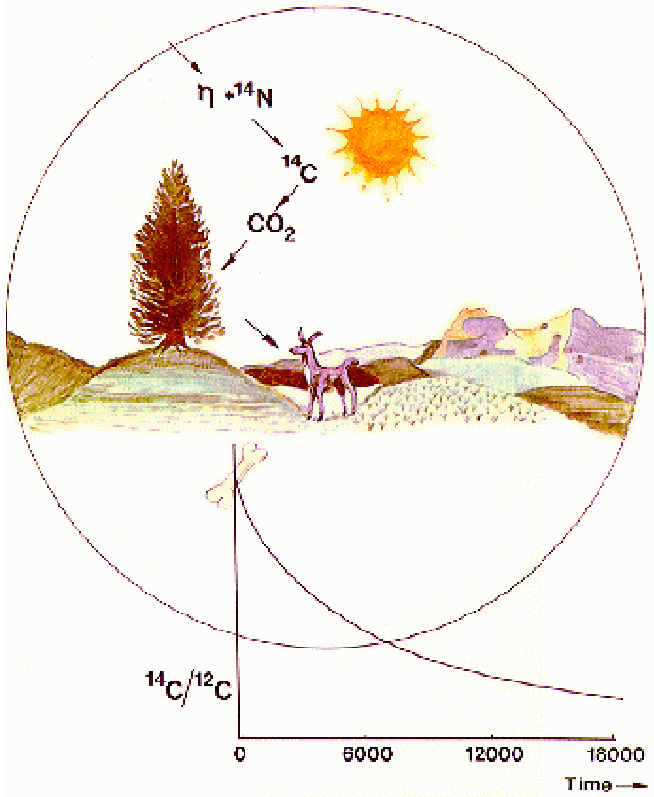


Figure 2.6 A diagram showing the production and life cycle of radiocarbon from production in the upper atmosphere to dispersal into terrestrial environments and incorporation by organic material. The graph demonstrates that the amount of $^{14}C/^{12}C$ reduces over time at a relatively constant rate for the last 12,000 years (Bronk Ramsey, 2009).

Photosynthesis takes ^{14}C atoms directly from the air and converts them back to ^{14}N through the process of radioactive decay. Then, ^{14}N atoms are converted to CO_2, which is a key requirement for photosynthesis (Wood, 2015). ^{14}C atoms are radioactive and have a half-life, which is the length of time it takes 50% of the nuclei to decay (Wood, 2015), and this is measured to interpret radiocarbon ages (Bronk Ramsey, 2008). The Libby half-life is defined as 5,568 ± 30 years (Wood, 2015; Stuvier and Suess, 1966), the Cambridge half-life is given as 5,730 ± 40 years (Godwin, 1962), and most recently, the half-life has been defined as being as long as 6,000 years (Chui *et al.,* 2007). So, radiocarbon ages interpreted from organic material are based on the date associated with the photosynthesis of green plants and the time the material was alive (Wood, 2015), not the date organic material died or was deposited.

After photosynthesis, radiocarbon passes into the food chain by being eaten by an animal or stays within the green plant until the organism's death (Bronk Ramsey, 2008; Curry and Schmidt, 2007). In the simplest form, photosynthesis in green plants and subsequently the start of ^{14}C decay occurs the same year or within a few years of consumption by animals (Bronk Ramsey, 2008). The transfer of radiocarbon in this way results in living organisms being in equilibrium with the same level of atmospheric radiocarbon as the time when an organism was alive (Blackwell and Buck, 2008). In more complex transfers, radiocarbon may pass through several different animals' post-photosynthesis, which means that the radiocarbon signature is passed on, and this could be over a number of years after ^{14}C decay has started (Bronk Ramsey, 2008). In some environments, for example in peat bogs, carbon is not taken directly from the air, but is taken from carbon reservoirs where carbon can be stored for hundreds of years (Wood, 2015). Carbon sequestration is the process of carbon being removed from the atmosphere and stored in a reservoir. A carbon reservoir stores large amounts of carbon that is produced during the carbon cycle; the largest carbon reservoirs are the oceans and other main forms of reservoirs include rivers and lakes (Bronk Ramsey, 2008). When carbon is taken from a carbon reservoir, a reservoir effect must be reported to incorporate an offset in radiocarbon age and minimise the possibility if age overestimations (Blaauw *et al.,* 2004). Blaauw et al. (2004) highlights when multiple samples of peat from different environments are measured, the reservoir effect may be different or not present in all location.

Carbon-bearing organic material used in radiocarbon dating is widely used to interpret the age of events across a wide variety of disciplines, for example archaeology, environmental monitoring, palaeoenvironment studies (Wood, 2015) and palaeoflood hydrology (Jones *et al.,* 2015). The preservation of organic material is dependent upon local environmental and climatic conditions, for example favourable preservation conditions are either very wet locations, for example peat bogs, or very dry environments, and these are characterised by wood and plant remains (see Table 1 in Bronk Ramsey, 2008 for a description of the main types of stable carbon and the associated sample material they are found in). In non-favourable conditions, the carbon content of an organism will convert from stable molecular form to small mobile molecules, such as humic acids (Bronk Ramsey, 2008). The radiocarbon age determined from organisms that have not been well preserved do not contain the isotopic signal as the time when photosynthesis originally occurred so should not be used to determine specific events (Bronk Ramsey, 2008).

2.2.1.2 Methods used for radiocarbon dating

Organic material that is used for radiocarbon dating must undergo pre-treatment before being dated to remove contamination (Wood, 2015; Bowman, 1990). When dead organic material is deposited on the ground it takes time to become fully covered underground within a sedimentary unit. During this process the organic material could come into contact with other sources of carbon, which could disrupt and alter the original radiocarbon signal, and therefore not provide a representable age of the sample (Wood, 2015). Contamination of organisms to exogenous sources of carbon affects organisms differently depending on the age (Wood, 2015). For example, if contamination occurs then the radiocarbon date could be overestimated (Wood, 2015).

Pre-treatment methods typically include physical and chemical processes that are applied using a three-stage process. For example, acid washes to eliminate carbonates, a base to get rid of humic acids, and another acid wash to remove dissolved carbon dioxide. This is often referred to as the ABA or AAA treatment in geochronological studies (Wood, 2015, Rodríguez-Rey *et al.,* 2015). Brock et al. (2010) provides a description of pre-treatments methods that are used for different sample materials. There is an international standard of practice applied by different radiocarbon laboratories across the world to maintain and deliver high quality radiocarbon dates in archaeological studies (Brock *et al*., 2010). However, there is a lack of consistency in reporting methods used in publications (Wood, 2015; Brock *et al.,* 2010; Bronk Ramsey, 2008), which introduces uncertainty. Some laboratories publicise their pre-treatment methods, for example the Oxford radiocarbon accelerator unit (Brock *et al.,* 2010), but it is not common practice to report pre-treatment methods in publications, so it is the responsibility of the user of the radiocarbon results to state them. When radiocarbon pre-treatment methods are not reported, the integrity of the data used can be questioned or when a robust method of quality control is applied, the data may be excluded from analysis (Rodríguez-Rey *et al.,* 2015).

In fluvial environments common sample materials used to radiocarbon date include wood, peat, charcoal and plant remains (Macklin *et al.,* 2012). The size of sample required for dating depends on the type of material and the technique of radiocarbon dating used. For example, when dating wood, it is common to use the full sample because the dating procedure can cause the sample to reduce to 20% of its original size (Wood, 2015). Whereas, up to 300 milligrams is recommended for peat samples (Beta Analytic, 2019b).

Radiocarbon dating involves measuring isotopic ratios of $^{14}C/^{12}C$, which is challenging because ^{14}C occurs in such low quantities (Bronk Ramsey, 2008). There are two main methods of measuring radiocarbon and these are decay counting and accelerator mass spectrometry (AMS; Wood, 2015; Bronk Ramsey, 2008). Decay counting, which is also referred to as the conventional method of radiocarbon dating, uses liquid scintillation or gas proportional counters to measure the amount of decayed ^{14}C atoms by converting the sample to CO_2 so the proportion of beta particles can be counted (Wood, 2015). There are three main factors to consider during the process: detection efficiency, instrument precision and sample purity (Bronk Ramsey, 2008). AMS is the most recently developed and applied method of radiocarbon dating and it works by directly counting the number of ^{14}C atoms and compares it to the number of ^{13}C and ^{12}C atoms in the sample, which is more efficient than gas proportional counting and increases the detection efficiency (Wood, 2015). Instrument precision involves the laboratory equipment and how it is set up, which determines the level of precision the instrument is capable of recording and this is different across different laboratories (Bronk Ramsey, 2008). The purity of a sample is determined by the preservation of a sample and pre-treatment methods (Wood, 2015). AMS can be carried out on small sample sizes (as small as 1mg, Baker, 2008) but as the sample size decreases, the effect of contamination during preservation and pre-treatment significantly increases (Wood, 2015).

Laboratory uncertainty is a measure of chemical and physical error that accompanies radiocarbon determinations (Scott *et al.,* 2007; Bayliss *et al*., 2004). Typically, when an object is radiocarbon dated, only 1 measurement of the object is taken because of the cost. However, through previous testing, it has been understood that if one sample were measured multiple times, each measurement would be different (Scott *et al.,* 2007). This would ultimately create a spread of potential ages for a given sample. However, it is often unrealistic and uneconomical to take more than one measurement, so this uncertainty needs to be numerically incorporated into analysis of radiocarbon measurements. Laboratory error is aimed to acknowledge this uncertainty, but it is not a reflection of all error or uncertainty associated with a sample (Scott *et al.,* 2007). Many radiocarbon laboratories have adapted the use of International radiocarbon laboratory inter-comparisons to try to reduce the uncertainty and error associated with radiocarbon measurements and to enable transparency of methods (Scott *et al.,* 2007).

2.2.1.3 Radiocarbon calibration

The amount of radiocarbon produced in the upper atmosphere or in situ varies over time (Brauer *et al.,* 2014; Blackwell and Buck, 2008), so in order to interpret a date to a radiocarbon measurement, the ^{14}C date must be calibrated (Bronk Ramsey, 2008). Calibration converts radiocarbon dates so that they are given in calendar years, which allows the dates to be translatable, so that they can be compared with other proxy records to infer palaeoclimatic and/ or palaeoenvironmental records (Bartlein *et al.,* 1995). Calibration is completed by using a calibration curve (Bronk Ramsey, 2008). This study refers solely to the Northern hemisphere terrestrial calibration curve constructed by the International Working Group IntCal13 (Reimer *et al.*, 2013), although there are other calibration curves available, such as CalPal and Fairbanks 0107 (Wood, 2015). The IntCal13 curve is updated every few years (Reimer *et al*., 2013) and was preceded by IntCal09 (Reimer *et al*., 2009) and IntCal04 (Reimer *et al.*, 2004). IntCal13 is characterised by showing the radiocarbon determination on the y-axis and the converted age in years cal BP on the x-axis (Figure 2.5; Reimer *et al.,* 2013). A worked example of this is provided in 3.3.1.

The IntCal13 curve is constructed of collated dendrochronologically dated trees for the last 14,000 years and data is available at five-year intervals (Reimer *et al.,* 2013). Older parts of the calibration curve (> 14,000 years) have data at 10 to 20-year intervals (Wood, 2015).

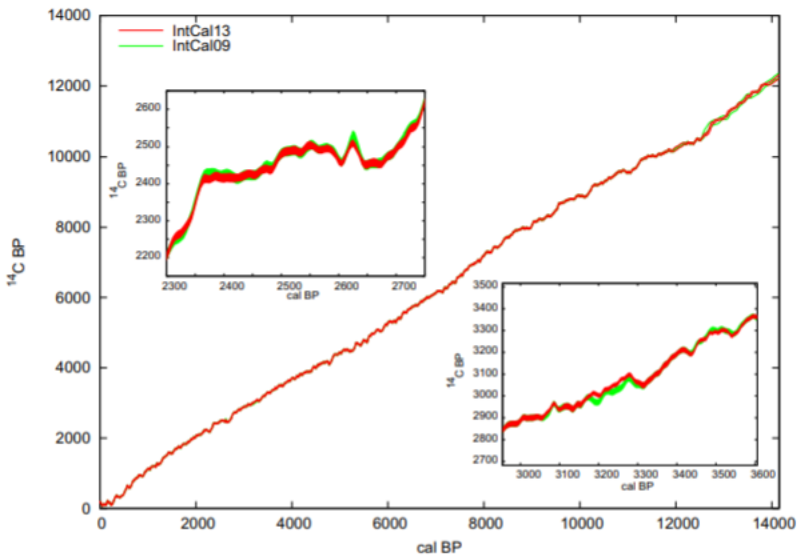


Figure 2.7 IntCal13 curve for terrestrial northern hemisphere samples. The inset images highlight changes to the structure of the curve with the update of data since the construction of Intcal09 calibration curve (Reimer et al., 2013).

The measurements of radiocarbon and associated ages that are used to construct the calibration curve (Intcal13, Reimer *et al.,* 2013) are a collection of data from different sources, for example, tree ring data from different countries in the Northern Hemisphere, including Germany, Ireland, USA (Seattle and Arizona) (Reimer *et al.,* 2004). Uncertainty associated with the data is present within the curve and the main causes are differences in the sampling method. For example, the use of a saw can destroy some tree rings and therefore cause error in the data as well as sample pre-treatment and laboratory operations (Reimer *et al.*, 2004). For the last 14,000 years the calibration curve is constructed of tree ring data, which consists of known absolute dates, calculated by counting tree rings, and radiocarbon measurements and unknown dates and measurements (Reimer *et al.,* 2004; Bronk Ramsey *et al.,* 2001). The different records are put together by using the known absolute age points and fitting the unknown ages of data to them using a Bayesian approach, which applies a Markov chain Monte Carlo sampling methodology using a Metropolis-Gibbs sampler (Niu *et al.,* 2013). For a full summary of this see Reimer et al. (2013), Niu et al. (2013) and Heaton et al. (2009). The statistical reliability of the fit of the datasets can be calculated by using one of the following methods: the X^2 test, Monte Carlo wiggle matching, and a Bayesian approach (Bronk Ramsey *et al.*, 2001). The effect of natural variations in the production of radiocarbon, causes the calibration curve to have fluctuations, which are present in the form of steps and plateaus (Williams, 2012; Bartlein *et al.,* 1995) and span up to several hundred years (Williams, 2012).

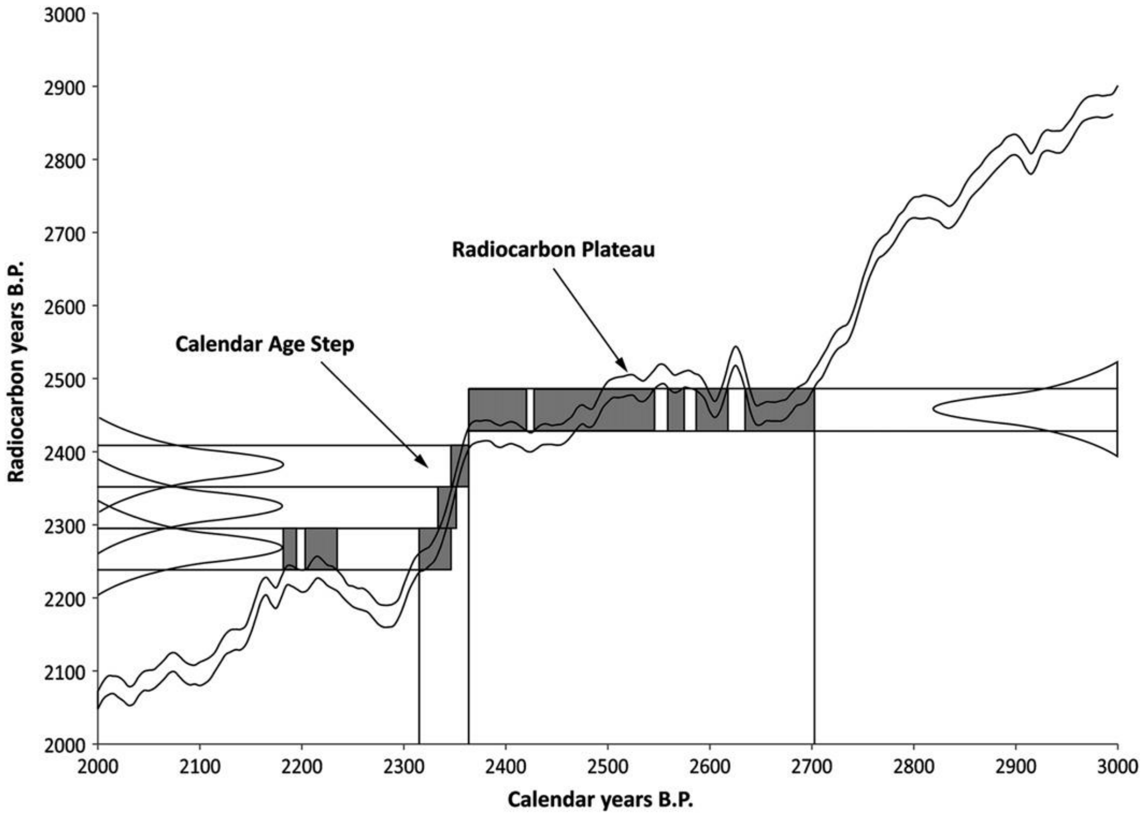


Figure 2.8 The identification of a radiocarbon step and plateau on IntCal09 calibration curve between 2,000 – 3,000 cal years BP (Williams, 2012).

Radiocarbon steps are steep sections of the radiocarbon calibration curve and cause calibrated radiocarbon periods spanning hundreds of years to be attributed to a single radiocarbon determination (see Figure 2.6; Williams, 2012). Radiocarbon plateaus relate to periods of time when the radiocarbon calibration curve remains flat over several hundred years, which results in a single radiocarbon age being converted into a wide range of potential calibrated dates (see Figure 2.6).

The presence of steps and plateaus in the calibration curve causes individual ^{14}C ages to be deformed when they are calibrated, and this is referred to as the 'suck and smear' effect (Ballie, 1991). 'Suck in' is when an unknown-age and radiocarbon measurement is matched to fit the calibration curve of known ages, and this results in clusters of radiocarbon observations across a variety of context that appear to show a correlation that is wrongly interpreted to explain causation (Ballie, 1991). ‘Smear’ of radiocarbon dates has the opposite effect of the ‘suck in’ of dates and causes the presence of radiocarbon dates at certain times to be smeared and effectively lost because the dates are put together into a period, so individual radiocarbon dates that do reflect the occurrence of an event are lost (Ballie, 1991). The ‘suck and smear’ effect of radiocarbon dates during calibration has a knock-on effect on the statistical analysis and interpretation of calibrated dates (Williams, 2012; Chiverrell *et al.,* 2011; Michczyński and Michczyńska, 2006).

Oxcal (Bronk Ramsey, 2009) is software widely used to calibrate radiocarbon dates (Wood, 2015) and applies a Bayesian approach to address the issues discussed previously (Bronk Ramsey *et al.,* 2001). The method involves matching the intercept of the single radiocarbon age on the y-axis to the calibration curve on the x-axis. There can be more than one interception point and this increases the uncertainty of the result. The resultant age ranges are probability-based (see Figure 2.9) and are presented as single probability distribution functions (Bronk Ramsey, 2009). Calibrated ages are given in Oxcal at 68% and 95% confidence intervals (Bronk Ramsey, 2009; Michczynska and Pazdur, 2004). Radiocarbon dates are therefore estimated ranges based on probability distributions around the true age of a sample (see Figure 2.9; Wood, 2015).

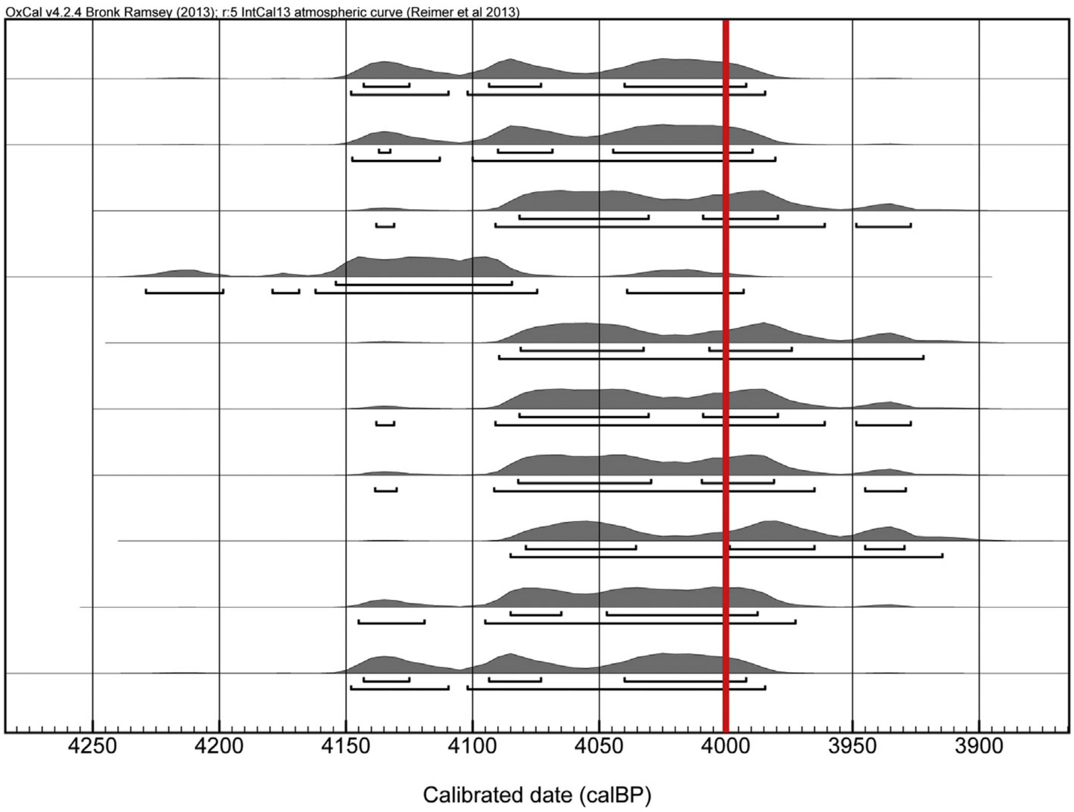


Figure 2.9 10 individual probability distributions of the ^{14}C determination 4,000 years and error ± 25 years show time ranges of individual dates can be overestimated (Wood, 2015).

2.2.1.4 Reporting of radiocarbon dates

The growing number of studies that use and apply radiocarbon dates has stimulated discussions for the reporting of radiocarbon dates and Millard (2014), recommends that radiocarbon dates are provided along with:

- Conventional ^{14}C age
- Conventional ^{14}C measurement and associated error
- Laboratory number where testing was carried out
- Pre-treatment methods
- Radiocarbon technique used
- Quality assurance/ quality control measures used

- Calibration curve used
- Software used to calibrate
- Calibrated age ranges
- Define whether dates are given in years BP or BC/AD
- Information of any modelling used on the dates

The recommendations above aim to increase the reliability of interpretations made using radiocarbon dating techniques and to allow the development of databases for meta-analysis, such as the British database, and to promote good practise of scientific reporting (Millard, 2014). Furthermore, the full reporting of radiocarbon dates will allow the dates to be used in new methods that have not been developed yet and be reassessed, for example, using the next updated version of the calibration curve (Bartlein *et al.,* 1995).

2.2.1.5 Quality control

Quality control of radiocarbon dates is aimed to strengthen the quality of interpretations made particularly of large databases of radiocarbon dates (Pettitt *et al.,* 2003). The use of quality control has been applied across different disciplines; for example, the use of Australian land vertebrates to identify and date megafauna extinction events (Rodríguez-Rey *et al.,* 2015) and the use of archaeological material to date human burial events (Pettitt *et al.,* 2003). As previously discussed in 2.2.1 there are several procedures used in radiocarbon dating, for example type of sample material, pre-treatment methods and calibration curves used, that could cause error and unreliability of the context of radiocarbon dates. The context of a radiocarbon date is essential to understanding how likely the date reflects a specific event rather than assuming the two are connected, so the application of quality control criteria increases the chance to identify specific events and reliably infer causal mechanisms (Rodríguez-Rey *et al.,* 2015).

The formal reporting of radiocarbon dates, as defined above after Millard (2015) is crucial to be able to apply quality control because if certain information about radiocarbon dates is not available quality control cannot be applied and cannot be used in future reanalysis as radiocarbon techniques develop. Table 2.3 highlights quality control factors that have been discussed and applied in previous studies.

Table 2.3 A summary of factors included in previous applications of quality control of radiocarbon dates.

Quality control criteria	**Definition**	**Reference**
Estimated ^{14}C date	Estimated age of sample including error bounds	Rodríguez-Rey *et al*. (2015) Jones *et al*. (2015) Pettitt *et al*. (2003)
Type of sample material	E.G. wood and charcoal	Rodríguez-Rey *et al*. (2015) Jones *et al*. (2015) Pettitt *et al*. (2003)
Direct ^{14}C dates	Ages taken from species/ material of targeted species/ event	Rodríguez-Rey *et al*. (2015), Wood (2015) Pettitt *et a*l. (2003)
Indirect ^{14}C dates	Ages not taken from target species/ event but could be used to date through association of material to the event	Rodríguez-Rey *et al*. (2015), Wood (2015) Pettitt *et al*. (2003)
Stratigraphic association	Stratigraphic relationship between sample material and event; if they occur in the same location, both are inferred to have the same age	Rodríguez-Rey *et al*. (2015) Jones *et al*. (2015) Pettitt *et al*. (2003)
Depositional context	Physical environment of sample material	Rodríguez-Rey *et al*. (2015) Jones *et al*. (2015)
Reworking of sample material within depositional setting	Displacement/ disintegration of sample material; could be caused by human activity, natural processes, erosion	Rodríguez-Rey *et al*. (2015) Pettitt *et al*. (2003)
Stratigraphic reliability	Preservation of undisturbed sedimentary layers	Rodríguez-Rey *et al.* (2015)
Sample pre-treatment	Method of laboratory pre-treatment methods used	Rodríguez-Rey *et al.* (2015) Pettitt *et al*. (2003)

The examples given in Table 2.3 are not exhaustive and have been generalised from their respective fields of study to be applicable across different disciplines. Rodríguez-Rey et al. (2015) applies a 'decision tree' to radiocarbon dates based on the criteria in Table 2.3 and further interrogations of each criterion to score radiocarbon dates, see Figure 2.8. Whereas, in other studies a simple inclusion/ exclusion protocol is applied instead of a scoring system, for example the British fluvial database has a robust classification and identification protocol but only two of these factors are considered when analysing and interpreting the dates (Jones *et al.,* 2015; Macklin *et al.,* 2012).

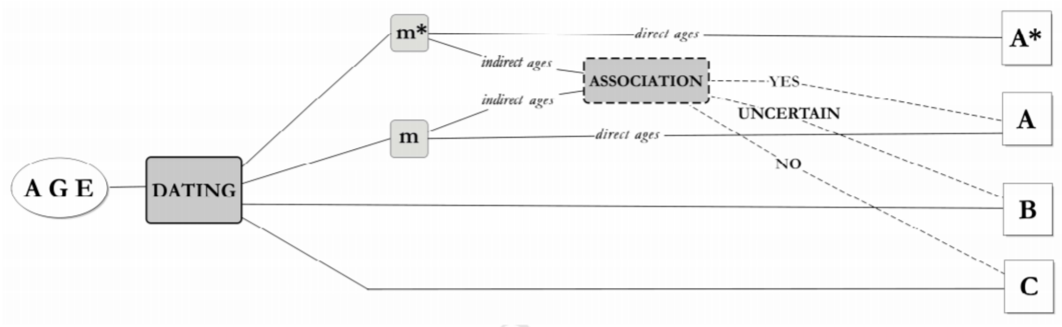


Figure 2.10 Decision tree used to score the quality of radiocarbon dates (Rodríguez-Rey et al., 2015).

The benefit of using a scoring system is that interpretations are justified by reliable data, although this can result in the exclusion of 70% of the data that fails to pass the quality control criteria (Rodríguez-Rey *et al.,* 2015). In the British database, 236 dates pass the two-stage protocol, which also leads to a loss of 70% of the data, so if a more robust scoring system were applied there may not be enough data to carry out any viable analysis and make interpretations. In some cases, the problem is that information about radiocarbon dates are reported but not incorporated into analysis. For example, Macklin and Lewin (2003) identified a classification system based on the stratigraphic location of sample material using 'change after' dates, which were discussed in 2.1.3.

When radiocarbon dates are collated into a database it is essential to distinguish the context of material because if this is not done, this could worsen the issue with ‘sucking’ and ‘smearing’ of the calibration curve, which was discussed in 2.2.1.3 and contribute to the unrealistic correlations between radiocarbon dates in different contexts (Bartlein *et al*., 1995).

The full reporting of radiocarbon dates, as suggested by Millard (2014), and the application of robust quality control procedures could enable a higher level of confidence in the use of radiocarbon dates across many different disciplines. Furthermore, as well as full reporting of radiocarbon dates and quality control criteria, it should also be clearly stated if any assumptions are made about data because this also effects the reliability of interpretation of events and their potential causes (Pettitt *et al.,* 2003). Without a formal protocol in place, the reliability of conclusions made by interpreting radiocarbon dates based on unjustified assumptions is questionable.

2.3 Methods of data analysis for groups of radiocarbon dates

This will focus on the development of methods used to analyse collections of radiocarbon dates and discuss their respective benefits and limitations in relation to representing flood frequency records.

2.3.1 Histograms

2.3.1.1 Overview

In the field of palaeoflood hydrology, groups of radiocarbon dates were first interpreted using histograms of uncalibrated radiocarbon dates (e.g. Figure 2.9; Macklin and Lewin, 2003). The use of uncalibrated dates allows the individual radiocarbon determinations to be represented but if histograms were to be remade using calibrated dates, an average value of the range of calibrated dates would need to be used, such as the mean or the median value. This would allow comparison to other records on a calendric timescale but could distort the representation of the true sample age. Often the mean is used but the median has shown to better represent the spread around the true age value (Telford *et al.*, 2004).

2.3.1.2 Histogram construction

Histograms require a class interval, or bin-width, and are constructed by counting the frequency of data that fall within each interval. They are then plotted typically with the class intervals on the x-axis and frequency on the y-axis (see Figure 2.9 below). The number of radiocarbon dates required to construct a reliable histogram was determined to be at least 40 ages per 1,000 radiocarbon years (Stolk *et al.,* 1994; Geyh, 1980). The optimal bin-width is defined as having a minimum of 25 dates per bin-width of two standard deviations (Geyh, 1980). Histograms are interpreted based on the patterns of the heights of peaks and also to identify and comment on trends of peaks, for example, clusters of data during a given time period. Stacked histograms are useful to classify different characteristics of data within the same class interval. For example, McFadget et al. (1994) used stacked histograms to represent different sample material within each class interval to determine the offset between material and age.

Histograms of calibrated radiocarbon dates can be generated using one of two methods; the KORHIS method or the CALHIS method (Stolk *et al.,* 1994). The KORHIS method applies a correction Equation *(f)* to an uncalibrated histogram: $f=dy/dx$, *w*here, dy = ^{14}C year and dx = calendar year. The Equation is applied at 5-year intervals by multiplying an uncalibrated histogram with the correction Equation. For the full methodology see Stolk et al. (1989). The CALHIS method can be carried out using the Groningen programme and is based on the selection of a smoothed calibration curve (Stolk *et al*., 1994). It must be recognised that this method is now out of date with the development of the IntCal curves. The smoothing of the curve is carried out within the programme using 12 curves that vary from high-resolution, where wiggles, which show the variation in ^{14}C, are prominent, or low-resolution, where wiggles are smoothed out completely (Stolk *et al.*, 1994). The CALHIS method is also based on the generation of a histogram of uncalibrated radiocarbon dates first, which is completed by normalising the Gaussian distribution of each individual ^{14}C age to the area of a standard Gaussian distribution. Then, a calibrated histogram is created by also normalising the x-axis from ^{14}C years to a Gaussian distribution and calibrated years and summing each individual distribution. Stolk et al. (1994) determine that the CALHIS method provides more reliable results than the KORHIS method, but also states that it is not possible to eliminate the effect of the calibration curve on the shape of the histogram.

2.3.1.3 Application and interpretation

The interpretation of radiocarbon dates by histograms is based on the assumptions that: the dates being used in analysis are related to the occurrence of the event being studied; peaks and gaps are likely to be believed if other histograms show similar patterns for the same time period using similar data; and the shape of the histogram is not affected by radiocarbon calibration (Geyh, 1980). Geyh (1980) presents a classification of the reliability of histograms based on known uncertainties that are still applied today; these are defined as reliable, common and unreliable histograms with respective associated statistical uncertainty of less than 20%, between 50-20% and more than 50% (Michczyńska and Pazdur, 2004).

The British database was initially represented by histograms with class interval sizes of 400 years. In initial analysis, there were 123 samples using different geochronological techniques were included in histogram construction (see Figure 2.9). Figure 2.9 shows how a stacked histogram characterises the forms of data within each class interval. In relation to palaeoflood hydrology, histograms were used to identify periods of river alluviation and fluvial episodes to correlate the relationship between flood frequency and climate change (Macklin and Lewin, 1993).

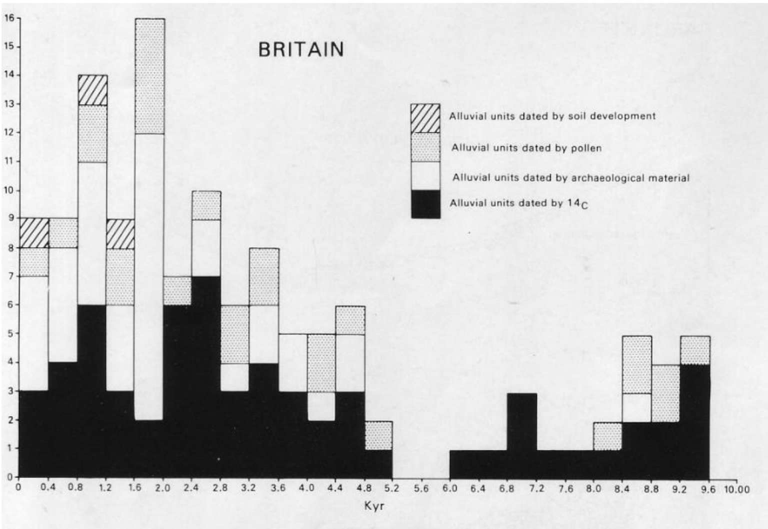


Figure 2.11 A histogram of alluvial units from British fluvial environments (Macklin and Lewin, 1993).

2.3.1.4 Limitations

Fluvial deposits are unique and each date or set of dates from a single site represent a different form of fluvial activity with different casual mechanisms. The application of using histograms to represent groups of radiocarbon dates does not allow for this information to be captured; for example, within the British database most available radiocarbon dates have no stratigraphic connection with each other and so leads to the question of whether they should be analysed and interpreted collectively (Chiverrell *et al.,* 2011).

Furthermore, a priori information is not always available to be able to make links or stratigraphic controls (Buck *et al.*, 1994). Bearing these issues in mind, Buck et al. (1994) developed a Bayesian methodology, which are explored in 2.3.3, to apply a single methodology to a 'global' set of dates, i.e. dates with no stratigraphic links. The Bayesian approach identifies outliers within a global set of dates so that they can be excluded from data analysis. Buck et al. (1994) applied this technique to archaeological data and certainly has potential to be applied to the recently updated British database.

The use of histograms to represent radiocarbon dates as a tool of analysis is limited for several reasons. Firstly, when data is either not available or absent within a class interval, there is no interpolation; so, the lack of data is represented by a clear gap. It is then up to the interpreter to determine the most likely cause of the missing data as histograms are not explanatory (Geyh, 1980). When examining palaeoflood data, gaps in the record are especially difficult to interpret (Chiverrell *et al.,* 2011) there may be many causes for gaps. For example, increased fluvial activity did not occur during a specific time interval or increased fluvial activity did occur and eroded evidence of preceding activity away. Furthermore, histograms are not capable of identifying individual events or the magnitude of a group of events; histograms only show the frequency of data that is available. Secondly, the definition of class intervals can greatly affect the representation and interpretation of data. For example, in Figure 2.12 below the same data are shown using different class intervals to represent annual discharge data. In Figure 2.12 (top), there are no gaps in the spread of data and a general increase followed by a decrease in discharge is observed. But when the class interval sizes are halved (in Figure 2.11, bottom), a more detailed interpretation can be made based on smaller time steps, for example the identification of a gap in the data and fluctuations on a smaller scale.

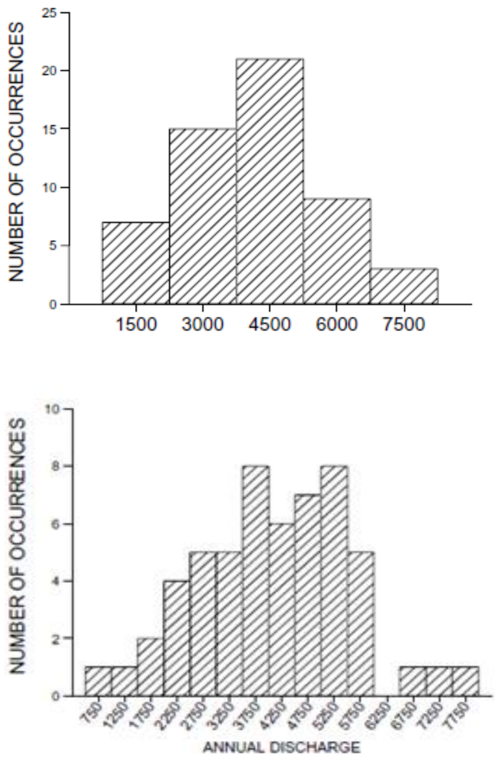


Figure 2.12 (top) Histogram showing the effect of using a large bin size (1,500 units) (bottom) compared with smaller units (750 units) (Helsel and Hirsch, 2002).

Thirdly, the calibration curve can affect the shape of histograms, which is referred to as the calibration stochastic distortion (CSD) effect (McFadget *et al.,* 1994). This is mainly a problem when analysing global datasets because the CSD effect can cause artificial peaks or gaps within a histogram, due to steps and plateaus in the radiocarbon calibration curve (2.2.1.3; Williams, 2012). To address this issue, McFadget et al. (1994) developed a correction method based on the standard deviation and the shape of the calibration curve to account for any influence.

As previously discussed the use of calibrated dates may be preferable, though Stolk et al., (1994) advises that two histograms are used; the first of uncalibrated radiocarbon ages and the second of calibrated radiocarbon ages.

2.3.2 Probabilistic approaches

2.3.2.1 Overview

In palaeoflood hydrology, the application of probabilistic techniques has superseded the use of histograms to represent large groups of radiocarbon dates in many studies (Jones *et al.*, 2015; Macklin *et al.*, 2010; Michczyńska *et al.*, 2007). There are three main probabilistic techniques that have been applied to analyse radiocarbon dates: probability distribution functions (PDFs), summed PDFs; and relative probability plots. It is important to acknowledge that the true age of a sample is unknown, which means that single events cannot be identified using groups of radiocarbon dates (Chiverrell *et al.,* 2011). However, it has been argued that the application of summed PDFs to fluvial data is not to identify individual events but periods of increased flooding (Jones *et al.,* 2015; Macklin *et al.,* 2011).

Individual PDFs are used during the radiocarbon calibration process to convert individual radiocarbon ages into radiocarbon dates on a calendric timescale (as discussed in 2.2.1.3). When more than one radiocarbon age is calibrated simultaneously, individual PDFs are summed to create a single summed PDF (Michczyńska and Padzur, 2004). Summed PDFs are used to represent and interpret collections of radiocarbon ages in the field of palaeoflood hydrology

(Macklin *et al.*, 2012; Michczyńska *et al.*, 2007; Thorndycraft and Benito, 2006; Michczyńska and Padzur, 2004) to interpret flood frequency records.

More recently, Jones et al. (2015) and Macklin et al. (2012) have adapted the summed PDF approach to construct relative probability plots, which have an extra step in the methodology to address issues of the calibration curve, as discussed in 2.2.1.3. Relative probability plots are constructed by dividing the probability at 5-year intervals of sub-datasets by the corresponding probability value for the full dataset (Jones *et al.,* 2015). For example, the analysis of 'change after' dates that were interpreted by a summed PDF (Macklin *et al.,* 2010), are now interpreted by a relative probability curve (Macklin *et al.,* 2012).

Each type of PDF is interpreted in the same way in that peaks are interpreted to show an increase, for example an increase in fluvial activity (Macklin *et al.,* 2012), and gaps or troughs in the curve are interpreted to show decreases in activity (Williams, 2012; Michczyńska and Padzur, 2004).

2.3.2.2 Curve construction

PDFs of radiocarbon dates are generated using modelling software during radiocarbon calibration (for example Oxcal; Bronk Ramsey, 2009). Individual radiocarbon ages are transformed into a range of potential ages characterised by two variables; time in years calibrated BC/AD (at 5-year intervals) and probability (Bronk Ramsey, 2009). The output data generated by calibration of multiple radiocarbon ages can be copied and pasted into a spreadsheet programme, such as Microsoft Excel, to create a curve using the graphs function (see Supplementary information, Macklin *et al.,* 2012). Relative probability plots are created by selecting a sub-dataset of data and dividing the output data with the output data for the entire dataset (Jones *et al.*, 2015; Macklin *et al.*, 2012). A step-by-step guide of how to construct each curve is given in 3.3.

2.3.2.3 Application and interpretation of probabilistic techniques

The application of using summed PDFs or relative probability plots to analyse ^{14}C dates is most common in European studies, where fluvial sediments are the dominant form of data. Three examples are provided in Figures 2.11 - 2.13 that show the application and interpretation of summed PDFs and relative probability plots to represent the flood frequency record. Firstly, 'change after' dates in the British database are analysed using a relative probability plot (as shown in Figure 2.11). The shape of probability curves are characterised by peaks and troughs, and the relative mean probability is annotated by a horizontal line. Periods of flooding are interpreted when the curve exceeds the mean probability (Jones *et al.,* 2015; Macklin *et al.,* 2012). The interpretation of peaks, gaps and troughs of probability curves is debated in the literature and is discussed in 2.3.2.4.

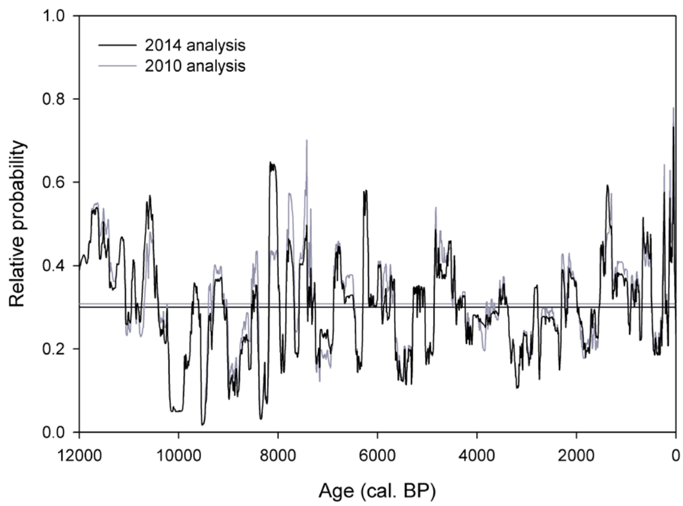


Figure 2.13 Relative probability plot for 'change after' dates from the British database. 2014 analysis reflects 252 dates and 2010 analysis reflects 236 dates. The solid horizontal lines signify the relative mean probabilities of each analysis (Jones et al., 2015).

A second example is the use of summed PDFs to represent ^{14}C dated palaeoflood data in Poland (Starkel *et al.,* 2006). Starkel et al. (2006) applied an adapted methodology for data selection: each sample was assigned to one of eight categories, which were indicators that the sample had a connection to a flood event, for example samples taken from the base of palaeochannels and samples from the top of overbank sedimentary units. Sub-datasets, which were created based on the eight criteria, were used to create summed PDFs. In total, 17 periods of increased fluvial activity are identified. Unlike the technique used in the first example of applying a mean probability, all peaks are interpreted to reflect changes in fluvial activity (Starkel *et al.,* 2006).

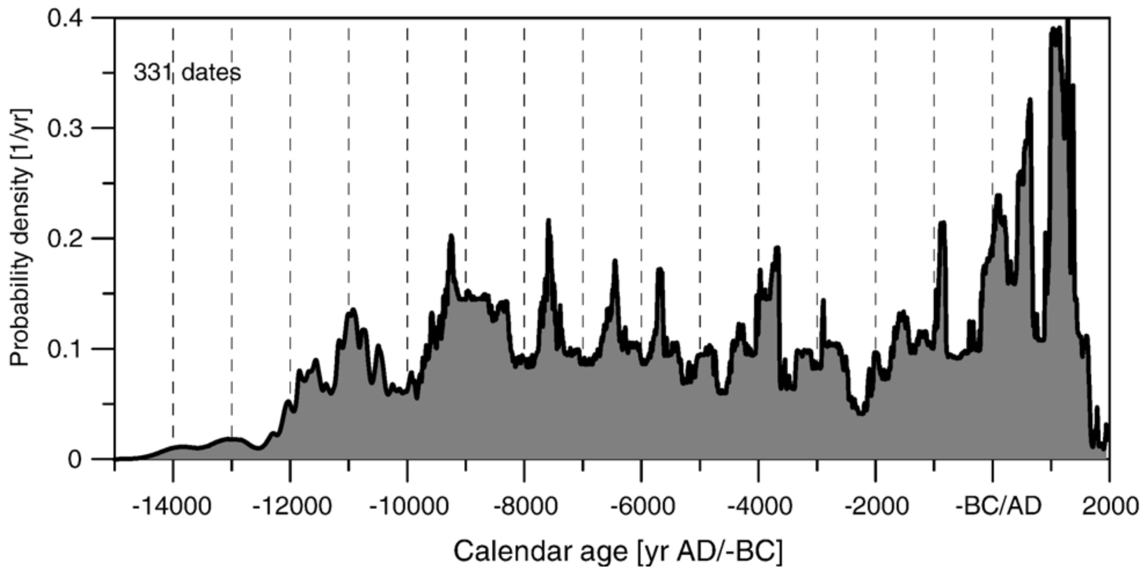


Figure 2.14 Summed PDF, also referred to as probability density curve, of fluvial deposits from Poland that have an association to a flood event (Starkel et al., 2006).

A third example is of the application of the Spanish fluvial record to create flood-frequency records for the Holocene (Benito *et al.,* 2008; Thorndycraft and Benito, 2006). Benito et al. (2008) used summed PDFs to analyse radiocarbon dated floodplain sediments alongside radiocarbon dated SWDs (as shown in Figure 2.13).

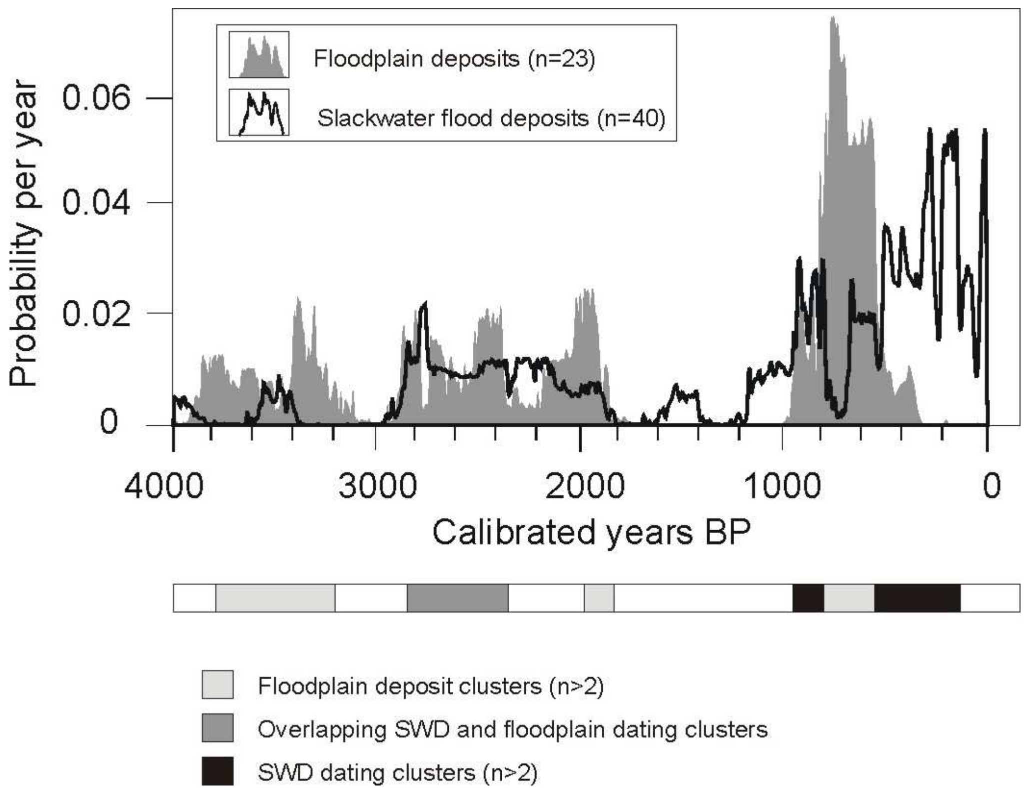


Figure 2.15 Summed PDF of Spanish palaeoflood data (Benito et al., 2008).

Despite the low number of samples (63 in total), three periods of increased fluvial activity are identified when the occurrence of SWDs coincide with clusters of radiocarbon dates (Benito *et al.,* 2008).

2.3.2.4 Comparison of summed PDFs to proxies

Each of the examples above also use probability curves generated using palaeoflood data to correlate with other records to infer human activity and/ or climate change. Holocene British floodplain sediments have been used to reflect anthropogenic impacts in relation to agricultural practices (Macklin *et al.*, 2010). A record of sedimentation of floodplain deposits was reconstructed by analysing radiocarbon dates using summed PDFs (Macklin *et al.,* 2010). Regional and national plots were produced and used to identify sedimentation rates by using information on sample depth and the midpoint of the ^{14}C calibrated date. An increase in sedimentation rates was observed in the regional and national reconstructions at around 1,000 years cal BP (Macklin *et al.,* 2010).

The interpretation of radiocarbon dates and sedimentation rates has been correlated with other data to infer a link between agricultural practices and fluvial activity. Hoffmann et al. (2008) applies a similar technique to identify an increase in sedimentation rates between 1,300 and 820BC in German fluvial deposits. Furthermore, summed PDFs are also used to analyse radiocarbon dates taken from Mediterranean fluvial deposits to infer changes in river aggradation (Hooke, 2006).

Fluvial activity and subsequently patterns in deposition are affected by changes in the climate. Blum and Törnqvist (2000) review fluvial responses to climate change by analysing channel and sediment deposition of fluvial sediments. Other studies have used radiocarbon dates to reconstruct fluvial responses to climate change using probabilistic techniques to 'wiggle-match' peaks and troughs across different records. For example, the British database is used to compare with palaeoclimatic proxies; North Atlantic drift ice record and European Neoglacial chronology (Macklin *et al.,* 2010). Based on the characteristics of each ^{14}C date within the British database, regional sub-datasets based on precipitation regions are compared with proxies of mire surface wetness and water table reconstructions. Macklin et al. (2010) conclude that increased fluvial activity during the Holocene, which is interpreted by peaks in the shape of summed PDFs, correlate with both palaeoclimatic and palaeoenvironmental proxies and that fluvial responses occur immediately after climatic changes.

Spanish fluvial sediments are correlated with pollen records to infer palaeoenvironmental records and the North Atlantic drift ice record to infer palaeoclimate (Thorndycraft and Benito, 2006); see Figure 2.14.

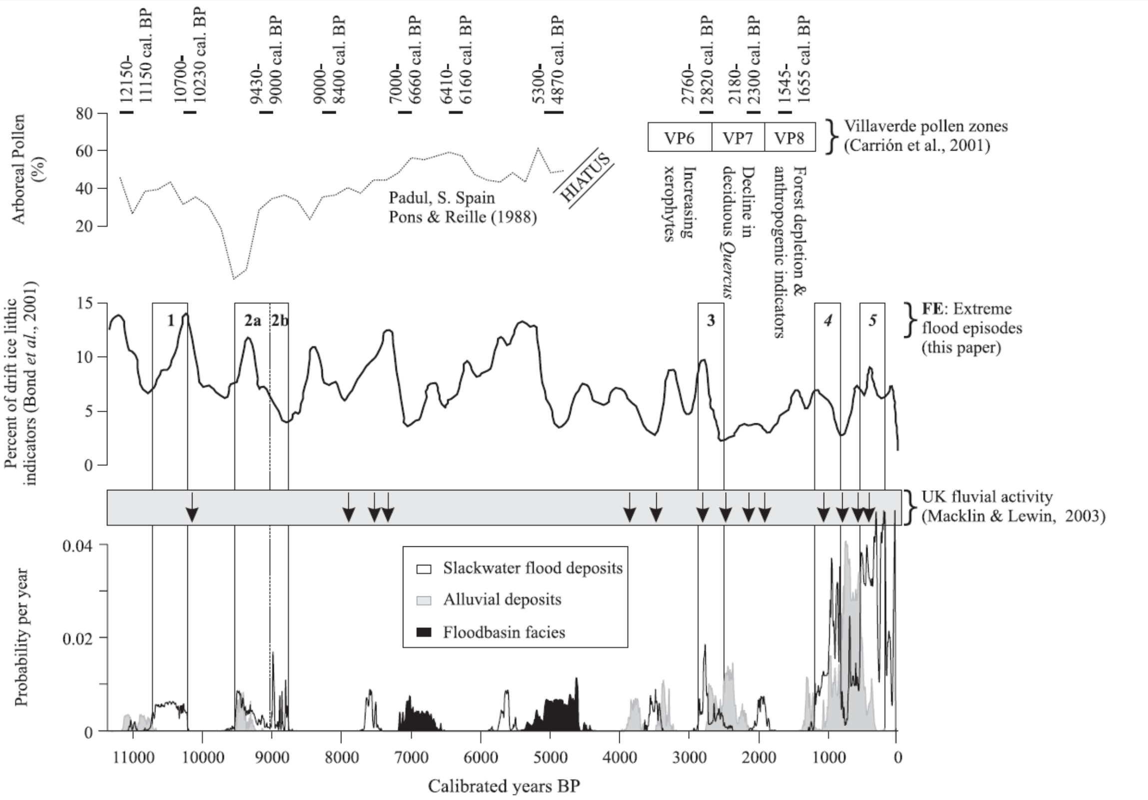


Figure 2.16 A comparison of Spanish palaeoflood data, pollen records and the North Atlantic drift ice record (Thorndycraft and Benito, 2006).

Thorndycraft and Benito (2006) determine that the combined probability of a fluvial deposit being present at specific times generally occur during stable conditions during the Mid-Holocene, which could reflect local palaeoenvironmental stable conditions, for example of slope stability. SWDs were used to compare with the North Atlantic drift ice record as opposed to radiocarbon dated fluvial deposits (Thorndycraft and Benito, 2006).

As well as comparing palaeoflood data with climate proxies, Macklin et al. (2012) compares palaeoflood data from the UK and New Zealand to test whether the anti-phasing of the northern and southern hemisphere observed during the last glaciation is also observed during the Holocene. A similar approach to the previous examples was applied: ‘change after’ dates were included in analysis to create summed PDFs and also relative probability plots. The resulting plots were compared to identify the statistical significance between the records.

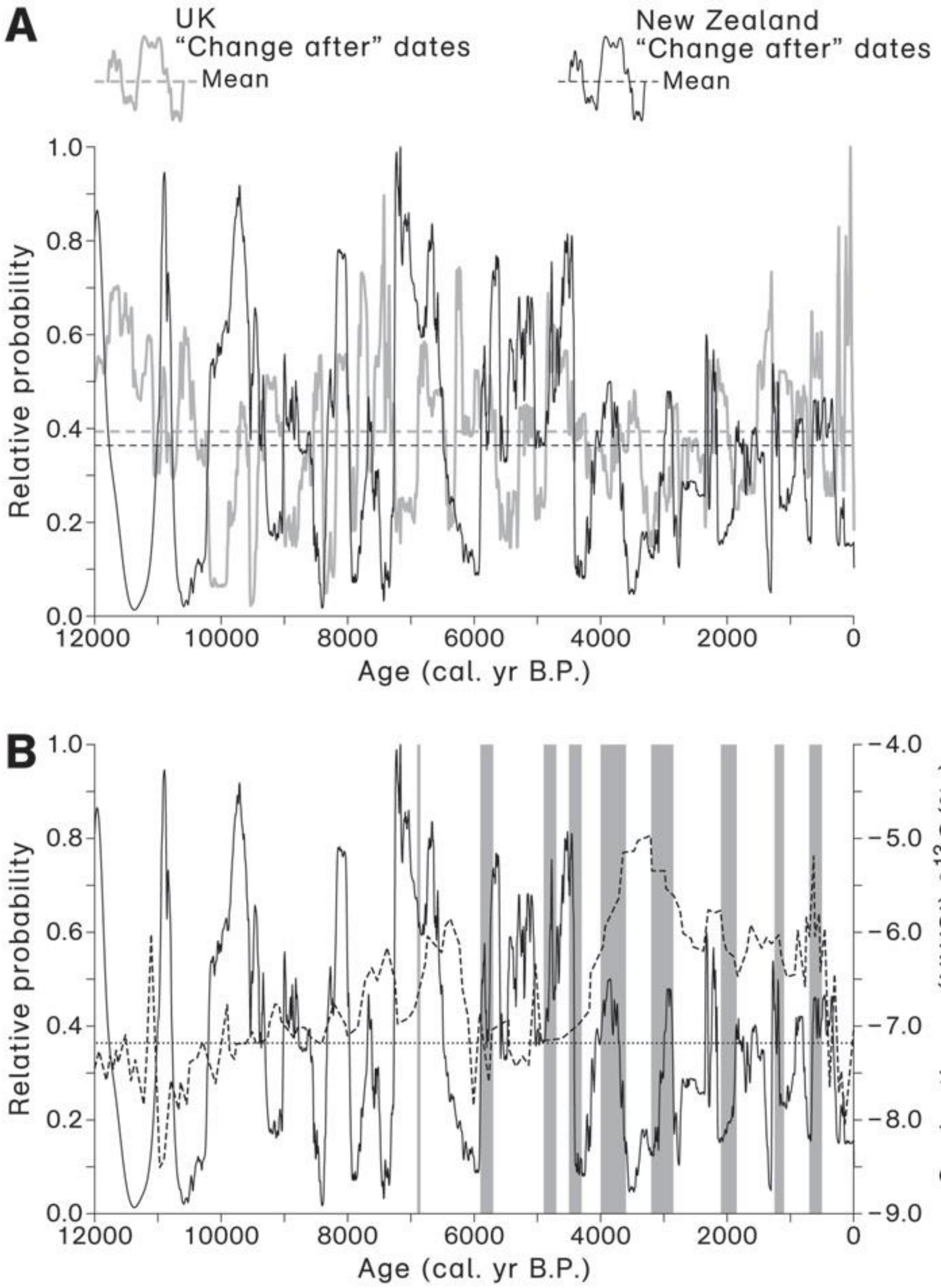


Figure 2.17 Relative probability plots generated using the British database (A) and the New Zealand fluvial record (B). Dotted horizontal lines indicate the relative mean probability. B is also characterised by including lake records (shaded vertical areas) and speleothem records (dashed curve) (Macklin et al., 2012).

From Figure 2.15 it was determined that periods of increased fluvial activity during the Holocene occurred in asynchronous patterns in the UK and New Zealand (Macklin *et al.,* 2012). As a result of the wide application of using probability curves as a tool to analyse and interpret radiocarbon dates, several limitations have been realised.

2.3.2.4 Limitations of applying probabilistic techniques

The use of summed PDFs and relative probability plots to represent radiocarbon dated palaeoflood data has provided a significant development for the field of palaeoflood hydrology as well as for other radiocarbon users; but there are limitations.

Probability curves represent the potential age of radiocarbon dates and do not capture other important information about a sample that could greatly aid interpretation, such as sedimentological context between multiple samples from the same site, sample depth, and site location (Thorndycraft *et al.,* 2011). This is a limitation when meta-analysis of databases is completed. For example, the data within the British fluvial database is made up of different sample materials from different depositional environments from different precipitation regions of the UK. There is also a large range in the size of the catchments. Clumping all of this data together makes it very difficult to interpret the cause behind changes to the shape of the corrected PDF. For example, floodplain deposits are caused by both hydrological conditions and sediment delivery, whereas SWDs are produced solely by single hydrological events (Faust and Wolf, 2017). Faust and Wolf (2017) recommend that data constructed of different depositional environments not be interpreted hastily and that a stratigraphic profile should be made and used to help infer floodplain dynamics. The availability of geomorphological fluvial deposits and the unavailability of other palaeoflood indicators, such as SWDs, sets British fluvial deposits apart from other palaeoflood studies and increases the uncertainty of interpreting the shape of summed PDFs. Chiverrell et al. (2011) have also recommended that regional databases should be put together first to test the validity of the data using a Bayesian approach (discussed in Chapter 2.3.3); and once the data has been corroborated it can be analysed using meta-analysis techniques. However, Jones et al. (2015) argues that Bayesian approaches should be complimentary to the probabilistic approach.

Lewin et al. (2005) discusses how to determine signal from noise with within fluvial data by comparing summed PDFs of each depositional environment within the British database to identify differences between each type of fluvial deposit. The importance of the stratigraphic location of a sample was also highlighted (Lewin *et al.,* 2005), which has been preceded by the classification of 'change after' dates (Macklin *et al.,* 2010). Faust and Wolf (2017) have suggested the need for filtering of data in order to distinguish local patterns from interregional patterns, which also links back to limitations of analysing meta-datasets. The analysis of stratigraphic records alongside statistical analysis of dates (Benito *et al.,* 2008; Chiverrell *et al.,* 2011) is a favourable method of distinguishing signal from noise by interpreting the local environment as well as the radiocarbon dates presented using summed PDFs alone.

There are further limitations associated with probability curve construction. The interpretation of probability curves is based on the assumption that the shape of the curve (peaks, troughs and gaps) reflect genuine changes, however it has been well-documented that the shape of summed PDFs are prone to systematic errors. In particular, the shape of the calibration curve influences the occurrence and extent of peaks and troughs (Williams, 2012; Guilderson *et al.,* 2005; Michczyński and Michczyńska, 2006). The influence of the calibration curve causes some peaks in the shape of probability curves to correspond to steep sections of the calibration curve (identified as radiocarbon steps in 2.2.2.4) rather than genuine occurrences (Faust and Wolf, 2017; Williams, 2012; Chiverrell *et al.,* 2011). Radiocarbon plateaus are related to smoothing over peaks in summed PDFs (Williams, 2012; Guilderson *et al*., 2005).

Another limitation associated with applying summed PDFs is the uncertainty of an appropriate sample number size required to create statistically reliable results (Michczyńska and Padzur, 2004). From the Polish database, Michczyńska and Padzur (2004) initially determined that 200 radiocarbon dates were needed to produce results with fewer than 40% statistical fluctuation and 820 dates to produced results with fewer than 20% statistical fluctuation. Statistical fluctuations are defined by Geyh (1980) when the statistical reliability of histograms was first defined. In a later study, Williams (2012) stated that 200-500 radiocarbon dates were needed and that the shape of PDFs generated using 200 dates would be expected the change with the addition of more data. Williams (2012) also determined that variability in PDFs was greater when dates from the last 500 years were included.

As well as the influence of the radiocarbon calibration curve, peaks and troughs also correspond with the number of data available (Williams, 2012). This leads to the question of: does a trough or gap reflect that no data is present or that no activity occurred? This is an important limitation when interpreting early-Holocene records because the unknown length of time it took for a sample to be preserved/ fossilised after the end of the last glaciation (Lowe and Walker, 2000). Williams

(2012) explores the use of taphonomic curves to correct the uneven temporal distribution of dates and determines that a corrected curve should always be presented alongside the uncorrected curve.

Fluvial datasets have been used previously to reconstruct palaeoenvironmental conditions (Chiverrell *et al.*, 2011; Macklin *et al.*, 2010; Michczyńska *et al.*, 2007). There is a growing demand for fluvial data to be able to increase our understanding of how climate change affects fluvial systems, in particular the frequency of flooding and Jones et al. (2015) argues that the British fluvial database provides a net gain of information. However, considering the limitations discussed above the ability for British Holocene fluvial data to be reliably and accurately used to represent past flood records should be investigated. We need to better understand the factors that influence and control fluvial sequences before we can reliably interpret the data within them collectively.

2.3.3 Bayesian analysis

2.3.3.1 Overview

The methods of data analysis that were discussed in 2.1 and 2.2 are limited in that they only represent radiocarbon dates as statistical values without incorporating supporting contextual information about a given sample (Chiverrell *et al.*, 2011; Thorndycraft *et al.*, 2011). Contextual information includes geomorphic and hydrologic data that could help to understand the context of the deposition and preservation of an organism better. Bayesian analysis allows such incorporation of supporting information into the analysis (Thorndycraft *et al.,* 2011).

2.3.3.2 Bayesian methodology

Bayesian methodologies are based on Bayes' Rule which states: $p(\theta,y)=p(\theta)p(y|\theta)$, where $p(\theta,y)$ is the interpretation to be made about a topic given the observed data and parameters $p(\theta)$ is the prior distribution, and $p(y|\theta)$ is the observed data distribution (Gelman *et al.*, 2014).

Bayesian data analysis is characterised by three main steps: firstly, the use of probability distributions of observable and unobservable data by bringing together the two variables into a single model; secondly, calculation of posterior distribution for unobserved data given the observed data; and thirdly, determining the likelihood of the posterior distribution given all observed data (Gelman *et al.,* 2014). Bayesian data analysis is used across different disciplines, including archaeological studies (Bachand, 2008; Buck *et al.*, 1991), geochronology, for example the generation of IntCal13 (Reimer *et al.*, 2013; Nui *et al.*, 2013) and the Oxcal radiocarbon calibration programme (Bronk Ramsey, 2009), and palaeoflood hydrology (Thorndycraft *et al.*, 2011; O'Connell, 2005).

2.3.3.3 Applications and interpretation of Bayesian analysis

Thorndycraft et al. (2011) applied a Bayesian approach to three river catchments in Spain to identify periods of flooding using contextual information to support each ^{14}C date. The study usedsummed PDFs and utilised other environmental information to produce a more robust flood-frequency record, which are shown in Figure 2.16.

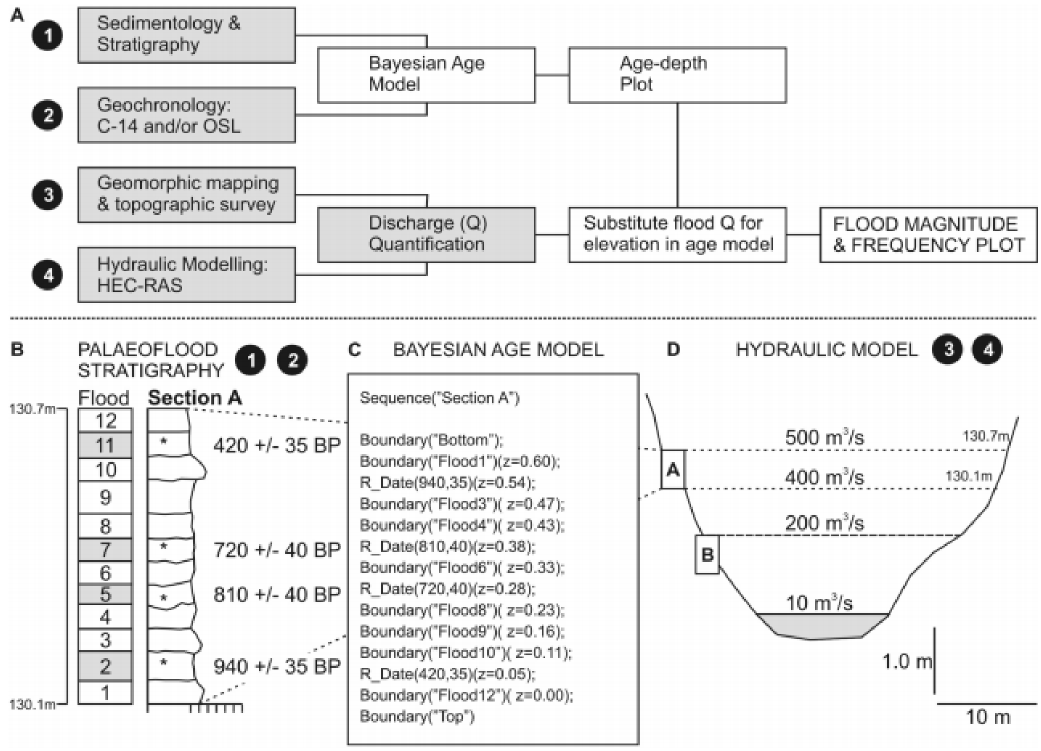


Figure 2.18 Four stage approach to analysing Spanish fluvial deposits (Thorndycraft et al., 2011).

In Figure 2.18, the data used to create the model in step A includes radiocarbon and OSL dates, sediment stratigraphy profiles and hydraulic modelling of a cross-section. Bayesian analysis differs from probabilistic approaches and histograms because multiple ^{14}C ages that have an association to each other are calibrated into ^{14}C dates based on stratigraphic data (as shown in step B and C in Figure 2.18) rather than calculating individual ^{14}C dates and collectively summing them without considering association.

Another example of the application of Bayesian analysis in palaeoflood hydrology is to identify annual exceedance probabilities (O'Connell, 2005). O'Connell (2005) uses historical palaeodischarge data from two rivers in Nevada, USA to identify flood frequency and palaeodischarge records while considering the uncertainties associated with the data and the methods used.

2.3.3.4 Limitations of applying Bayesian analysis

Bayesian analysis is used to collate multiple sources of data, for example multiple ^{14}C ages from one specific site location or one catchment area, to create a palaeoflood stratigraphy (Chiverrell *et al*., 2011). Chiverrell et al. (2011) argued that the methods of analysis used to interpret the British database is questionable because important contextual information is not available and/ or not used if available. Jones et al. (2015) argues that the purpose of applying of Bayesian analysis is to identify individual flood events not periods of flooding and so is not relevant to previous analysis of the British database (Macklin *et al*., 2012). However, a criticism of applying summed PDFs to fluvial deposits has been that individual events cannot be identified (Chiverrell *et al.,* 2011) and that the true age of a sample is lost (Wood, 2015). Therefore, the suitability of applying Bayesian analysis to the British database has not been formally explored but there are lots of examples of where it is used on other palaeoflood data, for example on SWDs (Thorndycraft *et al.,* 2011), but it was identified in 2.1.1 that the British database lacks the presence of SWDs and relies in the interpretation of sedimentary fluvial deposits (Jones *et al.,* 2015). O'Connell (2005) developed a nonparametric method, which differs from parametric Bayesian analysis, because the Bayesian model does not assume the data has a certain distribution. However, the application of parametric methods is a limitation when analysing radiocarbon dates because the real distribution of the data is unknown (Williams, 2012).

2.3.4 Spectral analysis

2.3.4.1 Overview

Methods of data analysis discussed in 2.3.1, 2.3.2 and 2.3.3 have focused on techniques used to identify periods of increased fluvial activity as a function of time. Spectral analysis is used to identify periodicities within datasets as a function of phase (Vanderplas, 2018), typically associated with signal processing or astronomy applications (Schimmel, 2001). Spectral analysis has been used to identify harmonic signal (cyclicities or periodicities) in palaeoclimate time series data (Schultz and Stattegger, 1997). There are several different methods that can be used to identify and characterise periodicities in time series data and the methods can be further refined depending on the characteristics of time series datasets. For example, point observations are a value (with an associated error) at a specific point in time for equally or unequally-spaced datasets; or other types of time series, such as binned event data that are event driven (Vanderplas, 2018). Vanderplas (2018) categorises methods used to analyse point observations into four types: Fourier methods, phase-folding methods, least squares methods, and Bayesian approaches. One specific method of spectral analysis incorporates all four categories: Lomb-Scargle periodogram. The Lomb-Scargle periodogram is the only spectral analysis technique that does not rely on interpolation of data, unlike the statistical methods of probabilistic approaches, which often involve making assumptions about data, when data is unavailable (Swindles *et al.*, 2012). Furthermore, Lomb-Scargle periodogram is the most popular method for analysing unevenly spaced datasets across different fields (Vanderplas, 2018).

Periodicities of changes in the terrestrial palaeoenvironmental record can be more diagnostic of causes than the timings of change because cycles with similar wavelengths across different datasets can be compared (Swindles *et al.*, 2012). Lomb-Scargle periodogram has been applied to peatland records to interpret Holocene palaeoclimate in northern hemisphere terrestrial environments and identifying millennial, centennial and sub-centennial periodicities (Swindles *et al.,* 2012); examples are given below.

2.3.4.2 Lomb-Scargle methodology

Lomb-Scargle periodogram is unique to other methods of data analysis discussed previously because it is designed to be used on unevenly sampled data and incomplete records (Hocke and Kämpfer, 2009), which could be especially useful for palaeoflood data. Data used in spectral analysis must be in the form of two variables, for example x = time and y = probability, and then spectral analysis estimates a function rather than two parameters (x,y). This makes Lomb-Scargle a complex method; see Trauth (2010) for a breakdown of the equations and calculations used to complete Lomb-Scargle spectral analysis.

Trauth (2010) provides a standard script that be used in data analysis software, such as Matlab or R, which has been adapted and modified for this study, and is provided in Appendix A. A Lomb-Scargle periodogram identifies harmonic signals and determines if they are significant by providing probability and frequency outputs (Mudelsee, 2010). In order to evaluate the outputs of Lomb-Scargle spectral analysis, parameters must be defined. Firstly, a rectangular window of the factors wavelength and probability set the area of study (Swindles *et al.,* 2012). Secondly, the probability chosen determines the level when cyclicities are considered statistically significant. For example, if the probability is set at 0.4%, then the false alarm probability (FAP) is 99.6%, this is referred to Siegel's test (Swindles *et al.,* 2012). This means that any frequencies that exceed 0.4 are unlikely to be a random occurrence and hence are significant (Trauth, 2010). Another essential factor to identify is the Nyquist value, which is the highest frequency identified in relation to the data available (Swindles *et al.,* 2012).

The output generated by applying the Lomb-Scargle periodogram comes in two parts: power and probability against frequency (Trauth, 2010). Figure 2.19 below shows an example of a power output with reference to the FAP value.

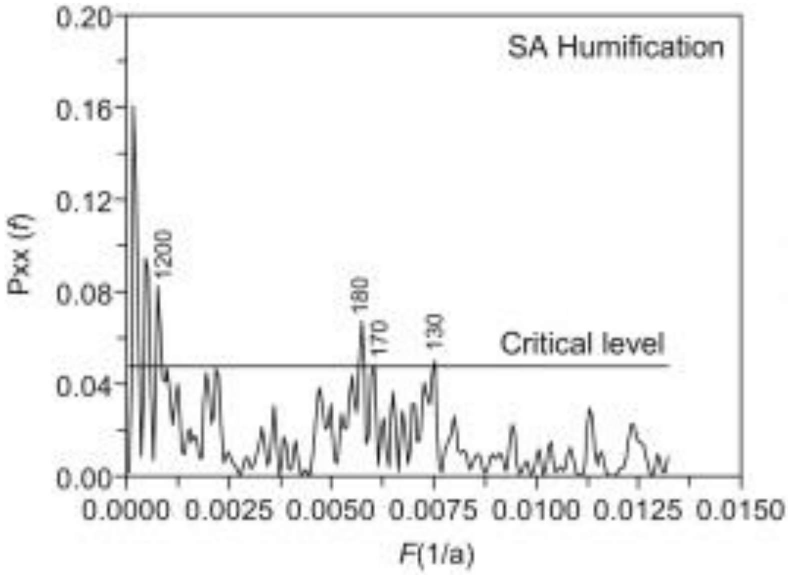


Figure 2.19 Lomb-Scargle spectral analysis power output for peat humification data from Northern Ireland (Swindles et al., 2012).

Figure 2.19 is characterised by peaks in power (y-axis) at given frequencies (x-axis). Peaks that occur above the critical level marked by the horizontal line are considered statistically significant. For example, in Figure 2.19 there are four cyclicities that occur with frequencies: ~1,200 years, 180 years, 170 years and 130 years. In the next section, three examples of the application and interpretations of spectral analysis are given.

2.3.4.3 Applications and interpretations of spectral analysis

Recent examples of the application of spectral analysis to palaeoenvironmental and palaeoclimatic studies use Lomb-Scargle periodogram alongside Bayesian age-depth models to identify cyclicities within datasets (Perez *et al.,* 2018; Swindles *et al.,* 2012). Swindles et al. (2012) applied Lomb-Scargle fourier transform to peat humification data and testate amoebae-derived water table reconstructions to two sites in Northern Ireland. Peat humification data was dated using tephrochronology and radiocarbon dates. Perez et al. (2018) applied Lomb-Scargle periodogram to fluvial data in the form of a sediment core taken from the continental shelf in southwestern South America to determine the link between terrestrial climate and sediment flux patterns over the last 1,000 years. Lomb-Scargle spectral analysis was applied to element ratios from the sediment core. Multidecadal and centennial cycles were identified in each element ratio output and interpreted based on age-depth models and palaeoclimatic proxies (Perez *et al.*, 2018). Each spectral analysis of element ratios was supported by wavelet analysis and age-depth profiles to identify the difference of power of frequency between each record.

Other techniques of spectral analysis have been used to interpolate data when data is unavailable. For example, historical flood records of high and low water marks on buildings along the River Nile were used to reconstruct periodicities of flooding to determine the link between river discharge and climatic influences (Kondrashov *et al.,* 2005).

2.3.4.4 Limitations of applying spectral analysis

There are limitations associated with the methodology and application of spectral analysis to time series data and limitations associated with the interpretation of

Lomb-Scargle periodograms. To the author’s knowledge, Lomb-Scargle spectral analysis has not been applied to large datasets of calibrated radiocarbon dates using the probability values generated and their respective time interval. Therefore, the limitations provided here are given in the context of how the radiocarbon dates could be represented in terms of palaeoflood data, similar to the limitations of methods discussed in 2.3.1, 2.3.2 and 2.3.3. Additional limitations on the technicalities of the method of Lomb-Scargle Periodogram are provided by VanderPlas (2018).

Studies that have applied Lomb-Scargle or another form of frequency analysis do not provide a clear description of the data used to input into analysis, so it is difficult to provide a direct comparison between studies that apply Lomb-Scargle to groups of radiocarbon dates. Only generic descriptions are given, for example peat humification data (Swindles *et al.,* 2012), and not the specific variables used in analysis. It would be useful for other users of frequency analysis to understand how other studies have formatted data for frequency analysis. A common protocol of reporting spectral frequency methodology would also increase confidence in the using the technique, especially for fields that are not familiar with it, to test new methods of analysis, for example in palaeoflood hydrology.

In relation to spectral analysis it is not clear how the number of samples affects the ability of the Lomb-Scargle periodogram to detect certain frequencies. This limitation is hindered by the lack of data description in studies (as discussed above). This is an important factor when considering the interpretation of Lomb-Scargle periodograms. Furthermore, the amount of noise within a dataset also affects the presence of power peaks; see Figure 2.20 below. A key question regarding this issue is: if spectral analysis was to be applied to palaeoflood data, which are known to be noisy (Lewin *et al.*, 2005), should the data be input unfiltered or not? In other words, could ‘change after’ dates (Macklin *et al.*, 2012) as discussed in 2.1.2.2 be used to reliably represent the British database or should the unfiltered British database be used?

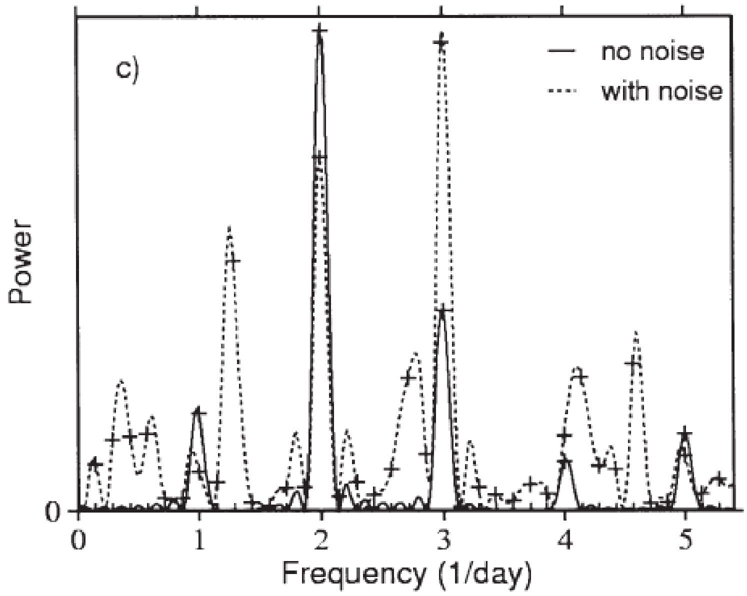


Figure 2.20 Lomb-Scargle periodogram for a dataset with and without noise (Schimmel, 2001).

The calculations behind spectral analysis do not account for chronological errors or outliers within the dataset being tested (Schimmel, 2001). The limitations associated with radiocarbon dates and of evidence used in palaeoflood studies have been discussed earlier in this thesis and demonstrate the variability within palaeoflood data. This means that data must be filtered before spectral analysis is applied if known uncertainties are present.

Lomb-Scargle frequency analysis is characterised by only working on observed data (Ruf, 1999). This means that the method of data interpolation does not occur in spectral analysis (Schimmel, 2001) unlike when probabilistic techniques are applied. This can cause bias in the results similar to histograms.

Another common limitation of applying spectral analysis is the presence of harmonic wavelengths. Harmonic wavelengths are mimics of larger wavelengths (Schimmel, 2001); for example, if a large peak in power had a wavelength of 1000, then a smaller peak in power may occur with a wavelength of 500. The wavelength of 500 would be the harmonic wavelength (Press and Rybicki, 1989). This is why the application of a rectangular window, as discussed in 2.3.4.2 is essential to filter the outputs.

Finally, similar to all the other methods of data analysis discussed here, spectral analysis does not explain causal mechanisms of periodicities. However, the analysis of spectral analysis alongside Bayesian age-depth models could improve the holistic interpretation on a wider context, which is essential when interpreting palaeoflood data.

2.4 Summary of literature review

Palaeoflood hydrology is the study of evidence of past floods that were not observed or recorded to construct flood frequency and flood magnitude records (Baker, 2008). Evidence of past floods occurs in many different types, for example SWDs, large boulders and of interest to this study is the sediment record. The sediment record contains evidence of past floods by preserving patterns of deposition in fluvial environments. It is important to remember that fluvial deposits are not proxies of past floods; they are direct evidence of past fluvial activity (Baker, 2016).

In the UK, the flood record has been extended by the analysis and interpretation of sediment records taken from a range of fluvial depositional environments, such as upland, lowland, flood plains or palaeochannel fills (Macklin *et al.,* 2012). As it is referred to in this study, the British database is a collection of radiocarbon dated organic material. It has been recognised that not all fluvial deposits are indicators of flood events (Lewin *et al.,* 2005). Therefore, the classification of samples based on their stratigraphical location was developed. Macklin et al. (2010) defines that samples that are taken from the base of abrupt changes in patterns of deposition, for example from fine to coarse sediment (Figure 2.3), are classified as 'change after' dates. 'Change after' dates are considered to be the most reliable indicator of the occurrence of flooding from sedimentological records and so samples that are not identified as 'change after' dates are excluded from data analysis (Jones *et al.,* 2015). Furthermore, samples that have an association to archaeological activity are also excluded from data analysis because the timing and location of the sample is more likely to be associated with human activity than fluvial activity (Macklin and Lewin, 2003). The exclusion of 'non-change after' dates and archaeologically associated material provide a broad method of quality control. The British database has previously been used to identify periods of increased fluvial activity (Jones *et al.,* 2015), which is indicative of flooding, and correlated with environmental and climatic records to determine fluvial responses to climatic and environmental changes (Macklin *et al.,* 2010). The latter is especially important in helping to understand the effect of climate change on the frequency of flooding to aid current and future flood risk management.

However, there are three main areas of uncertainty associated with the analysis and interpretation of radiocarbon dated palaeoflood data. The first area of uncertainty is around the quality of data in the British database. The second area of uncertainty is associated with the method of radiocarbon dating of organic material and the third area is associated with the methods of analysis used to interpret the radiocarbon dates.

Each fluvial environment identified in the UK is unique and responds to local environmental factors as well as climatic drivers (Macklin *et al.,* 2010). Therefore, the meta-analysis of a national dataset results in the loss or misinterpretation of information (Chiverrell *et al*., 2011). Given the existence of the database put together by Macklin et al. (2012) it would be appropriate to re-assess the record based on common factors, for example sample material, to determine how samples reflect the fluvial record, which could help to develop a formal quality control for the analysis of radiocarbon dated palaeoflood data.

Limitations of using radiocarbon dating as a method of geochronology include; the requirement of calibration of radiocarbon dates results in individual dates being converted into a range of potential ages, so the true age of a sample is not known (Chiverrell *et al.,* 2011) and the shape of the calibration curve is characterised by steps and plateaus, which can cause distortion of a radiocarbon date for example by having a very large age-range (Williams, 2012); or creating clusters of similar ages across different datasets that cause unrealistic correlations between different sources of data.

Methods of analysis applied to interpret groups of radiocarbon dates include histograms (Macklin and Lewin, 1993), summed probability distribution functions (Macklin *et al.*, 2010), relative probability plots (Jones *et al.,* 2015) and Bayesian analysis (Thorndycraft *et al.*, 2011). Histograms are an efficient and simple form of data analysis, but the interpretation of data is restricted. Summed PDFs provide a probabilistic method of data analysis, which include interpolation of data, but the application to large sets of radiocarbon dates are limited because the minimum number of samples required to generate statistically significant results for unevenly distributed data. Previous recommendations suggest that

200 is the critical minimum number of dates required for evenly distributed dates and this number has been used to justify the analysis of ‘change after’ dates; furthermore, the effect of radiocarbon error on sample number has also yet to be studied (Williams, 2012). The influence of the radiocarbon calibration curve influences the shape of the curve and hence this leads to the question what do peaks actually represent? The most recent application of relative probability plots claims to have overcome the issues surrounding the influence of the calibration curve, but it does not overcome the unknown uncertainties associated with individual sample characteristic and loss of contextual information. Bayesian analysis has not been applied to the British database but has been applied to Spanish palaeoflood data, where site specific datasets are favoured to represent and interpret regional flood records (Thorndycraft *et al.,* 2011). There is the potential to apply Bayesian techniques to analyse and verify regional or site-specific flood-frequency records as opposed to applying meta-analysis. The effect of uncertainties on the interpretation of palaeoflood data is unknown and this has been acknowledged in previous studies (Chiverrell *et al.*, 2011a; b) but has not been formally tested. Furthermore, spectral analysis has proved useful in other hydrological studies to identify periodicities and is a method specifically designed for unevenly noisy datasets. Given the limitations discussed here and in mentioned studies, an opportunity is present to test the sensitivity of the British database and to test an alternative method of data analysis. This study researches the gaps in knowledge discussed above to contribute to the growing number of palaeoflood hydrology studies and analysis of large sets of radiocarbon dates. However, it is clear that a formal study of the viability and reliability of the data and data analysis methods used to develop a British database is needed.

Chapter 3 Methodology

3.1 Introduction

This chapter sets out how this study accessed and prepared the data within the British database and explains the methodologies that were used to calibrate radiocarbon dates, produce summed probability plots, and test the statistical integrity of the data. Further methodologies are then applied in this study including the analysis of sub-datasets of the British database (provided in 4.2); determining how many dates are required to generate statistically reliable results in relation to the British database, and how the radiocarbon laboratory error affects the minimum number of samples (provided in 5.2); identification and comparison of the occurrence of radiocarbon plateaus and steps within the British database and the application of spectral analysis (provided in 6.2). Several studies have shaped the methodology chosen in this study to re-analyse the British database and these are highlighted below.

3.2 Data selection

3.2.1 British database

The British database was introduced and described in 2.1.3 but for the purposes of this chapter, a definition of the database is provided here. The British database is a collection of radiocarbon dated organic material from fluvial depositional environments with the United Kingdom (Macklin *et al.*, 2012).

3.2.2 Database conversion

The British database is open source and freely available to download as a supplementary document to Macklin et al. (2012) as Data Repository item 2012221 at: http://www.geosociety.org/datarepository/2012/. The database is presented as a table; the sample ID is provided in the first column and is accompanied by other information available about the sample, such as catchment name and sample depth. The document is only available as a PDF document so the table of 776 samples was converted into a Microsoft Excel spreadsheet. Most of the conversion was carried out using a PDF to Excel converting software, which is freely available at https://www.pdftoexcel.com/. But some data had to be hand-typed as the format of the table in the PDF is not consistent, so the conversion software did not recognise some of the information.

3.3 Data analysis

3.3.1 Radiocarbon calibration

The British database provides the radiocarbon determination and associated laboratory error for each sample. It was necessary to re-calibrate the radiocarbon determinations rather than use the calibrated age ranges from the original study (Macklin *et al.,* 2012) because the calibration curve used has been superseded by IntCal13 (Reimer *et al.,* 2013). Calibration of radiocarbon determinations is a necessary methodological step (as identified in 2.2.1). There are different software available to calibrate radiocarbon determinations (Wood, 2015), such as BCal, which enables dates to be ordered stratigraphically, in particular relation to archaeological data (Wood, 2015); and Bacon, which is set up to enable the user to construct age-depth profiles for palaeoenvironmental data (Blaauw and Christen, 2011). This study uses Oxcal (Version 4.2, Bronk Ramsey, 2009) because as discussed in 2.2.1.3 it is practical and Oxcal is set up to perform a variety of activities including generating summed PDFs, which is a primary focus of this study, and efficient to use. Below is a step-by-step explanation of how to use Oxcal to calibrate multiple radiocarbon determinations.

Radiocarbon determinations can be calibrated individually or as a collection of data. This study calibrates multiple dates at a time. Steps 1 - 14 are based on the methodology provided in the supplementary document to Macklin et al. (2012) as Data Repository item 2012221 at: http://www.geosociety.org/datarepository/2012/ so if additional guidance is required see original document.

1. Format data within Microsoft Excel into three consecutive columns: Sample name, radiocarbon determination, radiocarbon laboratory error
2. Open Oxcal online at https://c14.arch.ox.ac.uk/oxcal /OxCal.html?Mode=Input
3. Select 'New'
4. Use drop-down menu to select 'Sum'
5. Click on the symbol '>>'
6. Use drop-down menu to select 'R_Date'
7. Click on the symbol '>>'
8. Select 'Tools' from the tabs at the top of the panel
9. Select 'Import'
10. Copy data from Microsoft Excel in correct format and paste into yellow box
11. Click on the symbol '>>'
12. Select 'File' and 'Run'
13. Save the file to the Oxcal Online framework
14. Select 'Run'
15. The radiocarbon determination is presented in a table (see Figure 3.1) with two potential ages based on the IntCal13 calibration curve. A range of potential calibrated ages is presented on a calendar timescale (BC/AD).

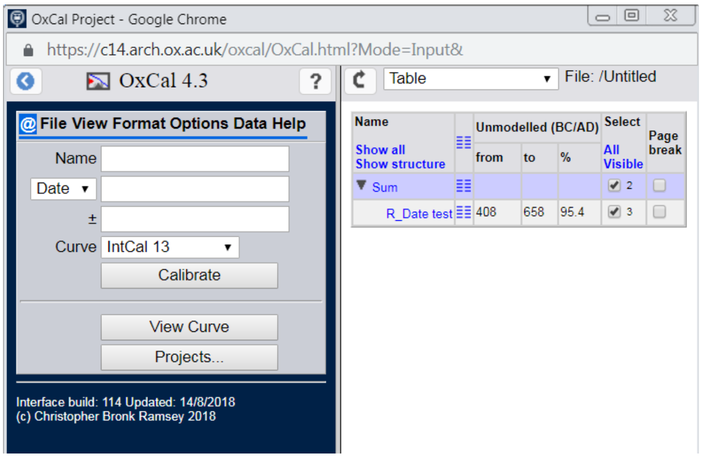


Figure 3.1 Screenshot of the output given in Oxcal for the calibration of a single ^{14}C age determination (Bronk Ramsey, 2009).

3.3.2 Generation of summed probability distribution function

Summed PDFs are generated using the outputs of the ‘Sum’ command in Oxcal. Step 4 above includes the function ‘Sum’ so no additional work is required to gain the values to construct a summed PDF. So, following on from step 15, the following methodology was used to construct summed PDFs.

16. Follow steps 1-15
17. Click on ‘≡ ≡‘ symbol to the right of the ‘Sum’ column
18. This will show two columns of data: years cal BP in 5-year increments and probability
19. Copy all values within yellow box (see Figure 3.2)

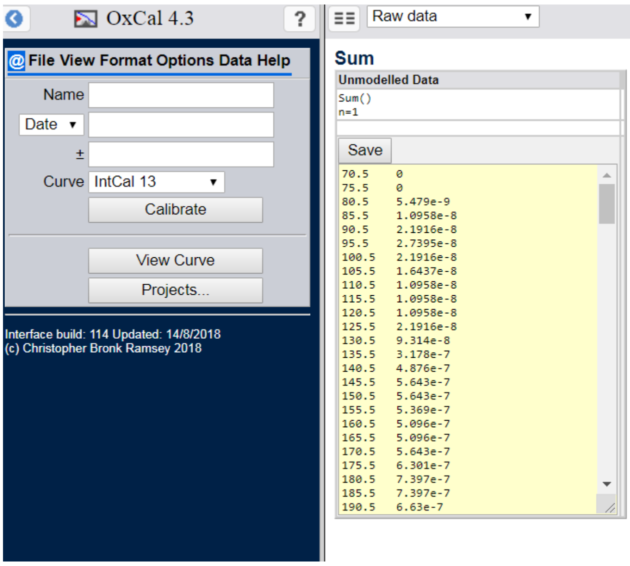


Figure 3.2 Screen shot showing output of 'Sum' function within Oxcal of a single ^{14}C age determination (Bronk Ramsey, 2009).

20. Paste values into two columns (cells A1 and B1) in Microsoft Excel (titled years cal BC/AD and probability respectively)
21. Create new column (years cal BP) in cell C1
22. Enter formula “=1950-B2” and drag formula down to convert all values in column B from years cal BC/AD to years cal BP
23. Create new column to show relative probability in cell D1
24. Divide all values in column C1 by the highest value in column C to make the relative probability values between 0-1
25. In Microsoft Excel, select ‘Insert’ from the tabs at the top of the screen
26. Select ‘Scatter with smooth lines’
27. Input data from column D into the x-axis and data from column B into y-axis to generate summed PDF

3.3.3 Generation of relative probability plots

Relative probability plots are a variation of summed PDFs as explained in 2.3.2. Following the methodology used in Macklin et al. (2012), RPPs were generated following the next steps.

28. Complete steps 1-15 for the British database and the sub-dataset
29. Arrange output data so that the probabilities for each dataset is aligned with the same 5-year increment, for example Table 3.1

Table 3.1 A demonstration of the structure of data following step 29.

	A	**B**	**C**
1	Years cal BP	Fluvial record probability	Sub-dataset probability

30. In the next column (D), insert the formula "=C2/B2"
31. In column E, calculate the relative probability by dividing al l values in column D by the highest probability value
32. To create relative probability plot, select 'Scatter with smooth lines'
33. Enter data from column A into x-axis and data from column E into y-axis

3.3.4 Calculation of number of calibrated dates

The number of dates per time interval has previously been represented by uncalibrated ^{14}C ages so when probability curves are graphed with the number of dates, they are out of sync because the probability curves are based on calibrated dates. Therefore it is difficult to visually link the availability of data with the shape of probability curves.

For this study, the number of calibrated dates was calculated using the following steps (based on methodology provided in personal correspondence; Bronk Ramsey, 2016):

1. Identify time interval, for example 200 years
2. Sum the probability values (as shown in Figure 2.3) within the chosen time interval
3. Multiply the summed probability value by 5

3.3.5 Statistical indicators of variance

Two formal statistical tests were applied to the output data generated by the summed probability distribution functions to provide a formal evidence base of the relative reliability of the radiocarbon dates and in turn the relative reliability of the palaeoflood data.

3.3.4.1 Sum of squares of deviation

Sum of squares of deviation (SSD) was used as an indicator of statistical variability in a given sub-dataset relative to the British database. SSD has been used in previous studies to show the relationship between SSD and the number of samples analysed in relation to radiocarbon dated peat deposits to determine the minimum number of samples required to generate statistically reliable results based on Geyh (1980) classification of histograms (Michczyńska and Padzur, 2004). Further details of this technique are provided in 4.2.

SSD was calculated by following the steps below:

1. Carry out steps 1 – 24
2. Insert Equation “DEVSQ” in Microsoft Excel
3. Insert all probability values of given dataset into ‘Number 1’ box (see Figure 3.3 below)

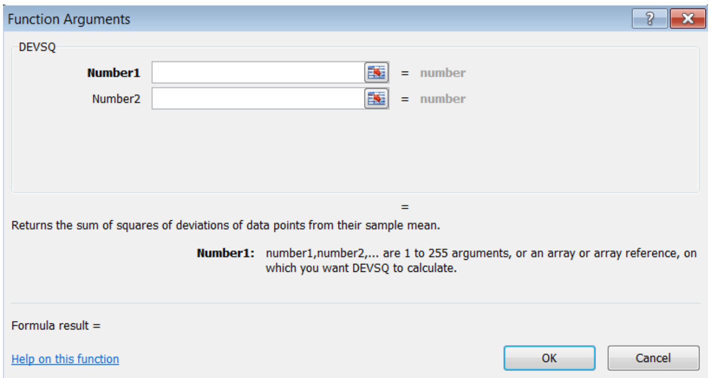


Figure 3.3 Screenshot of "DEVSQ" Equation input in Microsoft Excel.

4. Repeat steps 34-36 for each sub-dataset

Values of SSD are presented in tables in this study to compare values for different sub-datasets of the British database.

3.3.4.2 Summed mean square error

Summed mean square error (SMSE) was used to measure the similarity between a given sub-dataset and the British database. SMSE has been used previously in other studies to show the difference between two sample means, in particular sample size and variance (Williams, 2012). This statistical test was used to determine which sub-datasets are statistically similar or dissimilar to the British database, which was used to determine the dominating influences over the shape of probability curves generated using the British database. SMSE values are relative to each sub-dataset so it is not like correlation where a single value signifies a specific characteristic. For example, if the SMSE value of a sub-dataset is similar to the SMSE value of the British database then this can be interpreted to reflect that the British database is strongly influenced by a given sub-dataset. Whereas, if the SMSE value of a sub-dataset is relatively much lower than the SMSE value of the British database, this can be interpreted to show that the British database is not as sensitive to this sub-dataset. This study uses SMSE as an indicator for radiocarbon laboratory error on sample size.

The SMSE value was calculated by:

5. Create sub-samples of different sample sizes of the British database and of a randomly generated dataset by using the “RANDBETWEEN” Equation in Microsoft Excel for n= 50, 100, 200…..1,000
6. Calibrate the sub-samples using steps 1 – 24
7. Create a new sheet
8. Enter the probability values corresponding to 12,500 – 0 years cal BP into columns F - O
9. Insert column headings in A1 – D1 respectively: ‘cal years BP’, ‘mean probability density value’, ‘variance’, and ‘MSE’
10. Use five-year increments from 12,500 to 0 in column A; make sure probability values correspond to the correct year
11. Calculate values for column B by inserting formula “=AVERAGE(F2:O2)” into cell B2 and carrying the formula down to auto-fill the values for other years
12. Calculate values for column C by inserting formula “=VAR(F2:O2)” into cell C2 and carry down the formula again to auto-fill values
13. The mean sum of standard deviation is calculated by using the formula:

Equation 3.1 Mean sum of standard deviation (Williams, 2012).

$$MSE_n(t) = var[PD^n(t)] + [mean[PD_n(t)] - PD_{al\,l}(t)]^2$$

Where; $MSE_n(t)$ = mean square error at the specific time interval (t) based on sample size (n); $var[PD_n(t)]$ = variance of probability density values at the specific time interval (t) based on sample size (n); $mean[PD_n(t)]$ = mean of probability density values at the specific time interval (t) based on sample size (n); and

$PD_{all}(t)$ = probability density values at the specific time interval (t) based on the entire data set (Williams, 2012).

14. Finally, the individual values of SME for each 5-year increment is summed
15. This final value is compared with other SMSE values for other values of radiocarbon laboratory error and different number of samples by plotting a scatter plot
16. To plot the scatter plot, select number of dates for the x-axis and SMSE for the y-axis

3.4 Summary of methodology

The wide application and recent developments of radiocarbon dating (as discussed in 2.2.1) have led to developments in techniques used to analyse large numbers of radiocarbon dates. In particular, the ongoing development of the calibration curve (Intcal13, Reimer *et al.,* 2013) and the advancement in modelling techniques within Oxcal (Bronk Ramsey, 2009) has increased the accessibility and availability to radiocarbon users in general. This study used the Intcal13 calibration curve (Reimer *et al.,* 2013) and Oxcal (version 4.2, Bronk Ramsey, 2009) to construct summed PDFs using the methodology provided by Macklin et al. (2012).

After the initial application of summed PDFs to analyse radiocarbon dated fluvial units (Macklin and Lewin, 2003), several studies have focused on the reliability of the shape of summed PDFs in relation to the minimum number of samples required to create a summed PDF with a statistically reliable shape (Michczyńska and Pazdur, 2004), and the effect of the calibration curve on the location of peaks and troughs (Williams, 2012; Chiverrell *et al.,* 2011; Michczyński *et al.*, 2006). These studies have collectively provided new knowledge and identified gaps in knowledge, as summarised in 2.4 that this study aims to address. All of the methods chosen for this study have also been selected to be transferable to other disciplines and also easy and quick to use in any study involving groups of radiocarbon dates to determine the reliability of radiocarbon dates and to incorporate this into interpretations of radiocarbon dates.

Chapter 4 Sensitivity testing of the use of summed probability distribution functions in relation to the British database: Part 1 - Quality control of radiocarbon ages

4.1 Introduction

As described in Chapter 2, the British database has evolved since the first application of a probabilistic approach (Jones *et al.,* 2015). One key development was the identification of 'change after' dates, which has been used to determine the association of a radiocarbon age to a potential flood event, and hence determine the reliability of fluvial deposits. In the British database, there is an extensive classification system with information about each deposit, such as sample material – and the effect of this classification and its subsequent use for date selection has not been studied before. Yet in geochronology studies, the type of sample material is used as a criterion for quality control because the interaction of carbon with sample materials is different and therefore the resulting radiocarbon age reflects different 'events' (Rodríguez-Rey *et al.,* 2015). The use of summed PDFs to represent large numbers of radiocarbon dates in relation to fluvial deposits has been criticised because supporting information for each radiocarbon age, which could be used to interpret the age more reliably, is lost when meta-analysis was applied to (Chiverrell *et al.,* 2011).

Such differences in radiocarbon age determinations are of course highly relevant to palaeoflood hydrology because the radiocarbon dates are assumed to reflect a flood event, so previous interpretations of collective sample material could be wrong or at the very least alter the 'timings' of fluvial activity. Furthermore, when summed PDFs are used to represent and interpret large groups of radiocarbon dates, there is also a gap in knowledge of how data selection affects the shape of summed PDFs. If the use of radiocarbon dated fluvial deposits are to be considered a viable and reliable source of information, the fields of palaeoflood hydrology and geochronology need to be looked at closer together. Understanding the importance of data selection could provide an initial step into the development and application in future studies.

This chapter tests the sensitivity of the shape of summed PDFs to different characteristics of the British database by bringing together practices used in palaeoflood hydrology and geochronology studies. The characteristics chosen for each sub-dataset in this study were chosen based on their association, which is defined as a ^{14}C age having direct or indirect association to a specific 'event' (Rodríguez-Rey *et al.*, 2015).

4.2 Methodology

To test the role of sample selection as discussed above, the influence of four characteristics of the British database were studied. These characteristics were:

a. sample context to archaeological material

b. sample association to a flood event

c. sites with multiple samples

d. sample material.

The following methodology was applied:

1. Sub-datasets were created from the British database for based on the characteristics (a-d) defined above;

2. The radiocarbon determinations of each sub-dataset were recalibrated using the methodology set out in 3.3.1;
3. Summed probability distribution functions were plotted using the methodology set out in 3.3.2;
4. The mean relative probability of each sub-dataset was calculated by using the mean relative probability value and plotting it on each graph as a dotted black line;
5. The summed mean square error was calculated for each sub-dataset using the methodology set out in 3.3.4 and compared with the summed mean square error for the British database. A high value of summed mean square error will indicate the British database is highly sensitive to the sub-dataset, whereas a low value indicates a low influence on the shape of the summed probability distribution function;
6. The sum of squares of deviation was calculated using the methodology in 3.3.4. The sum of squares of deviation shows the variability in the sub-dataset from the sample mean; a low value indicates low variability from the sample mean and is considered more reliable than a high value.

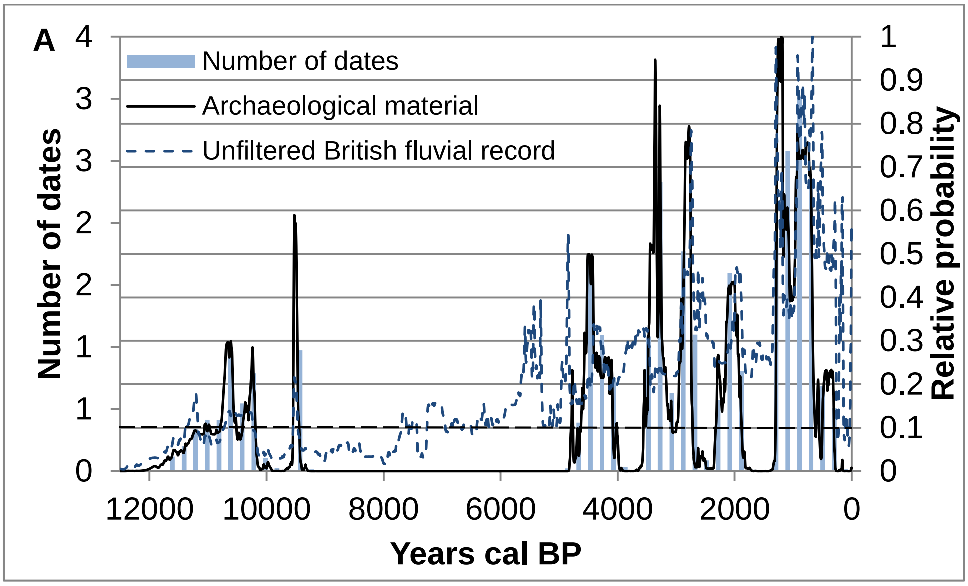
It is not possible to determine if the statistical variance for the unfiltered British database are high or low because the true shape of the summed PDF curve is unknown. The purpose of running these tests is to determine if and what data characteristics influence the shape of the curve constructed using the unfiltered British database dataset. These tests use the assumption that higher statistical indicators of variability reflect a high influence on the British database and that lower values reflect less variability within a sub-dataset and hence are more reliable.

4.3 Results

Below, summed PDFs for each sub-dataset are shown along with the summed PDF for the unfiltered British database. All of the graphs are presented on the same horizontal axis: from 12,500 years cal BP on the left-hand side, to 0 years cal BP on the right-hand side. The vertical axes are on different scales, which are dependent on the availability of data, and are discussed for each sub-dataset. The shape of each curve is compared and discussed in relation to the location of peaks in the curve and statistical indicators of variability as defined in 3.3.4. The end of 4.3 is characterised by a comparison of the location of peaks across each sub-dataset to the unfiltered British database.

4.3.1 Sample context in relation to archaeological activity

This section reanalyses samples material that has been interpreted to be related to archaeological activity (Macklin *et al*., 2012) using the most recent radiocarbon calibration curve – Intcal13 (Reimer *et al*., 2013); and provides the first formal test of the statistically reliability of probability curves generated using a sub-dataset based on archaeological context.



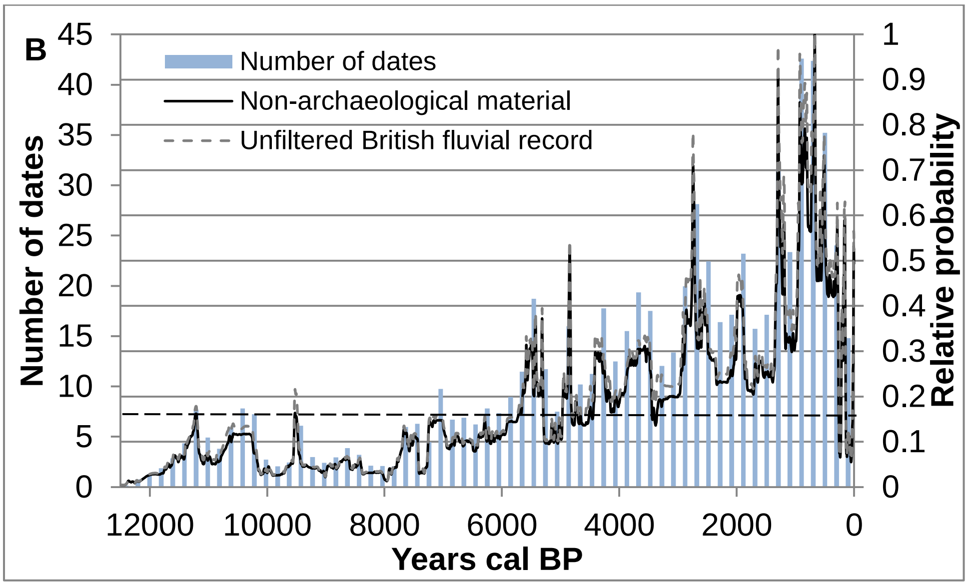


Figure 4.1 Summed probability distribution function of A: samples associated with archaeology and B: samples not associated with archaeology from the British database (Macklin et al., 2012). The number of dates in each sub-dataset per 200 years is shown by a histogram. The mean relative probability is shown by the black dotted line.

Figure 4.1A is characterised by six periods of intermittent increased probability throughout the Holocene, which all exceed the mean relative probability, and gaps in the curve (when the probability equals 0.00). This makes the shape of the unfiltered British database more noticeable in Figure 4.1A than in Figure 4.1B. Each of the six peaks occurs for different lengths of time ranging between 100 and 700 years and there is some cross-over between peaks in the unfiltered British database and the archaeological material. The longest period of time when a gap in the curve is present, which is reflected by a probability of 0.00, occurs between 9,000 – 5,000 years cal BP and the highest probability occurs in the Late-Holocene at around 665 years cal BP. There is not a clear increase in probability over time, the height of each peak varies and corresponds to the number of dates available.

The shape of the curve in Figure 4.1B exceeds the relative mean probability intermittently from around 5,700 - 0 years cal BP but the relative probability never falls to 0.00 like in Figure 4.1A. Even in the early to mid-Holocene (12,500 – 5,700 years cal BP) when there are noticeably fewer dates available, the relative probability falls below the mean relative probability but never as low as 0.00. The Late-Holocene is characterised by a continuous period of probability that exceeds the mean from around 3,350 – 0 years cal BP and the shortest period of increased probability is around 100 years long and occurs at 4,850 years cal BP. The shape of the unfiltered British database is barely noticeable in Figure 4.1B and is only noticeable when the relative probability of peaks within the unfiltered British database at 9,535, 4,850, 2,750, 1,295 and 925 years cal BP that exceed the relative mean of the non-archaeological sub-dataset. All five peaks are also identified in the archaeological dataset. The occurrence of peaks in each sub-dataset is presented and compared in Table 4.5.

The number of dates in each sub-dataset presented in Figure 4.1 differs; there are only 30 dates that are associated with archaeological material (0.4% of the total number of dates in the British database, Macklin *et al.,* 2012), and 746 dates that are not associated with archaeological material. In Figure 4.1B there is at least 1 sample per 200 years whereas in Figure 4.1A there are long periods of time when no data is available, and this is reflected by a probability of 0.00. Table 4.1 shows how the number of dates correlates with the statistical indicators of variability for each sub-dataset and the unfiltered British database.

Table 4.1 Statistical comparison of sample material context in relation to the British database.

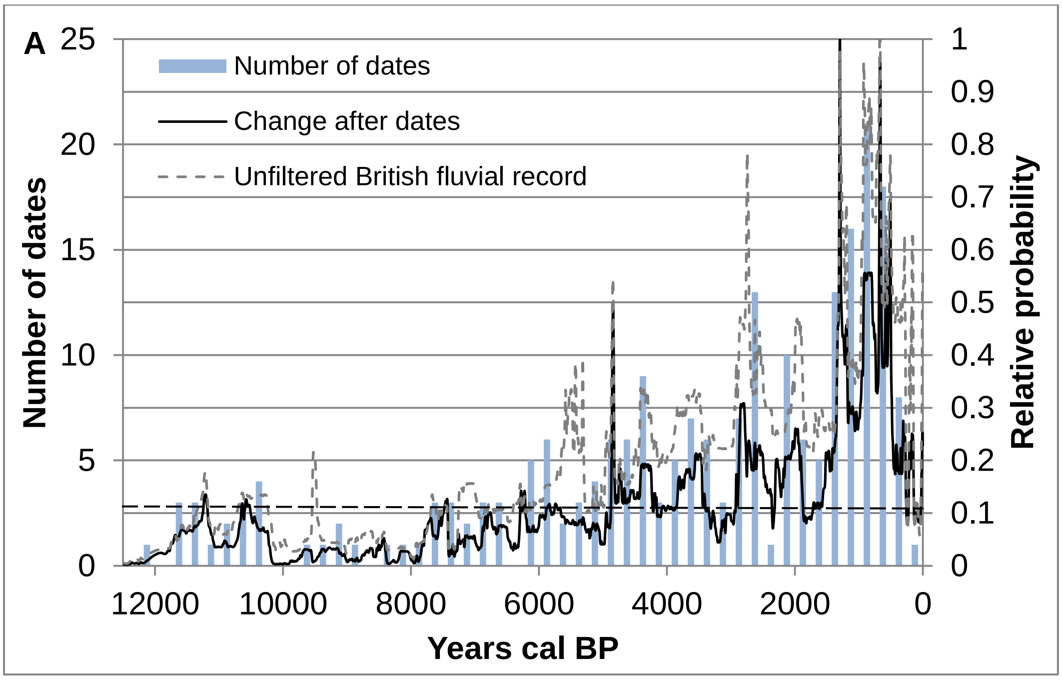
	Number of dates	**Sum of squares of deviation**	**Sum of mean square error**
Archaeological material	30	0.048	21.291
Non-archaeological material	746	6.926	144.838
British fluvial record	776	7.641	150.579

It was observed in Figure 4.1 that the shape of the curves for each sub-dataset increased in height as the number of dates per time interval increased, which suggested that the shape of the curve could be related to the number of dates. The dates that have an archaeological context have a relatively low variation from their sub-sample mean (SSD = 0.048) and the sub-dataset of dates that do not have an archaeological context have a relatively high variance from the sub-sample mean (SSD = 6.926); the non-archaeological material sub-dataset is more similar to the corresponding value for the British database. Table 4.1 shows that the statistical indicators of variability also increase as the number of dates increase. The number of dates in each sub-dataset based on archaeological context is very uneven so it is not possible from this table alone to determine the cause of the observed correlation; but the following results, which have a more even spread of data across sub-datasets, may be able to provide evidence to make a more justified interpretation.

The statistical values of SMSE shown in Table 4.1 suggest that non-archaeological dates are most similar to the unfiltered British database, which is unsurprising because 96% of the dates in the British database do not have an archaeological context. This finding supports the assumption that sub-datasets that have an SMSE value close to the British database reflect similarity between datasets, furthermore, the data used to calculate the statistical values of variability is the same as the data used to construct the summed PDF curve, it is fair to interpret the results from Table 4.1 to show that non-archaeological material has a similar shape to the curve generated using the unfiltered British database.

4.3.2 Sample association to a flood event

This section reanalyses ‘change after’ dates (Macklin *et al*., 2012) using the most recent radiocarbon calibration curve – Intcal13 (Reimer *et al*., 2013); and provides the first formal test of the statistically reliability of the ‘change after’ sub-dataset.



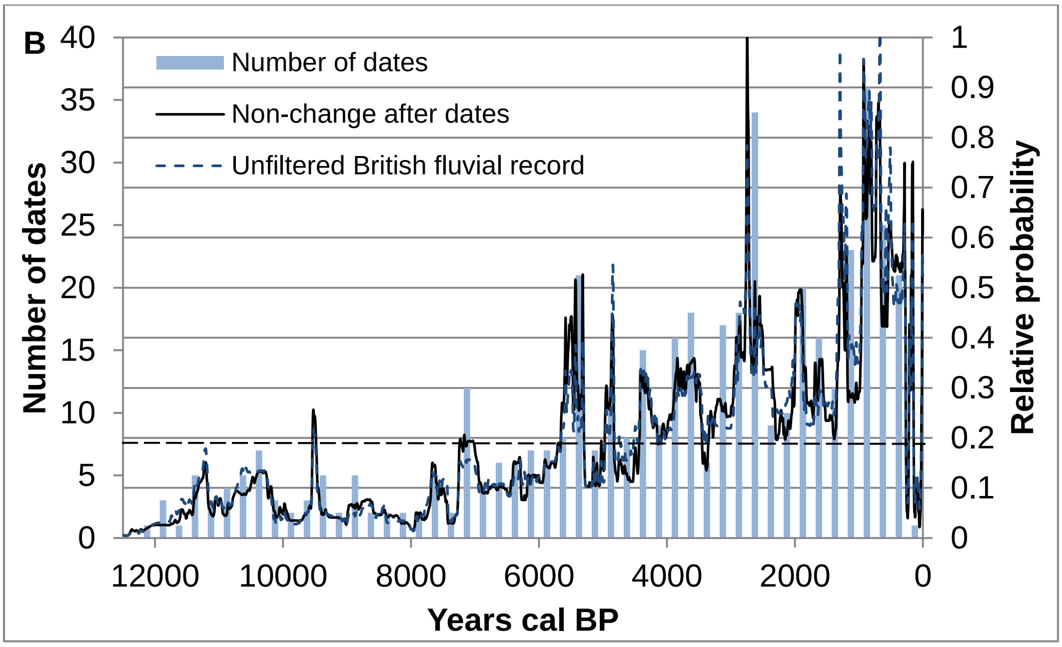


Figure 4.2 Summed probability distribution function of A: 'change after' dates and B: 'non-change after' dates from the British database (Macklin et al., 2012). The number of dates within each sub-dataset per 200 years is shown by a histogram.

Both Figure 4.2A and 4.2B show a similar pattern between time and probability; there are more periods when the relative probability exceeds the mean probability in the Late-Holocene (from around 6,000 - 0 years cal BP). The length of time that the relative probability exceeds the mean ranges between 100 – 1,450 years (in Figure 4.2A) and 100 – 3,200 years (in Figure 4.2B). There are 2 time intervals when there are no dates available in the ‘change after’ sub-dataset that occur between 12,000 – 11,800 years cal BP and 10,200 – 9,800 years cal BP; but unlike the curve in Figure 4.1A, the probability value does not drop as low as 0.00. The shape of the unfiltered British database is more noticeable in the Figure 4.2A because the height of peaks in the curve continuously exceeds the shape of the curve generated using ‘change after’ dates. Whereas in Figure 4.2B, the height of the curve using the unfiltered British database only exceeds the peaks in the ‘non-change after’ dates five times. These occur at 11,240, 4,850, 1,295, 675 and 515 years cal BP. In addition, the peaks that were identified in Figure 4.1 (9,535, 4,850, 2,750, 1,295 and 925 years cal BP) are represented differently in Figure 4.2A and 4.2B. The peak at 9,535 years cal BP is present in the ‘non-change after’ curve but not ‘change after’ dates curve, the peaks at 4,850 and 1,295 years cal BP are represented by a sharp narrow peak in both curves, and the peaks at 2,750 and 925 are illustrated by sharp narrow peaks in the ‘non-change’ after curve but occur as a smaller peak in a broad period of increased probability above the mean in the ‘change after’ curve. This suggests that when the British database is sub-sampled based on the association to a flood event inferred by the stratigraphic location, the shape of the summed probability curve is affected.

Similarly to Figure 4.1A and 4.1B, the shape of the curves in Figure 4.2A and Figure 4.2B correlate well with the number of dates available; the height of peaks in the curve increase when the number of dates increases. Despite the difference in the number of dates in each sub-dataset (as shown in Table 4.2 below) the availability of dates per 200 years shows a general increase in the availability of dates towards the Late-Holocene in both sub-datasets.

Table 4.2 Table showing a statistical comparison of the association of sample material to flood events in relation to the British database.

	Number of dates	Sum of squares of deviation	Sum of mean square error
Change after dates	236	0.978	54.397
Non-change after dates	531	3.453	104.320
British fluvial record	776	7.652	150.579

Table 4.2 identifies that 30% of the dates within the British database are classified as 'change after' dates and 70% are 'non-change after' dates. Table 4.2 also shows that there is the least variation within the 'change after' sub-dataset, which contains 87% less statistical variation (as measured by SSD) than the British database, and the sub-dataset consisting of 'non-change after' dates statistically contains 45% less variation than the British database. These findings are consistent with the suggestion made previously in this thesis that lower sample numbers correlate with lower statistical variability and the least similarity between the shapes of the curve of the unfiltered British database.

4.3.3 Cumulative analysis of multiple dates from single sites

The results presented in 4.3.1 and 4.3.2 have been based on the previous classification of dates used to reliably represent palaeoflood data that have not been supported by statistical justification. The next two analyses are based on further data characteristics that could also represent ^{14}C dates in additional contexts. Firstly, multiple dates from single sites are collectively analysed followed by sample material. The same methodology as given in 4.2 is applied to determine the statistical reliability of each sub-dataset.

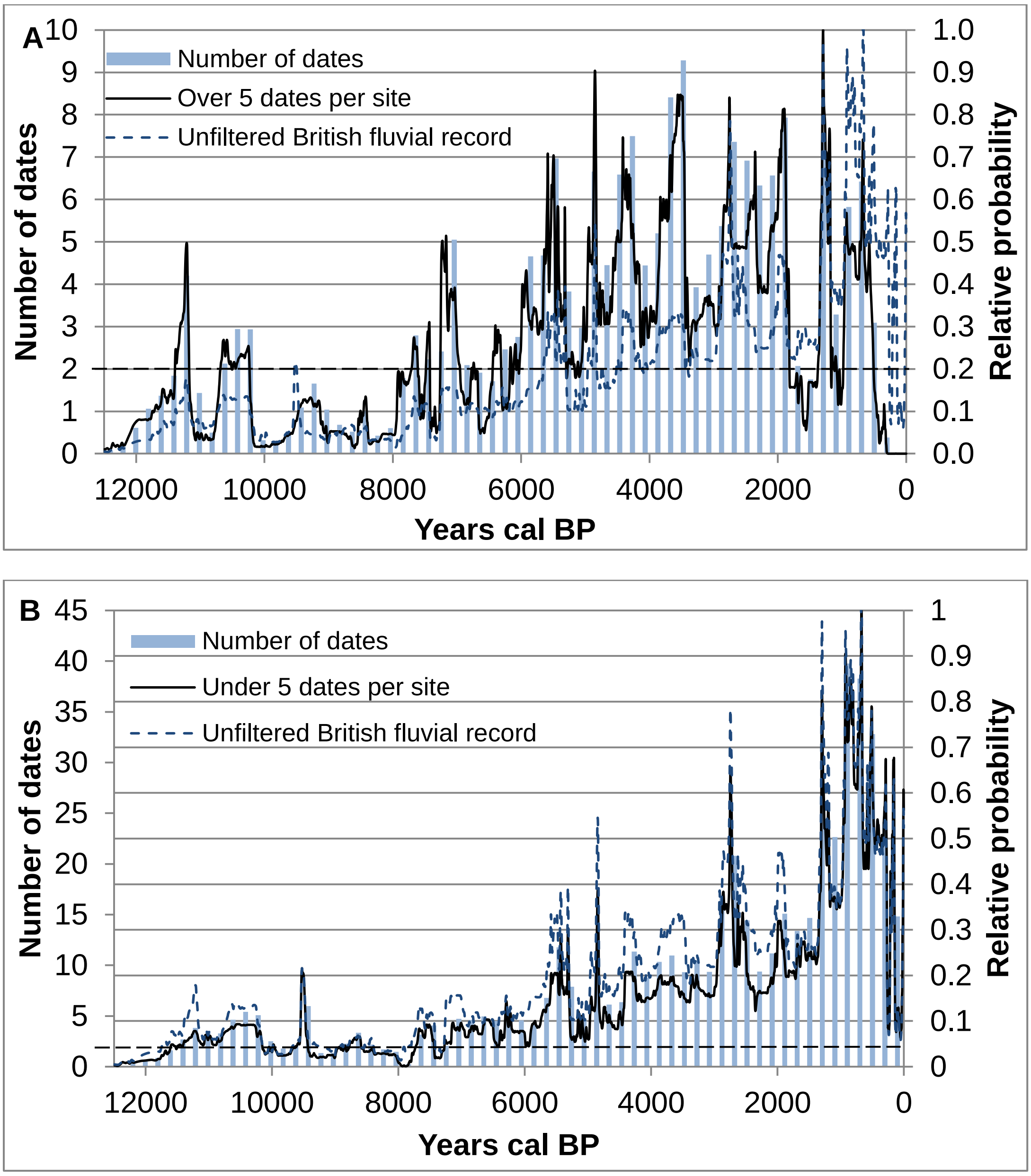


Figure 4.3 Summed probability distribution function of radiocarbon dated A: samples from sites with five or more dates and B: less than 5 dates per site from the British database plotted with the number of calibrated dates per 200 years.

Figure 4.3A and 4.3B shows the Late-Holocene is characterised by more frequent and longer periods of time when the mean relative probability is exceeded; but Figure 4.3A and 4.3B are also characterised by presenting periods of time in the Early-Holocene (between 12,500 and 8,000 years cal BP) when the mean is also exceeded. This pattern was also observed in the 'change after' and archaeological sub-datasets. There is a noticeable difference between when shape of the unfiltered British database and the sub-datasets in Figure 4.3A and 4.3B; in Figure 4.3A the height of the peaks in the sub-dataset is markedly higher than the unfiltered British database whereas in Figure 4.3B, the height of the peaks in the unfiltered British database exceed the sub-dataset. The peaks that have been observed across both sub-datasets occur are signified by sharp narrow peaks except for the peak at 9,535 years cal BP, which does not occur in Figure 4.3A.

There is also a noticeable different in the number of dates available over time between both sub-datasets. There are more data available between 8,000 – 0 years cal BP in Figure 4.3A that are characterised by long periods of time (over 400 years) whereas in Figure 4.3B there is a clear increase in the number of dates towards the Late-Holocene.

Table 4.3 Statistical comparison of samples taken from site with five or more dates and under five dates in relation to the British database.

	Number of dates	**Sum of squares of deviation (SSD)**	**Sum of mean square error (SMSE)**
Site with > 5 dates	219	0.457	51.890
Sites with < 5 dates	557	5.546	108.915
British fluvial record	776	7.641	150.579

The statistical variation, as represented by the SSD in Table 4.3, of the collective analyse of over five dates per site is significantly lower than from the analysis of less than five dates per site. In comparison with the unfiltered British database, the statistical variability in the sub-dataset of less than five dates per single site is relatively high. Furthermore, the SMSE analysis demonstrates that the probabilistic data generated using data from sites with more than five dates is much lower than in the other sub-dataset and the unfiltered British database. This suggests that the variability in the British database could be strongly influenced by dates that are characterised by the collective analysis of sites with less than five dates (i.e. single ^{14}C dates).

4.3.4 Sample material

For this section, the ^{14}C dates were classified into three broad categories based on sample material as defined in the British database (Macklin *et al*., 2012). Three categories were used: wood, peat and other sample material.

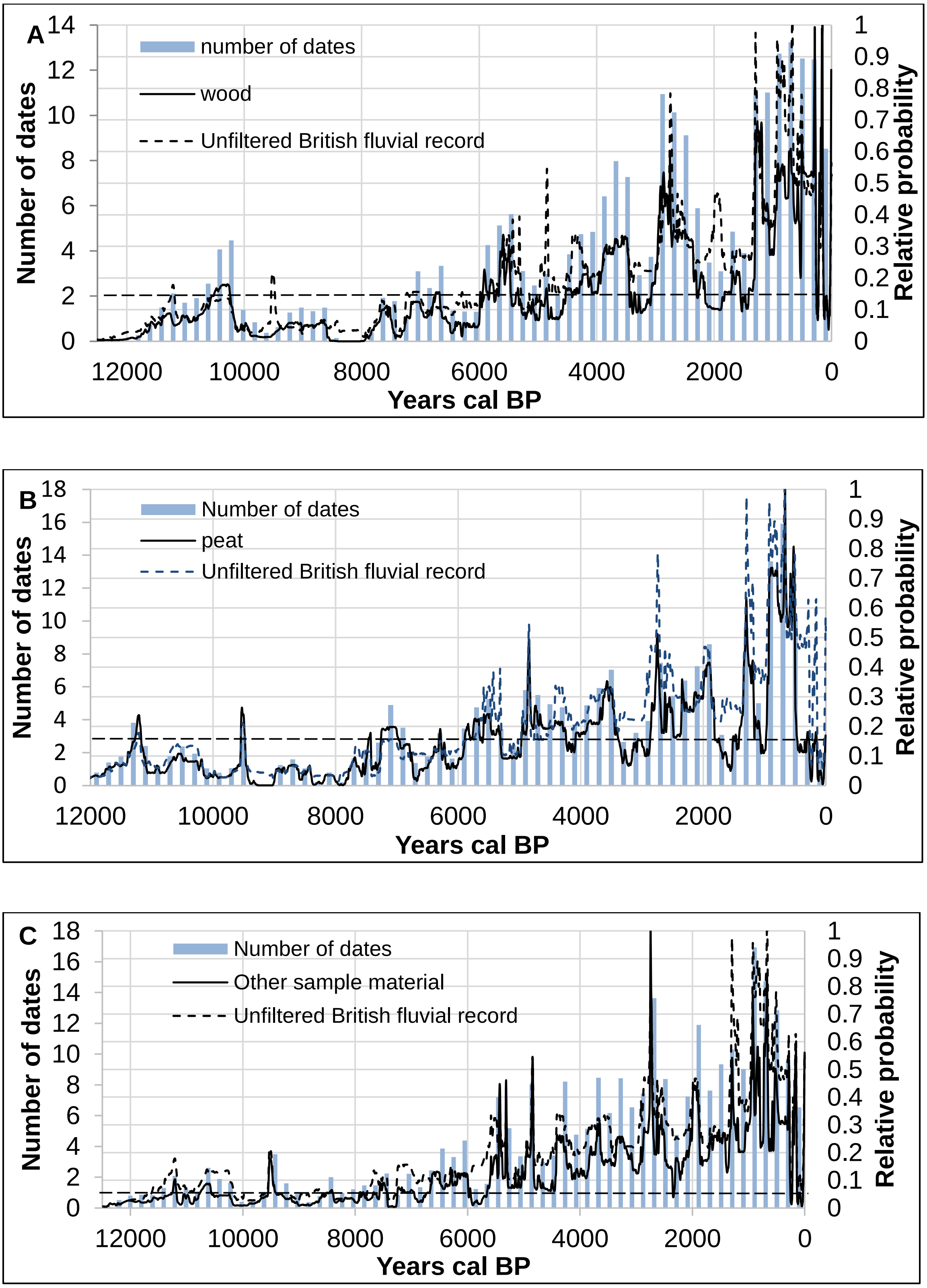


Figure 4.4 Summed probability distribution function of radiocarbon dated A: wood, sample material, B: peat sample material, and C: collective other sample material, from the British database plotted with the number of calibrated dates per 200 years.

The height of peaks in the curve generated by the unfiltered British database in Figure 4.4A-C are characterised by being higher than the curves produced by the sub-datasets, particularly between 6,000 – 0 years cal BP. In other words, the probability of a ^{14}C date occurring per given time interval is higher in the unfiltered British database rather than in sub-datasets of the British database. There is a noticeable difference in the shape of the curve in Figure 4.4A at around 9,530 and 1,985 years cal BP when there is a peak in the unfiltered British database but a decrease in probability in the wood sample material sub-dataset curve. The shape of the curves produced by wood and peat sample material are also characterised by fluctuating above and below the relative sample means between 6,000 – 0 years cal BP, rather than the height of the curve being maintained above the sample mean continuously like in the collective analysis of 'other sample material' (Figure 4.4C). In Figure 4.4A and 4.4B there are no dates available between 8,600 – 8,000 and 9,600 – 9,000 years cal BP respectively, which correlate with probability of 0.00, which is also consistent with other sub-datasets analysis. The peaks that have been identified in Figures 4.1 – 4.3 also occur in Figure 4.4 but are not characterised by sharp narrow peaks in each curve. For example, in Figure 4.4A sharp narrow peaks occur at 1,295, 285 and 165 years cal BP and there is an increase in relative probability above the mean that includes 2,750 and 925 years cal BP that are characterised by broader peaks instead of sharp narrow peaks. Figure 4.4B and 4.4C both show curves that identify sharp narrow peaks at 9,535, 4,850, 2,750, 1,295 and 925 years cal BP. Figure 4.4A also identified sharp narrow peaks at 285 and 165 which is consistent with the shape of the curves produced using sub-datasets of less than five dates per site and 'non-change after' dates.

Table 4.4 Statistical comparison of wood and peat sample material in relation to the British database.

	Number of dates	Sum of squares of deviation (SSD)	Sum of mean square error (SMSE)
Wood	250	1.052	57.199
Peat	225	0.762	53.416
Other material	301	1.375	63.129
British fluvial record	776	7.641	150.579

Table 4.4 highlights that each sub-dataset per sample material contains 30% – 40% of the total number of dates within the British database, which is a more even spread than was observed in other sub-dataset analysis and also mimics the previous suggestion made that lower sample numbers correlate with lower statistical variability, which is clear even when there is a difference of 25 – 51 dates. The statistical indicators of variability shown in Table 4.4 are significantly lower than the unfiltered British database; the SSD values for the sub-datasets are between 80 – 90% lower and the SMSE values are more than 50% lower. Based on the results presented in Table 4.4 it is interpreted that the shape of the respective curves based on sample material produce curves that are statistically similar, yet it was observed in Figure 4.4 above that the shape of the curves and corresponding peaks are represented differently.

The comparison of periods of increased probability, as defined by periods of time when the curve exceeds the relative mean probability, show that the periods, including peaks in the curve, occur at different times across every sub-dataset tested as well as in comparison with the unfiltered British database. There are some overlaps across every sub-dataset and the unfiltered British database, for example in the Late-Holocene (see Table 4.4), which also coincides with more dates being available. It is noticeable that periods of increased probability that are identified in the sub-datasets are not shown in the unfiltered British database in the Late-Holocene, for example between 12,000 – 10,000 years cal BP when periods of increased probability are identified in the sub-datasets containing archaeological material, 'change after' dates, over and under five dates per single site and across all sample materials and also between 9,500 – 5,500 years cal BP when periods of increased probability are identified in the sub-dataset containing more than five dates but again not in the unfiltered British database. This supports the statistical indicators of variance that identify an increase to mean that a specific sub-dataset has less influence over the shape of the unfiltered British database when analysed collectively, and also highlight the variability between sub-datasets.

Table 4.5 Identification of periods of increased probability for the British database and sub-datasets.

Unfiltered British database	archaeological material	Non-archaeological material	Change after dates	Non-change after dates	> 5 dates	< 5 dates	Wood	Peat	Other sample material
			11,320 – 11,300		11,390 – 11,175	11,635		11,235	11,710 – 11,100
	10,750 – 10,650					11,420 – 10,170			11,095 – 11,065
	10,500 – 10,250		10,695 – 10,655		10,680 – 10,315		10,485 – 10,345		10,865 – 10,155
9,520	9,550 – 9,540			9,530		9,680 – 9,435		9,535	9,690 – 9,295
						8,915 – 8,670			8,595 – 8,500
					7,695 – 7,615	7,820 – 7,520		7,460	8,210 – 7,420
			7,505 – 7,495		7,480				
				7,240	7,255 – 6,985	7,270 - 0		7,255 – 7,000	7,270 – 5,985
					6,745				
			6,385 – 6,305		6,440 – 6,315			6,305	
					6,170				
					6,100				
					6,000 – 5,260		5,920		5,730 - 250
5,650 – 5,330		5,725 – 5,295		5,655 – 5,380			5,700	5,885 – 5,295	
							5,675 - 5465		
4,960 – 4,745	4,780	4,970 – 4,815	4,975 – 4,235	4,960 – 4,900	5,125 – 1,820		5,295	5,025 – 4,245	
4,735									
4,480 - 265	4,630 – 4,085	4,565 – 3,395	3,840 – 3,450	4,515 – 3,445			4,810		
	3,550 – 3,150	3,385 - 260		3,415 - 265			4,585 – 3,390	4,085 – 3,440	
	3,000 – 2,700		2,920 – 2,415				3,065 – 2,300		
	2,300 – 1,900		2,410 – 1,880				2,290 – 2,175	2,865 – 1,815	
			1,695 - 250				1,800 – 1,550		
	1,300 - 635				1,370 – 1,170		1,410 - 0	1,515 – 1,075	
	625 - 255				975 - 510			975 - 480	
								260	
								220	
165									
5									

4.4 Discussion

4.4.1 Sensitivity of the shape of summed PDFs to data selection

The results presented in 4.3 demonstrate that when summed PDF curves are generated using a data selection method, the shapes of the curves generated are different and each curve is different from the unfiltered British database. Before discussing further – we should acknowledge that none of these records are correct -we do not know what the actual record of flooding is. Therefore, these are inter-comparisons and we may use the full record as the comparator (e.g. stating that other records are 'reliable' in comparison) – but it must be remembered that this is a relative reliability.

The statistical and graphical comparison of sub-datasets of the British database identified that the shape of the probability curve is clearly influenced by the data selected to construct it and by the number of dates used. When fewer dates are analysed, the statistical indicators of variability are lower reflecting both low variability within the sub-dataset and less similarity between the sub-dataset and the British database. The lowest relative statistical variability is present within the sub-dataset constructed of sample material that has an archaeological context, and the highest relative statistical variability occurs within the sub-dataset of sample material that does not have an archaeological context. To echo the point above about the effect of sample numbers, the archaeological sub-dataset has the lowest number of dates and the non-archaeological has the highest number. This finding does not justify the exclusion of archaeological sample material that has been applied in previous studies (Macklin *et al.,* 2012; 2010, Macklin and Lewin, 2003) because there are not enough dates to statistically increase the variability within the data. Furthermore, 9 dates that have an archaeological context are also classified as ‘change after’ dates, yet excluded from analysis because of their archaeological context, so could be associated with a potential flood event. So, ideally there is more to gain from including the archaeological material than discarding them - based on this specific characteristic - when other characteristics of the dates could provide useful information, such as the sample material if it is considered suitable by quality control. The sub-dataset with the second lowest statistical variability was the collective analysis of five or more dates per site. This finding is interesting as meta-analysis has previously been criticised (Chiverrell *et al*., 2011a, b) but this study shows that this is not true for this specific characteristic. Furthermore, our analysis questions the use of ‘change after’ dates as the most reliable indicator of flood events. In this study ‘change after’ dates do not produce more statistically reliable results despite there being sound justification in relation to site sedimentology and geomorphology. For example, after the analysis of archaeological material and the analysis of five or more dates per site, analysis of peat material produces less statistical variability than ‘change after’ dates.

Crucially, this demonstrates that no matter how the British database is sub-sampled (e.g. ‘change after’ as per Macklin *et al.,* 2012; 2010) the shape of the curve will change, often completely changing the distribution of peaks and this may in turn fundamentally change the interpretations made from them. Such selection is a subjective/qualitative decision and the impact of selection strengthens the argument about the unreliability of peaks within summed PDFs - that have been echoed across different disciplines (Chiverrell *et al.*, 2011a, b; Williams, 2012; Michczyński and Michczyńska, 2006). Studies of summed PDF’s are ultimately hampered as the ‘true’ shape of the curve that accurately reflects past flood events is effectively unknown. Therefore, it is only possible for the data to be analysed and interpreted as objectively as possible.

When meta-analysis is the primary method of data analysis, the process of filtering the data by identifying characteristics to determine the association of a ^{14}C age can cause a gain, and a loss of information. For example, a ^{14}C age could be classified as a 'change after' date, which is favourable (Macklin *et al.,* 2010), but could also be considered to have an archaeological context and so would be excluded from analysis therefore resulting in the loss of a ^{14}C date. Similarly, the cumulative analysis of multiple dates from single sites does not take into consideration the association to a 'change after' date or to the sample material. This study applied a single criterion of data selection but if a more robust data selection process based on ^{14}C age association were to be applied, there would be a loss of 91% of the data and this does not take into account the criteria of sample pre-treatment so there could be a further loss of data during the quality control process. This value was calculated by applying the data selection in the following order: wood, charcoal or bone sample material (316/ 776), 'change after' date (79/ 776) and not in an archaeological context (70 / 776). To put this into perspective of other studies that use collections of radiocarbon dates to reconstruct palaeoenvironmental and palaeoclimatic conditions, Michczyńska et al. (2007) use 330 fluvial deposits alongside over 1000 peat samples without reporting that a method of quality control was applied. Based on the findings of this study, it is fair to predict that there would be a significant loss of data if a robust quality control method was applied to the Polish database too. When more than one criterion for data selection is applied, there needs to be significant justification for the order they are applied because this significantly affects the amount of data that is excluded and the context of interpretation. Furthermore, the application of a robust data selection process could bias the shape of the curve produced by controlling the data that is included in analysis; but this is already happening in previous studies by including only ‘change after’ dates and excluding ^{14}C ages that have an archaeological context (Macklin *et al.,* 2012).

4.4.2 Quality Control Protocol

Uncertainty is present in individual and groups of radiocarbon dates and is represented in radiocarbon measurements as an absolute value (in years) that accompanies a radiocarbon measurement, which represents chemical and physical error (Bayliss *et al*., 2004). Laboratory uncertainty propagates into data uncertainty and this affects the precision and reliability of data, see Figure 4.5. Another source of uncertainty is the statistical analysis of radiocarbon dates, as discussed in 4.4.1. The application of a robust quality control protocol would assess the uncertainty at each stage of sample analysis before statistical analysis is carried out.

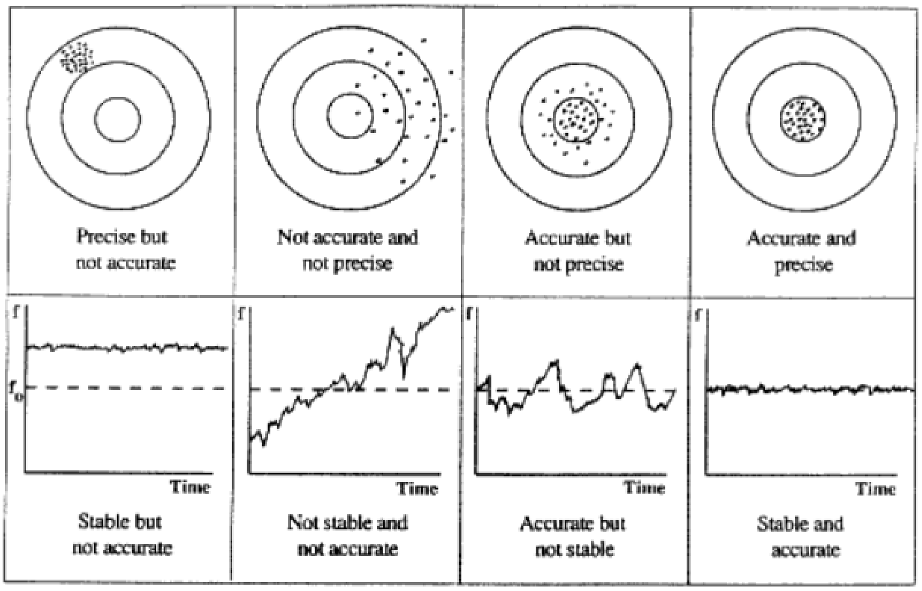


Figure 4.5 Graphic illustration showing the difference between accuracy and precision in statistical analysis (Vig, 1992).

Based on the interpretation of the results presented in this study and the practice of interpreting radiocarbon dates in other studies across different disciplines, this study recommends that a quality control protocol be applied to the unfiltered British database to work towards producing an objective curve. In archaeological studies, quality assurance is reported in great detail (Bayliss *et al*., 2008). Quality assurance that is reported includes the quality assurance procedures of individual radiocarbon laboratories used, reporting of offsets including specific values, and the reporting of duplicate measurements. The exercise of repeating measurements on single samples tests if measurements from the same sample are statistically significantly similar or different. If the measurements are different, supporting information about a sample is used to determine which measurement is likely to be the most accurate and the justifications of any decision made are reported in detail (Bayliss *et al*., 2008). Using the results presented in 4.3 and the information reported in the British database (Macklin *et al.,* 2012), it may be possible to bridge the gap of practice in radiocarbon studies in geomorphology, palaeoflood hydrology and geochronological studies to match the standard of archaeological studies.

Previous applications of quality control were discussed in 2.2.1.5 and to the author's knowledge the appropriateness of a robust quality control protocol to palaeoflood data - that consider the geochemical component of the radiocarbon ages being used, such as the protocol used in Rodríguez-Rey et al. (2015) - has not been tested. In this study, the British database was re-assessed to see how applicable the quality control protocol applied by Rodríguez-Rey et al. (2015) could be based on the results presented in 4.3. The first step involved scoring the sample material and radiocarbon methods of pre-treatment used on the sample; based on the criteria set by Rodríguez-Rey et al. (2015), wood, charcoal and bone would pass as acceptable sample material, but bulk soil organics would not because the interaction and source of carbon and structure are unknown. At this early stage of the quality control process, 59% of data within the British database database would be excluded, leaving 320 dates that would be considered as suitable to pass to the next stage of the quality control process. The methods of sample pre-treatment are not provided in the British database database, so it would not be possible to move to the next stage of the quality control process; however the laboratory codes for each sample are provided in the database so it is possible that this information can be gained through further investigation. Then, a robust method of quality control could be applied to the fluvial database.

The next stage of the quality control process involves determining if the sample has a direct or indirect association with the 'event' being reconstructed so that the radiocarbon age can be interpreted to reflect a specific 'event' (Rodríguez-Rey *et al.*, 2015). Radiocarbon dates of different sample material represent different dates, and this has not been considered in previous palaeoflood studies where different types of sample material have been analysed collectively (Macklin *et al.,* 2012). In relation to the results presented in this study there are four ways that the British database could be sub-sampled and analysed and each method would represent different 'events' that could be considered in the interpretation rather than the collective analysis of radiocarbon dates that represent different 'events'. For example, 'change after' dates represent the sedimentary context and identify abrupt changes in deposition patterns that could be attributed to a flood event (Jones *et al.,* 2015). The analysis of 'change after' dates has determined that the cumulative presence of 'change after' dates is interpreted as periods of increased fluvial activity (Macklin *et al.,* 2012). The exclusion of sample material that is in the context of archaeology is based on the assumption that the sample material is more likely to have been deposited by human activity than by fluvial activity (Macklin and Lewin, 2003) but this excludes sample material, such as bone and charcoal that is considered as suitable sample material in geochronological studies (Rodríguez-Rey *et al*., 2015). The final example of how different collections of radiocarbon dates could represent different 'events' relates to sample material. Wood represents the date of sample burial but is interpreted collectively with peat and other bulk organic material so show increased fluvial activity. However, the radiocarbon age of peat and other bulk organics are not considered suitable material to make interpretations from (Rodríguez-Rey *et al.*, 2015) as discussed in 2.1.4. Also, the time interval between a sample of wood material and the date of deposition can be problematic, which is referred to as the 'old wood' effect, and can result in an overestimation of radiocarbon ages (Pettitt *et al*., 2003). Therefore, sample material such as leaves and seeds are preferred as they have a relatively short lifespan and so the time between radiocarbon incorporation and deposition is likely to be shorter and reduce the uncertainty associated with age overestimation (Pettitt *et al*., 2003). There are evidently many ways that radiocarbon dates from fluvial environments could be sub-sampled and interpreted but the data is still limited by lacking local and regional environmental and climatic factors, such as rainfall patterns. This limitation could be addressed once the database is at the required standard of best practice and with the addition of data from future studies.

4.4.3 Summary

The implications of these findings show that when the British database was re-analysed it is not possible to apply a robust quality control protocol using methods used by the geochronology community because the radiocarbon pre-treatment methods information is not reported in Macklin et al. (2012). It is fair to support

Jones et al. (2015) in that the British database provides a collective net gain of information, but in order for the radiocarbon dates within the database to be considered suitable and hence reliable and be studied on par with other geochronology studies, the geochemical pre-treatment methods must be retrieved. The application of robust quality control criteria would validate the radiocarbon dates included in analysis, which would be a significant advancement for palaeoflood hydrology studies, in particular when sedimentary records are the predominant source of data, as well as in other fields that use radiocarbon dates.

A final point for discussion is the correlation of the shape of the curve to the availability of dates, which shows that most of the peaks in curves correspond to a relative increase in sample numbers, regardless of how the data is selected and analysed. The pattern occurred in every curve presented in 4.3. This leads to the fundamental question: is there enough data to analyse the British database using summed PDFs?

4.5 Conclusion

This chapter tested the sensitivity of the shape of summed PDFs to different characteristics of the British database by bringing together practices used in palaeoflood hydrology and geochronology studies. When the British database is analysed using sub-datasets, which were selected based on unique characteristics of the data such as sample material, the shape of curves in the summed PDFs are influenced by the data selected and the number of dates. Despite the findings presented above, the reliability of radiocarbon ages presented in the British database are questionable (Macklin *et al.,* 2012) because the reliability cannot be probably assessed because the meta-data information on the geochemical status of the dates is unavailable. Without this information, a robust method of quality control cannot be applied (Rodríguez-Rey *et al.*, 2015). We found that regardless of the data selection process, the shape of the curves and the variability is strongly influenced by the number of samples so the effect of sample numbers on the shape of the curve needs to be studied further.

Chapter 5 Sensitivity testing of the use of summed probability distribution functions in relation to the British database: Part 2 Sample size and ^{14}C mean relative uncertainty

5.1 Introduction

Previous analysis of the British database has worked on the assumption that 200 dates are enough to generate statistically reliable results and this assumption was used to justify the data analysis of only ‘change after’ dates to interpret the British database (Macklin *et al.,* 2012). Chapter 4 – and previous studies have determined that the number of dates used can affect the shape of summed PDFs and that in order to achieve a ‘reliable’ shaped curve (Williams, 2012; Michczyńska and Pazdur, 2004). Michczyńska and Pazdur (2004) suggested that for equally spaced datasets containing 1000 ^{14}C dates between 0 – 14,000 years, 200 dates are needed to generate a probability curve with statistical fluctuations under 50% and that 780 dates are needed to reduce statistical fluctuation to under 20%. These definitions were based on a randomly generated dataset of 1,000 radiocarbon ages that were equally temporally spaced and had a mean uncertainty of ± 115 years (Michczyńska and Pazdur, 2004). Williams (2012) determined that 500 dates are required to generate a statistically ‘reliable’ curve based on a total sample size of 2,000 radiocarbon dates with a mean uncertainty of ± 170 years. However, it has already been identified in Chapter 4 that there are 776 dates in total in the British database, the dates are not equally spaced over time, and the mean uncertainty of the database has not been defined. Therefore, the applicability of the assumptions of minimum sample number size is unknown.

The reliability of the shape of summed PDF curves determines if the occurrence of peaks could be attributed to environmental/ climatic causes and/or methodological /sampling errors (Williams, 2012). If the assumption that 200 dates is incorrect for British database this could infer that summed PDFs generated using data from the British database should not be treated as reliable indicators of past fluvial activity.

Another factor that could cause uncertainty in the shape of the curves is the radiocarbon uncertainty because different mean uncertainty values have been shown to cause the shape of curves to change, however this finding has not been applied to real datasets before for low uncertainty values (under ± 50 years) or low sample numbers (less than 1000 dates). The reanalysis of the British database offered the opportunity to study this because the mean uncertainty of the radiocarbon dates is ± 76 years and sub-datasets presented in Chapter 4 highlighted low sample numbers and variability over time.

The effect of mean radiocarbon uncertainty on sample size was discussed previously and it was observed that when a dataset of 500 dates was analysed, the statistical fluctuation was lower when the mean uncertainty was higher (Michczyńska and Pazdur, 2004). For example, when the mean uncertainty was ± 50 years, the associated statistical fluctuation was 40% and when the mean uncertainty was ± 200 years, the statistical fluctuation was 20% (Michczyńska and Pazdur, 2004). This infers that that mean uncertainty could influence the shape of probability curves and the number of dates required to generate statistically reliable results.

This chapter tests the sensitivity of the effect of the number of samples used per sub-dataset identified in Chapter 4 (‘change after’ dates, ‘non-change after’ dates, five or more dates per site, less than five dates per site, wood sample material and peat sample material), and then tests the sensitivity of the relationship between radiocarbon mean uncertainty and sample size on the location of peaks and troughs in each probability curve.

5.2 Methodology

5.2.1 Assumptions made in this study:

It is important to first define statistical reliability in the context of this study. Statistical reliability refers to the variability between a control dataset and sub-datasets of the control dataset. In this study, the control dataset is the unfiltered British database (Macklin *et al*., 2012). A low value of variability reflects good correlation between a sub-dataset and the full dataset; hence high statistical reliability. Statistical reliability is discussed based on definitions provided by Geyh (1980) for histograms, see Table 5.1 and the Monte Carlo experiment (Michczyńska and Padzur, 2004)

Table 5.1 Definition of different types of histograms based on statistical fluctuations (Geyh, 1980).

Type of histogram	**Statistical fluctuation**
Reliable histogram	Less than 20%
Common histogram	Between 20-50%
Unreliable histogram	More than 50%

5.2.2 Methodological approach

5.2.2.1 Sensitivity to sample number size

The following steps were taken to establish the sensitivity to sample number size.

1. 20 Sub-datasets were created out of the British fluvial database by using a random number generator for different total numbers of samples (n= 50, 100, 150, 200, 300, 400, 500, 600 and 700) creating 180 in total (n=20 x 9 sample numbers);
2. Each sub-dataset was calibrated using the methodology given in section 3.3.1 to create 20 summed PDFs for each value of n;
3. The mean squared error (MSE) was calculated for each 5-year increment within each sub-dataset using the methodology outlined in 3.3.4;
4. Then the summed mean square error (SMSE) was calculated by summing the MSE values per sub-dataset;
5. Steps 1-4 were applied to sub-datasets of the British database to generate 10 sub-datasets per sub-dataset, which created 280 sub-datasets. The sub-datasets are defined below:

    - ‘change after’ dates (n = 50, 100, 150, 200; 10 x 4 sample numbers = 40 sub-datasets);
    - ‘non-change after’ dates (n = 50, 100, 200, 400, 500; 10 x 5 sample numbers = 50 sub-datasets);
    - Over 5 dates per site (n = 50, 100, 150, 200; 10 x 4 sample numbers = 40 sub-datasets);
    - Under 5 dates per site (n = 50, 100, 150, 200, 300, 400, 500; 10 x 7 sample numbers = 70 sub-datasets);

- Wood sample material (n = 50, 100, 150, 200; 10 x 4 sample numbers = 40 sub-datasets);
- Peat sample material (n = 50, 100, 150, 200; 10 x 4 sample numbers = 40 sub-datasets);

6. As a tool for comparison a randomly generated dataset of 1,000 dates with uniform error ± 60 years were generated, which created an additional 10 sub-datasets;
7. In total, 470 sub-datasets were used to test the sensitivity of the shape of summed PDF curves to sample numbers.

5.2.2.2 Determining relationships between radiocarbon laboratory error and sample number size

1. The relative ^{14}C uncertainty values, which represents the radiocarbon laboratory machine error (± 'n' years), were classified as relative percentages of radiocarbon dates in the British database (for example, 1-2%, 3-4%);

2. Then the quantity of radiocarbon dates with associated relative laboratory error of the following percentages were counted and tabulated: ±0.0 – 1.0%, 1.1 – 2.0%, 2.1 – 3.0%, 3.1 – 4.0%, 4.1 – 5.0%, 5.1 – 10.0%, 10.1 – 15.0%, 15.1 – 20.0% and over 20%;
3. 19 sub-datasets were created using the British database based on the relative uncertainty values: 0.00 – 2.0%, 2.1 – 5.0%, 5.1 – 10.0%, 10.1 – 20.0% (^{14}C dates with relative errors over 20% were considered outliers and were excluded from analysis);

4. For each of the four sub-datasets, up to 50 further sub-datasets were created using a random number generator for different total numbers of samples (n = 50, 100, 200 …. 600);
5. Each sub-dataset was calibrated using the methodology provided in 3.3.1 to create 10 summed PDFs for each value of n;
6. Then the summed mean square error (SMSE) was calculated using the methodology given in 3.3.5;
7. In total, 190 sub-datasets were generated and these are defined below:

    - Relative error 0.0 – 2.0% for n = 50, 100, 200, 300 (10 sets of randomly generated numbers per 4 values of n = 40 sub-datasets);
    - Relative error 2.1 – 5.0% for n = 50, 100, 200, 300, 400, 500, 600 (10 sets of randomly generated numbers per 7 values of n = 70 sub-datasets);
    - Relative error 5.1 – 10.0% for n = 50, 100, 200, 300, 400, 500, 600 (10 sets of randomly generated numbers per 7 values of n = 70 sub-datasets);
    - Relative error 10.1 – 20.0% for n = 50, 100, 200, 300, 400, 500, 600 (10 sets of randomly generated numbers per 7 values of n = 70 sub-datasets);
8. As a tool for comparison, an additional 35 sub-datasets were constructed using a random number generator to create 1,000 dates with the same relative uncertainty values identified in step 3, with the addition of 0.00 – 0.5% and 0.6 – 2.0% classification. In total 385 randomly generated sub-datasets were created and are defined below:

- Relative error 0.0 – 0.5% for n = 50, 100, 200, 400, 600, 800, 1000 (10 sets of randomly generated numbers per 7 values of n = 77 sub-datasets);
- Relative error 0.6 – 2.0% for n = 50, 100, 200, 400, 600, 800, 1000 (10 sets of randomly generated numbers per 7 values of n = 77 sub-datasets);
- Relative error 2.1 – 5.0% for n = 50, 100, 200, 400, 600, 800, 1000 (10 sets of randomly generated numbers per 7 values of n = 77 sub-datasets);
- Relative error 5.1 – 10.0% for n = 50, 100, 200, 400, 600, 800, 1000 (10 sets of randomly generated numbers per 7 values of n = 77 sub-datasets);
- Relative error 10.1 – 20.0% for n = 50, 100, 200, 400, 600, 800, 1000 (10 sets of randomly generated numbers per 7 values of n = 77 sub-datasets).

5.3 Results

5.3.1 Sample size

The results presented below test the relationship between sample size and statistical variability of palaeoflood data using the British database and sub-datasets of the British database.

5.3.1.1 British database

Figure 5.1 shows the steady decline in the value of summed mean square error (SMSE) as the total number of dates increases. The curve flattens, and the rate of decrease is reduced when there is a minimum of 600 dates. Figure 5.1 indicates that the best match to the curve generated for the unfiltered British database is achieved when a minimum of 600 dates out of the total 776 dates are used. The collection of dates could have an assortment of different contextual characteristics, as identified in Chapter 4.

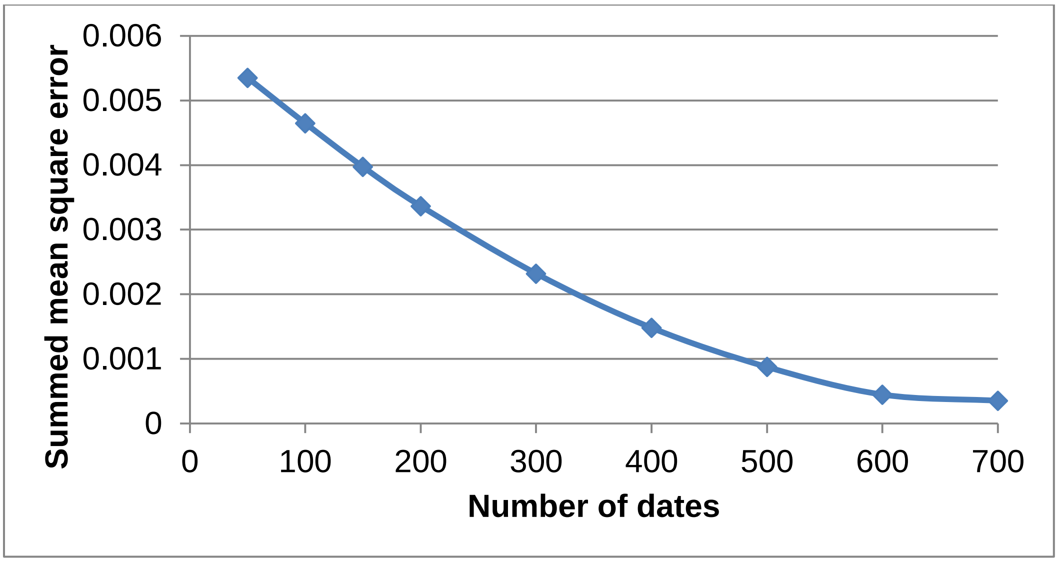


Figure 5.1 The variability between randomly generated sub-datasets (n=50, 100, 200….700) of the British database and the full British database (n=776).

5.3.2.2 Sub-datasets of the British database

Looking at the subsets of the British database, Figure 5.2A and 5.2B suggest that there is more statistical variability in larger datasets (over 500 dates) than there are in smaller datasets (under 200 dates). This is indicated by a higher SMSE value when n = 50 – 400 and the lowest SMSE occurs when the highest number of dates is included in analysis. This is indicative of SMSE as an indicator of statistical variance with a given dataset because it is expected that the least statistical variance is present when more data are used, and more statistical variance occurs when fewer dates are included. To achieve a relatively low SMSE value of around 1.00, 50 dates are needed out of 200 dates (25%), and around 400 dates are needed out of 600 dates (60%). To achieve the closest match to the curve of each sub-dataset, around 80% of the total numbers of dates are needed.

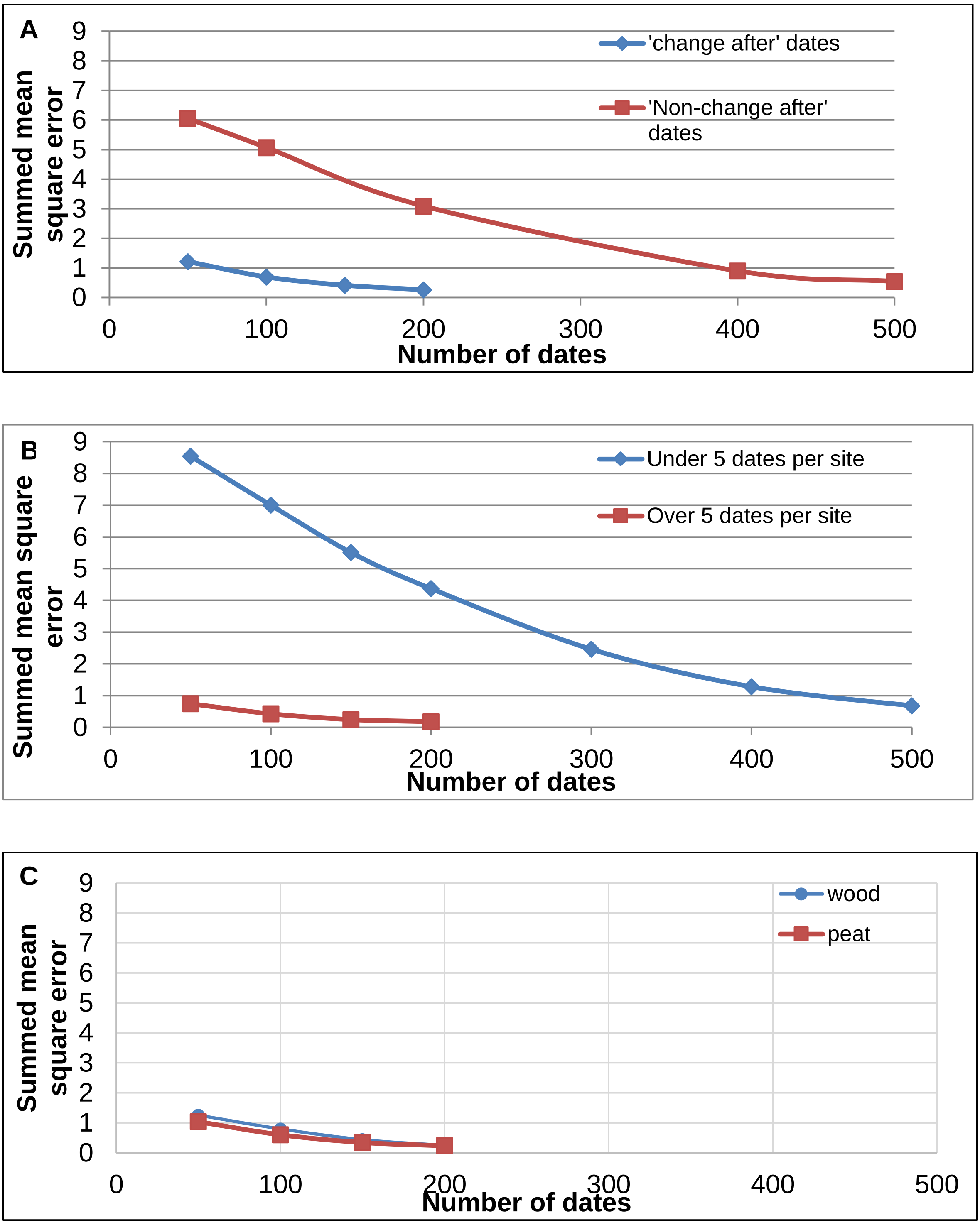


Figure 5.2 The variability of sub-datasets as determined by the summed mean square error of A: 'change after' dates, 'non-change after' dates; B: sites with five or more dates, sites with less than five dates; C: wood and peat sample material.

5.3.1.2 Simulated dataset

The range of SMSE values shown in Figure 5.3 is noticeably lower than the values observed in Figure 5.2. Regardless of this, there is a constant decrease in the value of statistical variability as the number of dates increases and the rate of the decrease in SMSE flattens when a minimum of 800 dates are used. Both of these findings are similar to Figure 5.2 and suggest that an important threshold for 80% of the numbers of samples required to statistically reflect a full dataset, regardless of the total number of dates or the relative SMSE value.

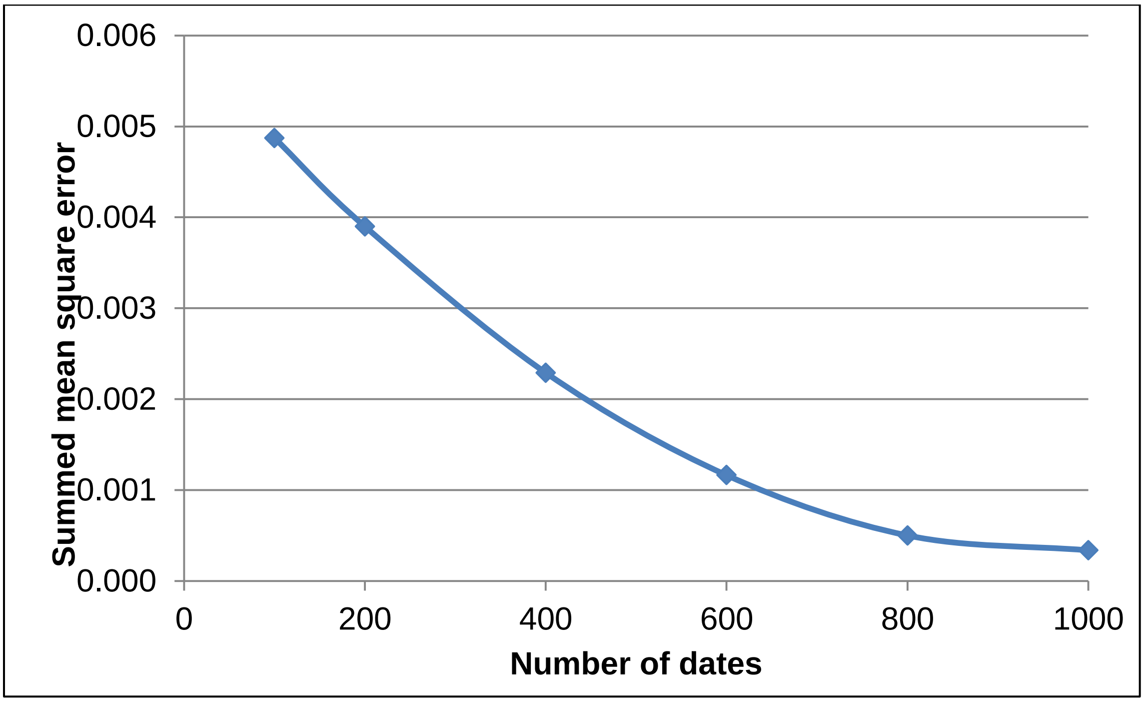


Figure 5.3 Summed mean square errors calculated for 1,000 randomly generated radiocarbon dates with uniform error ± 60 years.

5.3.2 Laboratory uncertainty

The next set of results examines the influence of relative radiocarbon laboratory uncertainty on the relationship between sample size and statistical variability.

5.3.2.1 The British database

Figure 5.4 shows that the value of relative uncertainty in recent radiocarbon ages is higher than in older radiocarbon ages (r^2 = 0.1435) with the exception of four outliers at 3,778 ± 4,140 years, 3,570 ± 3,851 years, 1,040 ± 945 years and 1,010 ± 923 years, where the radiocarbon uncertainty is between 90 – 110% of the ^{14}C age.

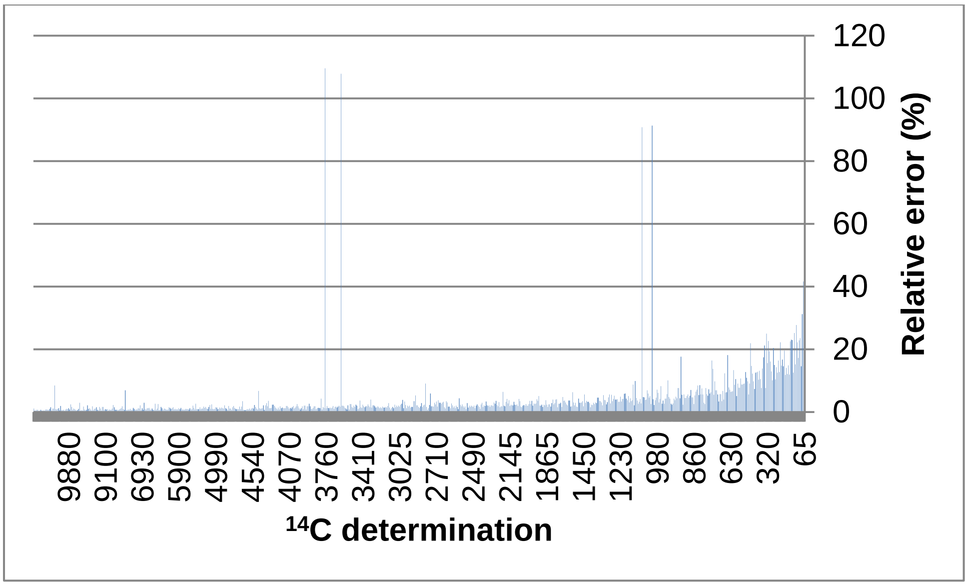


Figure 5.4 The relative uncertainty of ^{14}C ages within the British database shown as a percentage.

Table 5.2 shows that more than 50% of the ^{14}C ages within the British database have relative uncertainties that are under 2% of the value of the ^{14}C age. The lower the relative uncertainty, the higher the precision of the ^{14}C measurement based on laboratory equipment so the ^{14}C age is more reliable. This section tests the feasibility of using relative radiocarbon uncertainty as a factor of quality control.

Table 5.2 The number of ^{14}C ages classified with specific relative uncertainty values.

Relative uncertainty (%)	**Number of dates**
0 - 1.0	164
1.1 - 2.0	242
2.1 - 3.0	114
3.1 - 4.0	74
4.1 - 5.0	38
5.1 - 10.0	78
10.1 - 15.0	33
15.1 - 20.0	12
20.0 +	21

In terms of total number of dates available for analysis in the British database based on the classification of upper limits of relative ^{14}C uncertainty (as specified in Table 5.2), there would be a loss of 3% of dates if the value was set at 20%, 9% loss of dates if the relative uncertainty was limited to 10%, a loss of 19% of dates if the limit was 5% and a loss of 56% of dates if the limit was 2%. Clearly, the lower the value the more reliable the dates because they are associated with less laboratory uncertainty, as discussed in 2.2.1, however, it is unknown how setting a limit could affect a dataset; so this is explored next in relation to the British database and a randomly generated dataset for comparison.

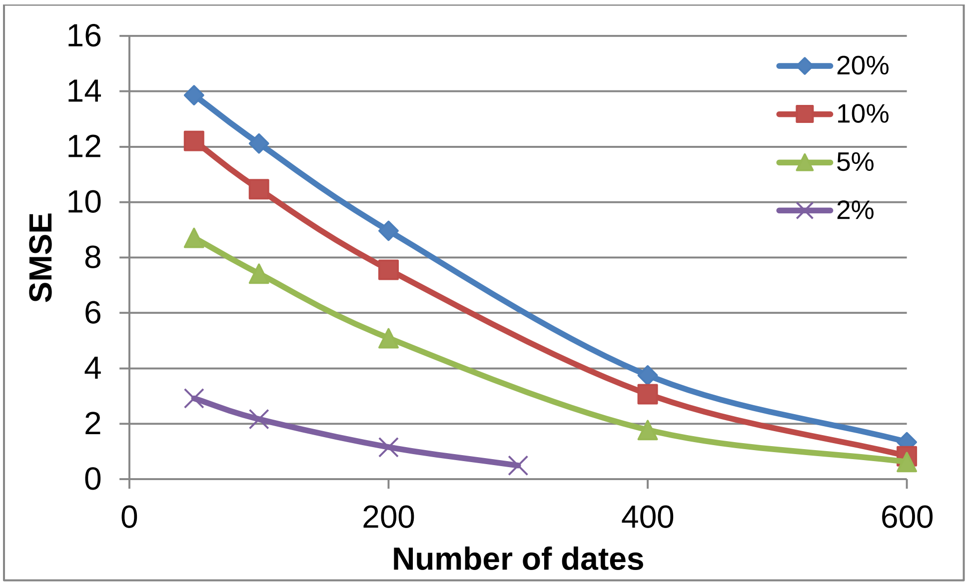


Figure 5.5 Statistical variation as shown my SMSE for 19 sub-datasets of the British database in relation to relative ^{14}C uncertainty values.

Figure 5.5 clearly shows that the statistical variability reduces when a higher number of dates are used for all values of relative uncertainty. The data also shows that when ^{14}C dates with a corresponding relative uncertainty value of less than 2% of the ^{14}C dates are analysed, the SMSE is significantly lower for all values of n used. This suggests that if the relative uncertainty is capped, for example at 2%, which results in a loss of 56% of the dates available, the overall statistical variability is lower, and it was shown in Figure 5.2 that there is statistically less variability when fewer dates are used and this observation is still valid when the relative uncertainty value was interrogated. This suggests that radiocarbon uncertainty could be influential in determining the number of dates required to generate a statistically reliable probability curve in relation to the British database; so for comparison the influence of relative ^{14}C uncertainty on the relationship between the number of dates used and statistical variability is tested on a randomly generated dataset.

5.3.2.2 Simulated dataset

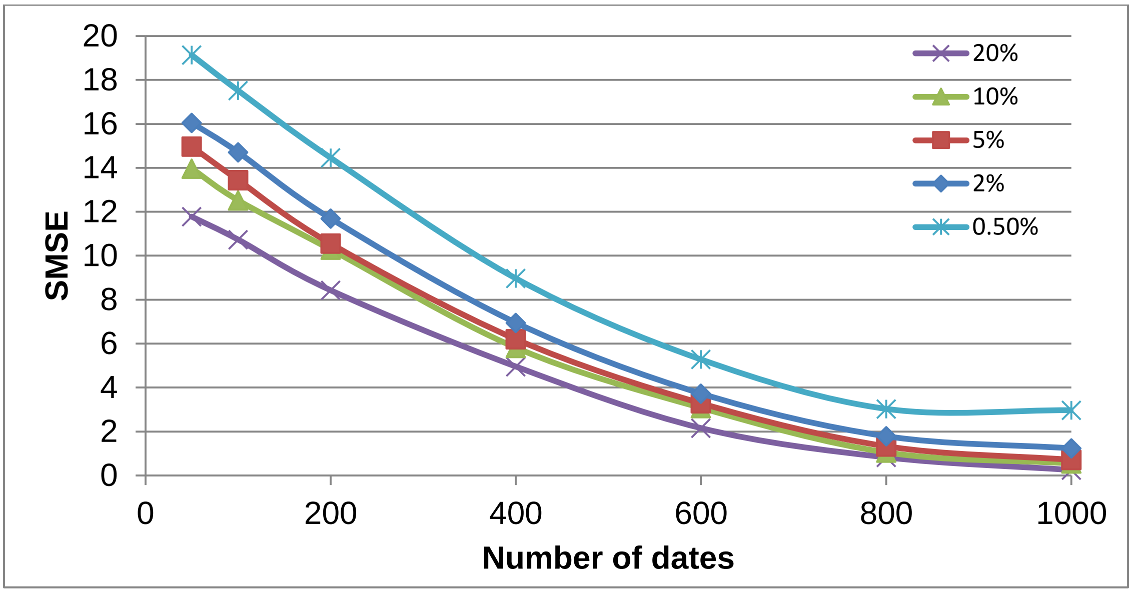


Figure 5.6 Plot of 1,000 randomly generated numbers with relative uncertainty values ± 0.00 – 0.50%, 0.51 – 2.00%, 2.10 – 5.00%, 5.10 – 10.00%, 10.10 – 20.00% for different values of n (n= 50, 100, 200, 400, 600, 800,1000).

Figure 5.6 demonstrates that when more dates are used the statistical variation decreases which is consistent with observations made based on Figure 5.5. However, Figure 5.6 suggests that there is greater statistical variability when the relative uncertainty is lower, which contradicts the findings in Figure 5.5. It can also be observed in Figure 5.6 that when 50 – 200 dates are analysed, the overall SMSE value is higher than in Figure 5.5. For example, when 50 ^{14}C dates with relative uncertainty from the British database is 2% or less, the SMSE value is 2.9, whereas when 50 ^{14}C dates with the same relative uncertainty are analysed out of 1000 randomly generated dates, the SMSE value is 16.00.

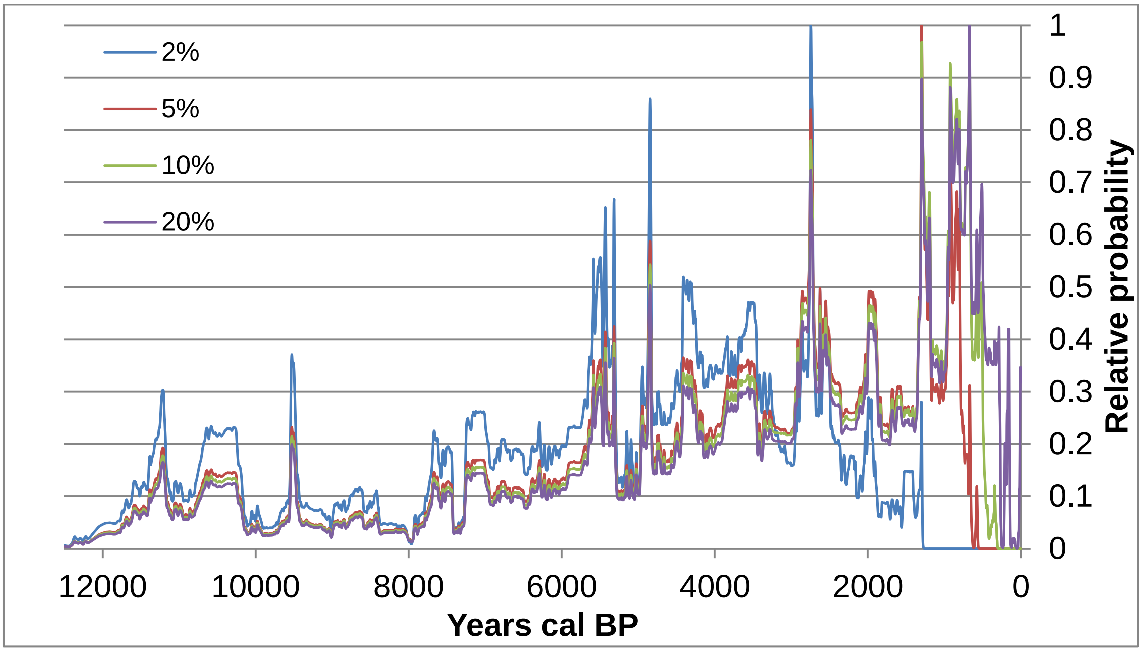


Figure 5.7 Summed PDFs for 4 sub-datasets of the British database classified by relative ^{14}C uncertainty (2% n = 406, 5% n = 226, 10% n = 78, 20% n = 45).

In terms of what the statistical analysis presented in 5.3.1 mean for the shape of the summed PDFs produced using sub-datasets from the British database based on relative uncertainty, it can be seen in Figure 5.7 that peaks are more pronounced when the relative uncertainty value is lower. This is because high relative uncertainty causes a wide but low distribution of calibrated ^{14}C dates, which results in a broad and flatter Gaussian curve; whereas when the relative uncertainty is lower, the individual Gaussian distributions of ^{14}C dates are more narrow and steeper and this is reflected in the summed PDF in Figure 5.7 as the relative uncertainty reduces.

Each of the curves presented in Figure 5.7 have different values of n but still follow the same general pattern; the shape of the curves are similar but the higher the relative uncertainty, the broader the shape of the curve and the lower the relative uncertainty, the shape of the curve is narrower. This observation is not consistent for ^{14}C ages younger than 1000 years cal BP. For example, for the curve generated by ^{14}C dates with relative uncertainty of less than 2%, there are no ^{14}C ages below 1275 years cal BP; when the relative uncertainty is capped at 5% there are no ^{14}C ages below 600 years cal BP; and when the relative uncertainty is limited to 10% there are no ^{14}C ages below 300 years cal BP. This suggests that the relative uncertainty is higher in younger ages than in older ages.

5.4 Discussion

The results in this study present the first sensitivity analysis of the British database by sub-datasets (e.g. sample material and the number of dates per site) to identify how the number of dates influences the statistical variability of the shape of summed PDFs. This study recognises that the total amount of statistical variability is relatively lower in smaller datasets (around 200 dates). This was also identified in 4.3 where it was shown that the amount of statistical variability reduces as the number of dates increase. Figures 5.2A, 5.2B and 5.2C consistently showed that better statistical matches are achieved when 80% of the total number of dates are analysed, when the total number is between 200 and 550 dates. This means that a sub-sample of n dates is capable of producing comparable curves to the full dataset when 80% of the total numbers of dates are used. Importantly, this pattern was also identified in the unfiltered British database (Figure 5.1) and the randomly generated dataset (Figure 5.3), which infers that this finding could be transferable to other datasets. This is also consistent with Michczyńska and Pazdur (2004) determination that 780 dates are required out of 1,000 dates to produce a probability curve with less than 20% statistical variability. This also shows that the analysis of ‘change after’ dates is statistically unreliable because 236 out of 776 dates is only 30% of the total number of dates within the British database.

The results presented in this study suggest that certain sub-datasets of the British database are considered statistically ‘reliable’ because the statistical indicators of variability are relatively low, but only when studied in isolation from the full unfiltered British database. This both supports the application of a quality control procedure suggested in Chapter 4 and confirms that there is a high level of variability and hence uncertainty present within the unfiltered British database.

This study also presents the first robust analysis of the influence of relative radiocarbon uncertainty associated with radiocarbon laboratory machine error, which is a factor that has been used in this study to determine the effect of sample number size on the statistical reliability of summed PDFs. The key finding from the results presented in 5.3.2 was that statistical variability is higher when the relative uncertainty is higher and this results in a broad low shaped summed PDF curve, whereas statistical variation is lower when the relative uncertainty is lower resulting in a narrow shaped curve with steep peaks. 5.3.1 identified that statistical variation is affected by the characteristics of sub-datasets, such as sample material, and Table 5.4 below highlights that the lowest relative probability occurs within the sub-dataset of ^{14}C dates measured from sites with five or more dates (which will be a variation of sample material, depositional environments and various other contextual characteristics), or peat sample material. This is an interesting find because Rodríguez-Rey et al. (2015) recommends excluding bulk organic material. This draws attention again to the issue of improving precision but reducing the accuracy of data analysis. Furthermore, the mean relative uncertainty of all the ^{14}C dates in the British database is lower than 3 of the sub-datasets tested, which highlights the important of assessing the mean relative uncertainty in sub-datasets defined by other characteristics that are identified to improve reliability of data.

Table 5.3 The mean uncertainty and total number of dates in sub-samples of the British database.

Dataset/ sub-dataset	Mean relative uncertainty (%)	Number of dates
Wood sample material	4.5	250
Peat sample material	2.9	225
Over five dates per site	2.1	219
Under five dates per site	45.0	550
'Change after' dates	3.9	236
'Non-change after' dates	4.3	531
Full British database	4.1	776

In previous studies, a uniform error has been applied to randomly generated ^{14}C dates to determine the effect sample numbers have on the statistical reliability of the shape of summed PDFs (Williams, 2012; Michczyńska and Pazdur, 2004) however, as discussed in 5.1, this is an unrealistic representation of ^{14}C uncertainty and therefore the relative uncertainty value provides a better representation. To put this into context, Williams (2012) applied a mean ^{14}C uncertainty value of ± 170 years, which to a young ^{14}C date of 1500 years would be 11% relative value, whereas to an older date of 10,000 years it would be much lower at 1.7%; and to compare with the British database, the mean relative uncertainty is 4.1%. Therefore, this study supports that relative uncertainty is an influential factor of controlling the statistical variance and hence the shape of probability curves. Williams (2012) recommended that higher sample numbers are required when datasets with low mean uncertainty are analysed in order to reduce statistical variability. The results from this study support this recommendation as shown in Figure 5.6. Based on the findings from this study and the recommendations of previous studies, this study does not support that 200 dates are sufficient to generate reliable summed PDFs is not applicable to the British database.

The findings presented in this study do not show a clear threshold where the statistical variability significantly reduces when a certain number of samples are analysed or when a specific range of mean uncertainty values are identified. This makes it difficult to provide a recommendation for the minimum number of samples required for datasets that have smaller mean uncertainties. However, it is clear from the results presented in 5.3 that each observed dataset is unique, even when a sub-dataset from a control dataset is analysed, so it may not be appropriate to base the statistical reliability of a specific dataset of generic thresholds.

This study recommends that the mean uncertainty of datasets should be reported, which was also recommended by Williams (2012), along with the relative mean (as a percentage) to be able to incorporate the influence that sample number size may have on the shape of summed PDF curves. Importantly, this could enable users of summed PDFs to reflect this information in their interpretations and it would also give confidence to the use of radiocarbon dates and the use of summed PDFs to have the uncertainties acknowledged and discussed.

Radiocarbon uncertainty represents chemical and physical errors associated with sample material used to measure radiocarbon (Bayliss *et al*., 2004). Both the ^{14}C determination and uncertainty value are input into statistical modelling to calibrate a ^{14}C age into a radiocarbon date (Bronk Ramsey, 2009). So, this effects the range of calibrated dates that are produced by probability curves and hence the location and size of peaks and troughs.

5.5 Summary

The sensitivity analysis of summed PDFs presented in this study is important because it shows that summed PDF are highly sensitive to data selection, the number of dates and the mean relative uncertainty, in relation to the British database. It was identified in 2.3.2 that the calibration curve also influences the shape of summed PDF curves and this leads onto the final part of this study. The most recent analysis of the British database applies a correction process to summed PDFs to account for the influence of the calibration curve. However, the correction process involves dividing a sub-dataset by the whole dataset to generate a relative probability plot and then periods of increased fluvial activity have been interpreted from the shape of the relative probability plot. However, this chapter has identified that when sub-datasets of the British database are statistically analysed relative to the unfiltered British database, statistical variability is higher than if the sub-datasets are analysed independently. The influence of the calibration curve and the use of relative probability plots are tested in the next chapter.

5.6 Conclusion

The results presented in this chapter show that the British database is sensitive to the mean radiocarbon laboratory error value; the lower the mean uncertainty, the higher statistical variability is present. The relative mean ^{14}C uncertainty has a direct effect on the shape of summed PDFs and so should be reported and considered when analysing groups of ^{14}C dates. As a result of the findings presented in this study, it is recommended that sub-datasets of the British database, for example using the criteria identified in Chapter 4, should be analysed in isolation from the unfiltered British database to avoid reintroducing variability into analysis. Furthermore, this study suggests that a minimum of 80%

of the total number of dates within a dataset/ sub-dataset should be included in analysis to achieve minimal statistical variability.

Chapter 6 Radiocarbon calibration curve and the use of relative probability plots as environmental/ climatic proxies

6.1 Introduction

There are many recommendations that have been made to address the issue of 'suck and smear' of the calibration curve on datasets of ^{14}C dates (Williams, 2012; Macklin *et al.,* 2012; Hoffman *et al*., 2008) as discussed in 2.3.2. For the British database, Macklin et al. (2012; 2010) applied a correction procedure, which was originally used by Hoffman et al. (2008), whereby the frequency distributions of sub-datasets (specifically, 'change after' dates) are divided by the corresponding frequency distributions of the full British database to normalise grouped data. The resulting frequency distribution is then used to construct a *relative* probability plot (RPP), that are characteristically similar to summed PDFs with time on the x-axis plotted with relative probability on the y-axis (see Figure 2.11). As identified in Chapters 4 and 5, the consistent analysis of 'change after' dates fail to address the potential influences of data selection, and statistical uncertainty associated with ^{14}C laboratory mean relative uncertainty, therefore the influence of data selection on the shape of RPPs is also unknown. The RPP correction procedure aims to reduce the influence of the calibration curve so that peaks in the shape of the curve that exceed the mean relative probability can be interpreted to reflect fluvial activity only (Jones *et al*., 2015; Macklin *et al*., 2012; Macklin *et al*., 2010). However, there has been no study that robustly compares correlations between peaks identified in RPP's and any patterns in the data used to construct the ^{14}C calibration curve. This step is important to establish whether or not RPPs do indeed remove the influence of the calibration curve as intended.

The link between fluvial activity and climate is important to hydrologists and geomorphologists because of the uncertainty in the understanding that global warming may have on the frequency and magnitude of future flood events. The shape of the curve in RPPs has previously been compared with other datasets to establish any fluvial response to past climatic changes (Macklin *et al.,* 2010; Hoffmann *et al*., 2008). The linking or correlation of proxy records is a commonly used method to assign causation to observed patterns in real datasets, but justifying causation of correlation is still a big concern (Swindles *et al.,* 2012) because supporting contextual information is often not available or not used to its full potential. If the calibration curve is still able to influence the shape of RPPs, then this questions the interpretations and reliability of RRPs and especially correlations made between fluvial datasets with other datasets. Therefore, it is key that more research is focused upon identifying and understanding uncertainty within statistical analysis of palaeoflood records to determine their viability to be used as reliable recorders of the link between climate and fluvial activity.

This chapter first tests the shape of curves of RPPs that are generated using sub-datasets of the British database. This enables us to compare the periods of time when the relative mean probabilities are exceeded to compare how data selection affects the shape of RPPs. Second, the periods identified in the RRPs are compared with steps and plateaus identified in the calibration curve during the Holocene to identify if there are any common periods. Last, this chapter applies the first frequency analysis to the British database and the data used to construct the calibration curve to identify and compare the occurrence of cyclicities that could influence the shape of RPPs.

6.2 Methodology

6.2.1 Relative probability plots

Relative probability plots were generated for six sub-datasets of the British database following the methodology outlines in 3.3.3 for the following sub-datasets:

1. Wood sample material (n = 250)
2. Peat sample material (n = 225)
3. Over five dates per single site (n = 219)
4. Under five dates per single site (n = 550)
5. ‘Change after’ dates (n = 236)
6. ‘Non-change after’ dates (n= 531)

The mean relative probability per sub-dataset was calculated and indicated on each figure. The number of dates per 200-year calibrated intervals was calculated following the methodology outlined in 3.3.4 and plotted on a secondary axis so that the shape of the curves could be commented on in terms of sample number availability. Using the procedure identified by Macklin et al. (2012), peaks that exceed the mean probability and are associated with three or more dates per 200-year calibrated BP time interval were classified as significant periods that could suggest an increase in fluvial activity. The locations of significant periods were compared for each sub-dataset so that the similarities and differences in the location of peaks and troughs in the shape of each probability curve could be identified.

6.2.2 Radiocarbon plateaus

Plateaus are sections of the radiocarbon calibration curve where the shape of the calibration curve remains flat (horizontal) for hundreds of years, and this causes

^{14}C determinations of single ^{14}C dates to be calibrated into a large range of ages (Williams, 2012). For example, a single ^{14}C age determination of between 2,500 – 2,400 years could be calibrated to a date with a range of over 300 years, resulting in a broad shaped probability curves. The shapes of RPPs generated using sub-datasets of the British database were studied to identify any ^{14}C dates that could be influenced by plateaus during the calibration process. To test this, the calibrated range of ages for the full British database was analysed (n=776) and compared with a randomly generated dataset (n=1,000), where:

1. 1,000 randomly generated dates were constructed using a random number generated between 0 and 12,500 with uniform error ± 70 years;
2. All ^{14}C ages were calibrated using Intcal13 in Oxcal 4.3 (see chapter 3 for full methodology);
3. The output that Oxcal generates provides the potential minimum and maximum calibrated age ranges;
4. The minimum and maximum calibrated ages for each ^{14}C age was plotted against each other;
5. Horizontal spikes are interpreted to represent plateaus in the calibration curve;
6. The difference between the minimum and maximum calibrated ^{14}C date ranges was calculated and plotted against the mean age range for each ^{14}C age to identify any systematic patterns within the age range data.

6.2.3 Radiocarbon steps

Steps are sections of the radiocarbon calibration curve where the shape of the curve is steep, which causes a long range of single dates to be 'sucked' into the curve, resulting in a small range of calibrated ages (Williams, 2012). For example, if a ^{14}C age between 2,400 – 2,200 years is calibrated, the resulting calibrated ages could be reduced to a range of only 100 years. Radiocarbon steps cause steep, narrow sections in the shape of probability curves. 6 sub-datasets of the British database, as identified in 6.2.1, were recalibrated using the methodology provided in 3.3.1.

1. The probability was plotted against the calibrated years BP;
2. Peaks were tabulated;
3. Peaks within summed PDFs identified in this study were compared with peaks identified in previous studies; it is assumed that common peaks are likely to be caused by systematic sampling errors rather than reflect actual environmental/ climatic events.

6.2.4 Lomb-Scargle frequency analysis

The Lomb-Scargle (L-S) Fourier Transform method was applied to the British database to test for cyclicities in the frequency domain. The L-S method was chosen because it is designed to be used on unevenly sampled data (Swindles *et al*., 2012; Mudelsee, 2010), L-S identifies harmonic signals and determines if they are significant by providing a probability output (Mudelsee, 2010) as discussed in Chapter 2.3.4.

Three datasets were analysed:

1. Frequency distributions of ‘change after’ dates used to create summed PDFs (Macklin *et al*., 2012);
2. Frequency distributions of ‘change after’ dates used to create RPPs (Macklin *et al*., 2012);
3. Residual $\Delta^{14}C$ dataset (Supplemental data, Reimer *et al*., 2013).

Residual $\Delta^{14}C$ was tested as ^{14}C data is used as a proxy for solar activity so residual $\Delta^{14}C$ reflects the change in the production of ^{14}C over time (Mudelsee, 2010), which is a key area of uncertainty and is associated with plateaus and steps in the calibration curve (Williams, 2012; Guilderson *et al.,* 2005).

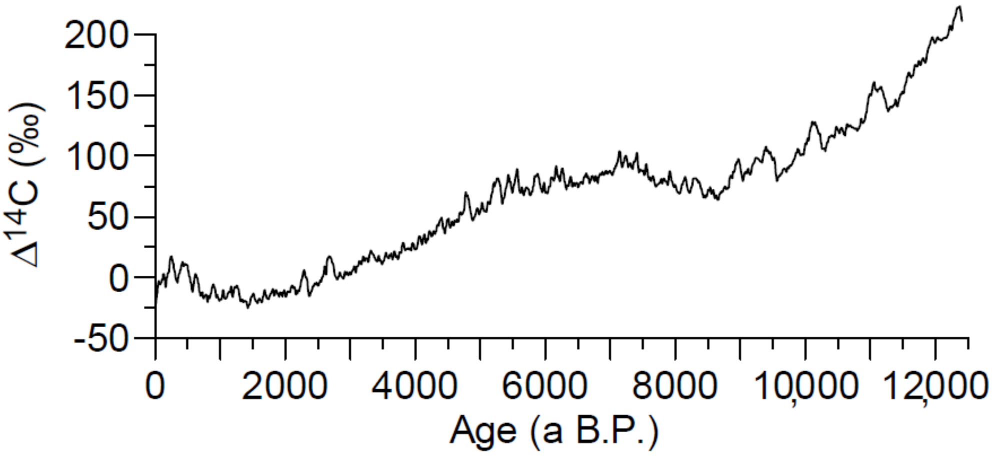


Figure 6.1 Delta 14C data measured from tree rings at 5 year intervals for 0 - 12,500 years BP (Mudelsee, 2010).

Residual $\Delta^{14}C$ ($d^{14}C$) represents the amount of depleted ^{14}C in a sample before isotopic fractionation correction (Stuiver and Polach (1977). Radiocarbon data is a proxy for solar activity so a high value of residual $\Delta^{14}C$ is interpreted to reflect low solar irradiance, and this can be used to check the presence of correlations between $\Delta^{14}C$ and climate proxy records (Mudelsee, 2010).

The following methodology was followed:

1. L-S frequency analysis was applied to each dataset using an edited version of the standard script provided by Trauth (2010) within MATLAB®. For all datasets the probability critical level is set at $\alpha = 0.05$. This means that anything below $\alpha = 0.05$ is likely to be significant; false alarm probability (FAP) means it is unlikely not to be a cycle;

2. X-axis set from 0 – 4,000 years BP for the observing window to exclude wavelengths that are similar lengths to the whole of the data that could be misinterpreted to reflect genuine cyclicities (Vanderplas, 2018);
3. The length of significant cycles within each of the dataset were compared;
4. The null hypothesis for the test is that all cyclicities are not significant.

Two simple questions are tested in this section of the study:

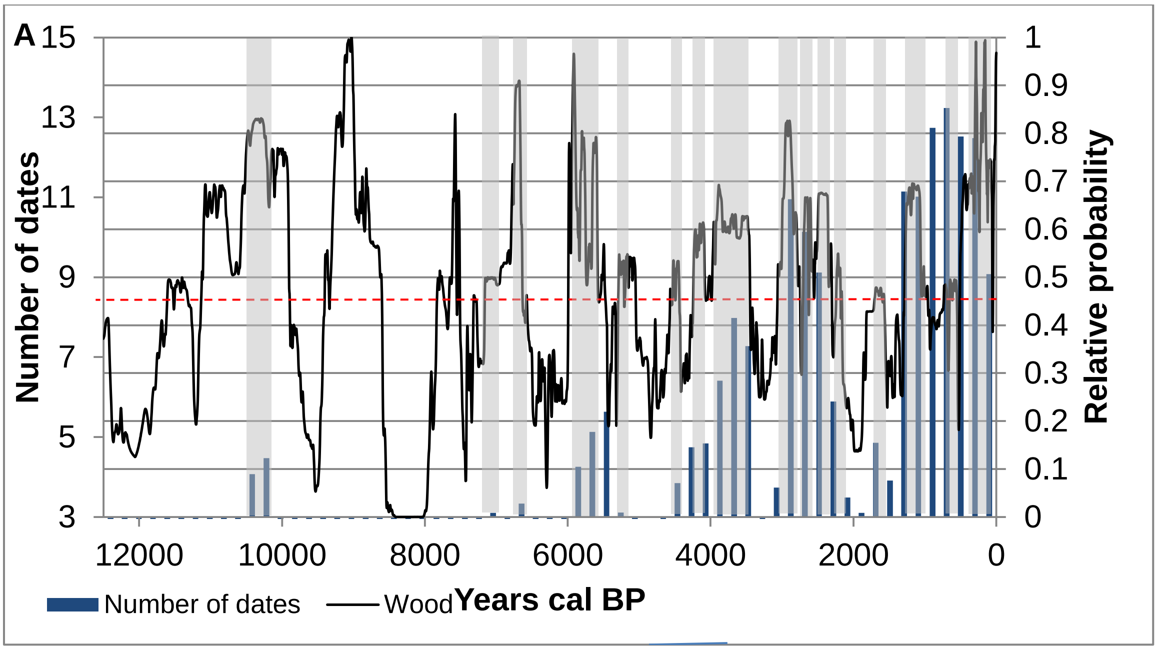
- if cycles are present in both the fluvial plot and the residual $\Delta^{14}C$ plot, are they likely to be caused by the influence of ^{14}C curve?
- if cycles only occur in the fluvial dataset and not in the residual $\Delta^{14}C$ plot, is it likely a signal reflecting something in the fluvial record that does not occur in the ^{14}C data?

6.3 Results

6.3.1 Relative probability plots

The graphs presented below show the relative probability plots generated per sub-dataset of the British database as identified in Chapter 4. Periods of time when the relative probability exceeds the mean relative probability and is associated with three or more dates in 200-year period in a sub-dataset are highlighted in grey.

6.3.1.1 Sample material



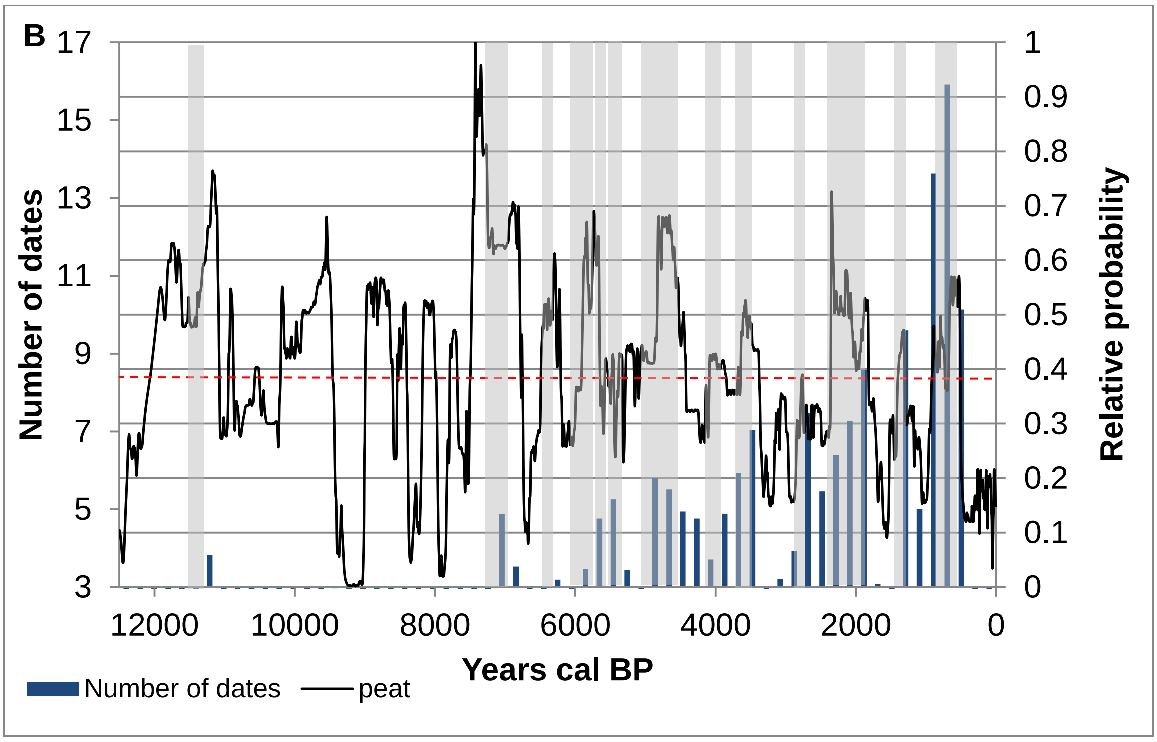


Figure 6.2 Relative probability plot of A: wood and B: peat sample material shown by the black curve. The number of dates per 200-year interval is shown in blue. The mean relative probability is shown by the dotted red line (wood = 0.46, peat = 0.39).

Figure 6.1A and 6.1B shows that there are 16 statistically significant periods that are identified in the wood sub-dataset and 13 statistically significant periods in the peat sub-dataset. Both datasets are similar as most of the periods that exceed the mean relative probability and are associated with three or more dates in a 200-year time interval are clustered between 7,000 to 0 years cal BP, with the exception of the peat record showing no significant periods between 400 and 0 years cal BP. Notably, the significant periods do not occur at the same time, for example between 5,000 – 4,400 years cal BP there is a significant period identified in the peat record but not in the wood record. There are also periods of time when there are three or more samples in a 200-year time interval, but the relative mean is not exceeded, for example in both datasets this occurs between 4,400 – 4,200 years cal BP, which could be related to the shape of the calibration curve. To contrast this there are also periods of time both records show the opposite pattern when the relative mean is exceeded when there are less than three dates per 200-year time interval, for example the peat record shows a relative probability below the sub-sample mean between 9,400 – 9,000 years cal BP but the wood record shows a relative probability above the mean. This could suggest that the unfiltered British database is influencing the shape of the RPPs when the correction procedure to address the influence of the calibration curve is completed, which would defy the point of the probability curve correction procedure.

6.3.1.2 Number of dates per single site

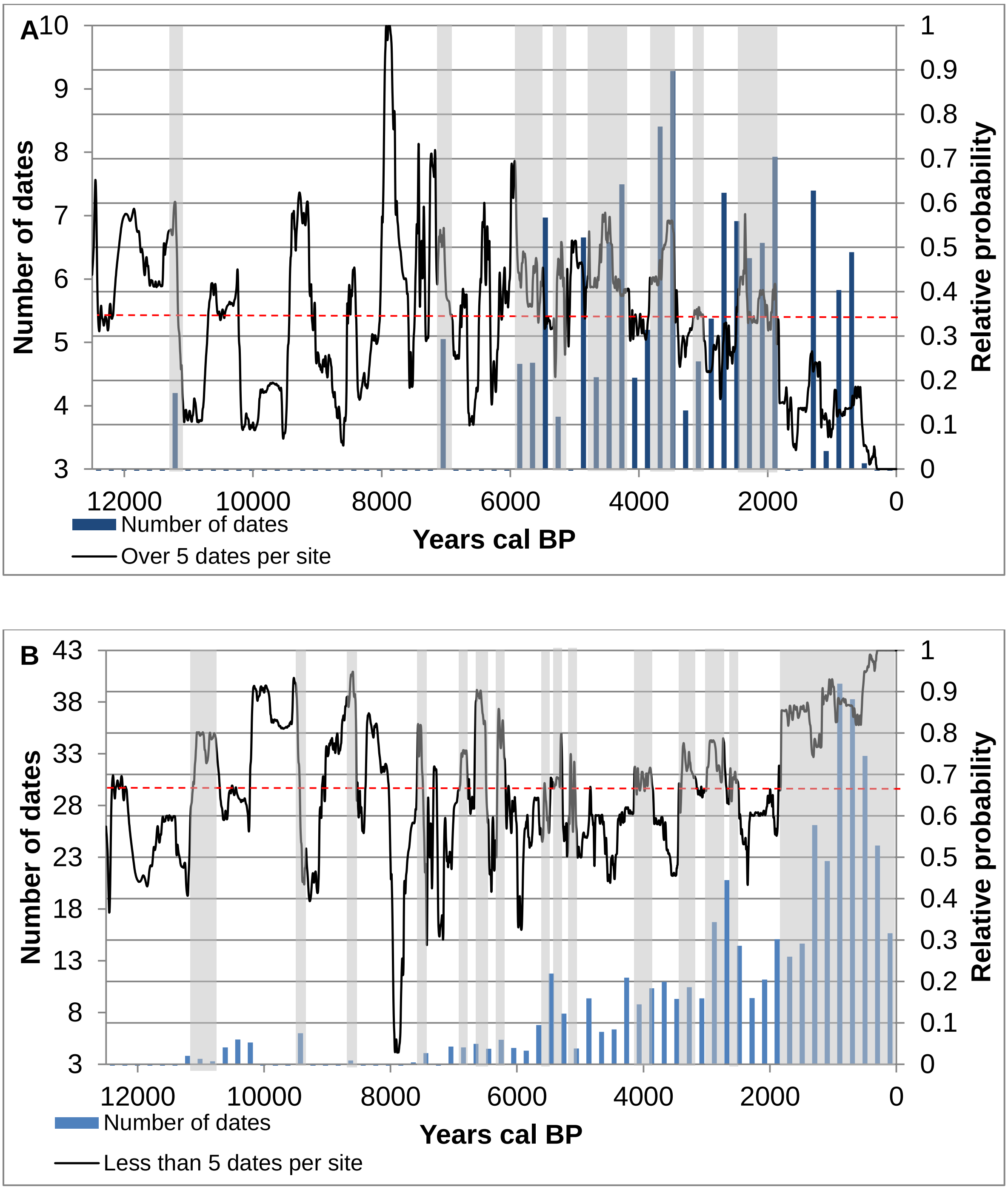


Figure 6.3 Relative probability plot of the collective analysis of A: more than five dates per single site and B: less than five dates per single site shown by the black curve. The number of dates per 200-year interval is shown in blue. The mean relative probability is shown by the dotted red line (> 5 dates = 0.33, < 5 dates = 0.67).

The collective analysis shown in Figure 6.2A and 6.2B highlight that the length of significant periods within each sub-dataset vary; Figure 6.2A is characterised by 8 statistically significant periods with lengths between 200 and 600 years, whereas Figure 6.2B is characterised by 15 periods, which consist of 10 periods that are 200 years long and 5 periods that are between 400 and 1,800 years long. There are also opposing trends that occur in the Early-Holocene as was observed in Figure 6.1; between 12,200 and 11,400 years cal BP, both records indicate there are less than three dates available per 200-year time interval yet the curve in Figure 6.2A exceeds the mean whereas the curve in Figure 6.2B does not exceed the mean. This suggests that the fluvial record is represented differently in each of the two datasets when analysed using RPPs, which was also observed in 4.3 when analysed using summed PDFs. This finding infers that the shapes of RPPs are sensitive to the data used to construct them in the same way summed PDFs are and the extent of influence varies between sub-datasets and possibly also sample number availability.

6.3.1.3 Association to a fluvial event

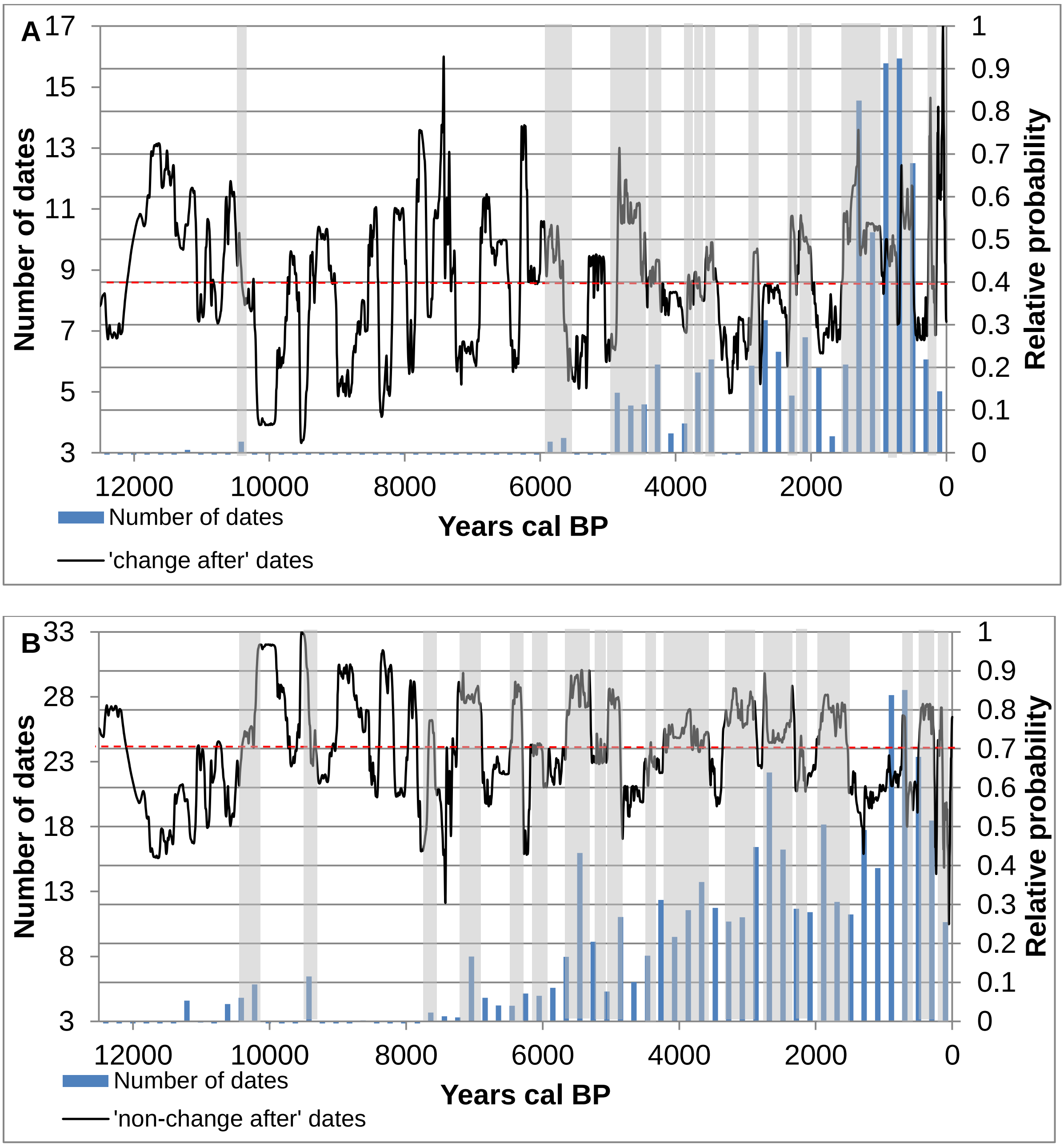


Figure 6.4 Relative probability plot of A: 'change after' dates and B: 'non-change after' dates shown by the black curve. The number of dates per 200-year interval is shown in blue. The mean relative probability is shown by the dotted red line ('change after' dates = 0.40, 'non-change after' dates = 0.70).

Figure 6.3A is characterised by a noticeable lack of data (less than three dates per 200-year interval) between 11,200 and 6,000 years cal BP and the shape of the curve fluctuations continuously above and below the relative mean probability. The periods when the mean is exceeded are not considered significant periods because they are not associated with enough data. Therefore, the peaks in the shape of the curve could be associated with the calibration curve or characteristics of the British database because there are more ^{14}C dates classified as 'non-change after' dates. Figure 6.3B does identify four significant periods during the same time interval because there are three or more dates per 200-years and the relative mean is exceeded. There are also more significant periods that are identified in the 'non-change after' dates than the 'change after' dataset (18 and 3 significant periods, respectively). The timing of significant periods tends not to occur at the same time in the two sub-datasets; hence again suggesting that the shape of RPP are vulnerable to the data used to construct them.

6.3.2 Radiocarbon calibration of palaeoflood data

This section tests how the process of calibration may influence the interpretation of palaeoflood data, which can be caused by the suck and smear effect of the calibration curve (as defined in 2.2.1.3), by comparing the location of plateaus and steps in the calibration curve with the ^{14}C dates available within the British database.

6.3.2.1 Radiocarbon plateaus

Each horizontal spike in Figure 6.5 represents the minimum and maximum calibrated date range of an individual ^{14}C age in the British database, and the larger the gap between the minimum and maximum dates are interpreted to reflect plateaus in the calibration curve.

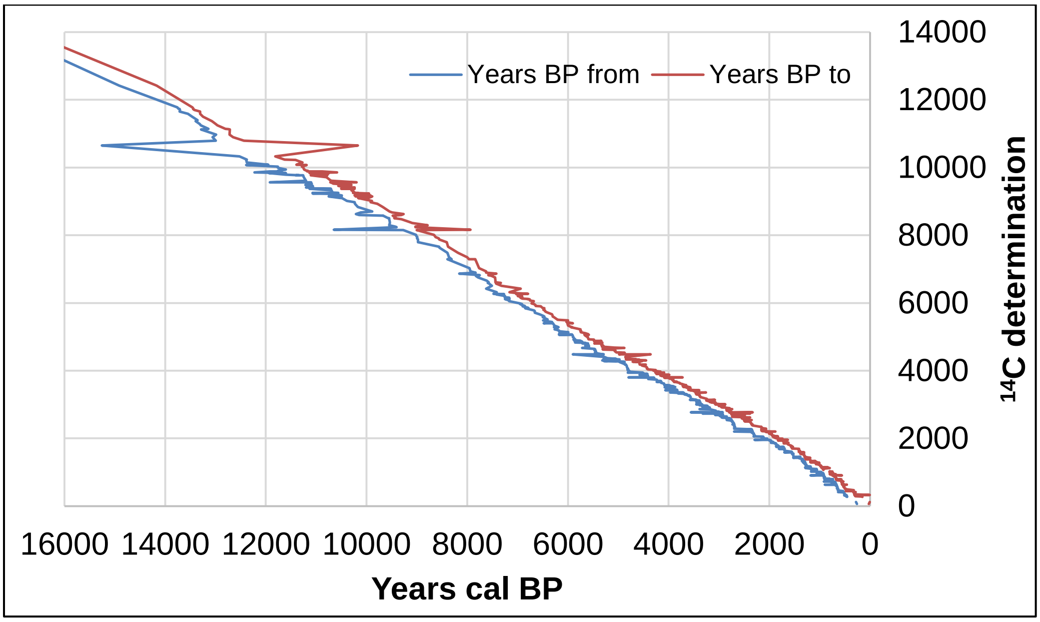


Figure 6.5 Plot of the minimum and maximum calibrated ^{14}C dates in the British database.

It is clear that the length of time between the minimum and maximum calibrated ages is not constant, which is likely to reflect of the variability within the calibration curve and the productivity of atmospheric ^{14}C. Figure 6.4 is characterised by two relatively large plateaus of samples with ^{14}C ages around 10,647 ± 900 years and 8,160 ± 560 years. These two ^{14}C ages have a relatively high uncertainty (in relation to the uncertainty values presented in Table 5.2) but not the highest. The highest uncertainty accompanies the following ^{14}C dates from the British database: 3,778 ± 4,140 years, 3,575 ± 3,851 years, 1,040 ± 945 years, and 1,010 ± 923 years, which suggests that a high range of potential ages is not influenced by the ^{14}C uncertainty, and so is likely to be influenced by the calibration curve.

The effect of plateaus in the calibration curve on the British database was tested by identifying significant periods identified in the RPPs of the sub-datasets presented in 6.3.1 that occur for the corresponding range of calibrated ages (^{14}C

10,647 ± 900 years = 15,259 – 10,174 years cal BP; ^{14}C 8,160 ± 560 years = 10,651 – 7,938 years cal BP) and identified that peaks in the shape of the RPP correspond with significant periods of fluvial activity in every sub-dataset tested (see Table 6.1 below). This provides further support that some peaks are attributed to the ^{14}C calibration curve, and so are not indicators of increased fluvial activity.

Table 6.1 The location of statistically significant peaks above the relative mean for sub-datasets of the British database in relation to two ^{14}C ages associated with plateaus in the calibration curve.

Sub-dataset	**^{14}C age 10,647 ± 900 years (15,259 – 10,174 years cal BP)**	**^{14}C age 8,160 ± 560 years (10,651 – 7,938 years cal BP)**
Wood	10,600 – 10,200	10,600 – 10,200
Peat	11,185	
> 5 dates per site	11,235	
< 5 dates per site	11,200 – 10,600	9,600 – 9,400, 9,545,
‘Change after’ dates	10,600 – 10,400	
‘Non-change after’ dates	10,600 – 10,200	9,470

Table 6.1 shows that there are overlapping significant periods and peaks in RPPs for both ^{14}C ages identified that have large calibrated age ranges (5,085 – 2,713 years) regardless of the sub-dataset used. This finding suggests that the shape of RPPs are influenced by the calibration curve in relation to plateaus even though the methodology used to generate RPPs was aimed at addressing this

issue. This finding can be studied further by comparing the range of calibrated dates produced for the ^{14}C ages in the British database with a randomly generated dataset to test if there are any correlations that could provide further evidence to support the influence of the calibration curve on datasets.

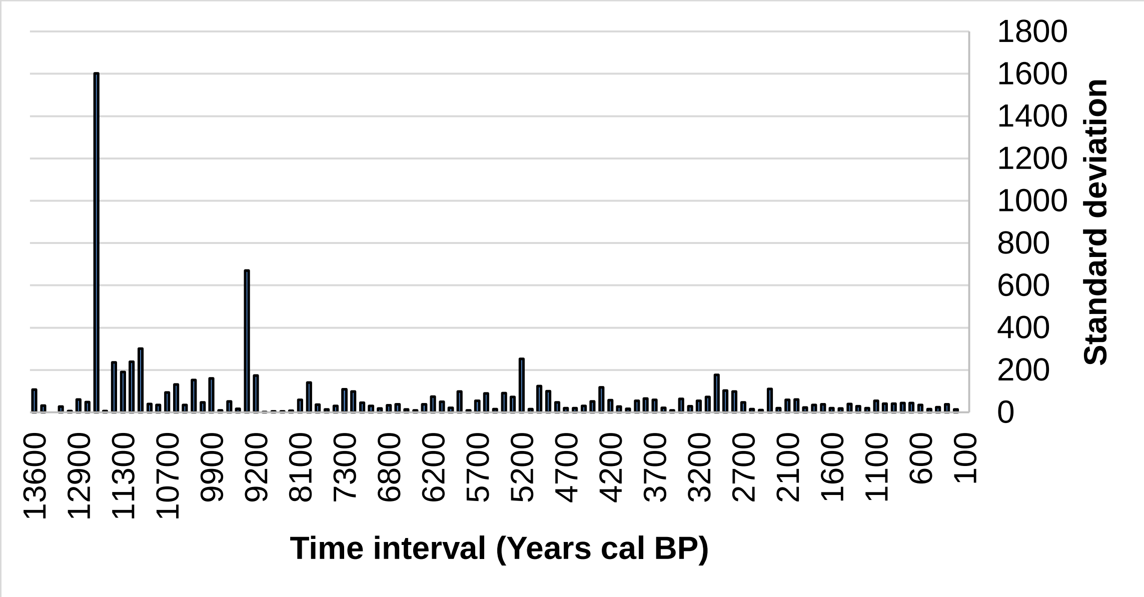


Figure 6.6 Standard deviation of the range of calibrated dates in the British database at 100-year intervals.

Figure 6.6 represents the similarities and differences between the ranges of ages produced during calibration for ^{14}C ages within the British database. To echo the observations in Table 6.1, there is a noticeable increase in the standard deviation of the range of ^{14}C dates that are calibrated to around 9300 and 12, 700 years cal BP. This is an important finding because it provides the first demonstration of how fluctuations in age ranges could be attributed to plateaus in the calibration curve in relation to the British database. This in turn reflects that a single sample cannot be represented by a single date and the effect of 'smearing' a single ^{14}C age can significantly alter the interpretation of datasets.

6.3.2.2 Radiocarbon steps

Williams (2012) and Michczyński and Michczyńska (2006) documented the occurrence of peaks that were linked to steps in the calibration curve. Peaks in the shape of the curve generated by summed PDFs were discussed in Chapter 4 and 5, so the results below tested to see if the documented peaks associated with steps in the calibration curve are present in the sub-datasets of the British database to further test the reliability and viability of the application of RPPs.

Table 6.2 Identification of peaks associated with steps in the calibration curve in comparison to significant peaks of sub-dataset of the British database. The presence of peaks at 11,220, 4,850, 2,750 and 520 years cal BP in each sub-dataset is marked with an 'X'.

Radiocarbon step peak (cal. BP)	**Wood**	**peat**	**> 5 dates per site**	**< 5 dates per site**	**'change after' dates**	**'non-change after' dates**
520 (Williams, 2012)		X		X	X	
2,750 (Michczyński and Michczyńska, 2006)	X	X		X		X
4,850 (Williams, 2012		X	X		X	X
11,220 (Williams, 2012)		X	X			

Table 6.2 identifies that the peaks identified in previous studies that are associated with steps in the calibration curve are also present in RPPs of sub-datasets of the British database, and the peaks occur during periods identified as significant based on a minimum of three dates per 200-year interval. All four peaks are present within significant periods of the peat record but only the peak

at 2,750 years cal BP occurs in the wood record. The collective analysis of sites with more than five dates contains the peaks at 4,850 and 11,220 years cal BP whereas the analysis of sites with less than five sites show the occurrence of the peaks at 520 and 2,750 years cal BP.

6.3.3 Lomb-Scargle frequency analysis

This study has established that the shape of probability curves, both summed and relative probability plots are strongly influenced by data selection *and* the shape of the radiocarbon calibration curve. This finding undermines previous studies use and interpretation of probability curves to reflect robust fluvial records because areas of uncertainty identified in this study have not been considered. Frequency analysis offers the opportunity to analyse data in the frequency domain rather than using probabilistic analysis of time series data. The resulted presented next show the first frequency analysis of the British database.

6.3.3.1 British database: ‘change after’ sub-dataset

The frequency analysis presented in Figure 6.6 identifies the presence of cyclicities in the frequency distribution created by ‘change after’ dates. The false alarm probability (FAP) is shown and peaks in power (Figure 6.6A) that correspond with probability under the FAP is classified as significant. The midpoint of each significant peak in power is also identified. The interpretation of significant cycles is discussed in 6.4.2 with reference to potential causal mechanisms.

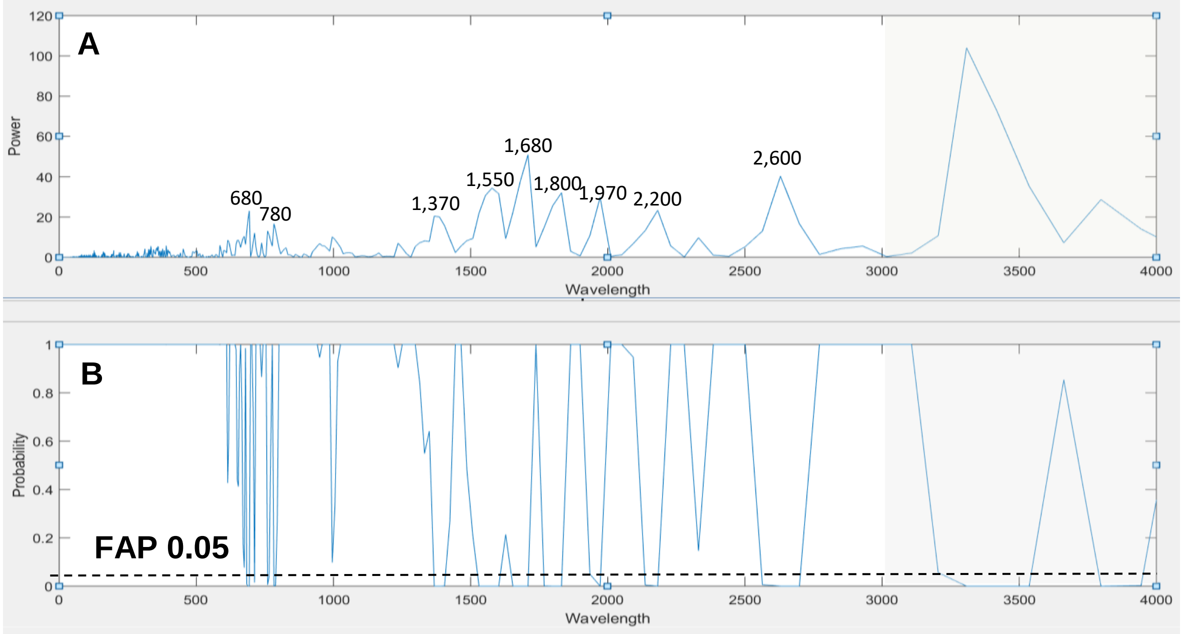


Figure 6.7 A: Power output of the Lomb-Scargle spectral analysis for 'change after' dates analysed using summed probability distribution function. Wavelengths greater than 3,000 years have been greyed out but included to show the 2 harmonic of the '1,500 year cycle'. B: Probability outputs for 'Change after' dates. FAP = α < 0.05 shown by the dotted black line.

Figure 6.6A shows a cluster of peaks with wavelengths between 1,300 – 2,600 years that pass the FAP shown in Figure 6.6B. The cluster of peaks in power are characterised by relatively broad peaks that appear to be connected because the power value does not return to 0.00 between peaking again. The cycles identified in Figure 6.6A have wavelengths that range from 1,370 – 2,600 years. The most noticeable peak in power in Figure 6.6A has a wavelength of 3,500 years but given that the full dataset spans 12,500 years, even though this peak passes the FAP, it is likely to be a representation of the length of the 'change after' dataset (4 x 3,125 = length of fluvial dataset). Therefore the c.3,500 wavelengths could be harmonics rather than a true reflection of a signal within the data.

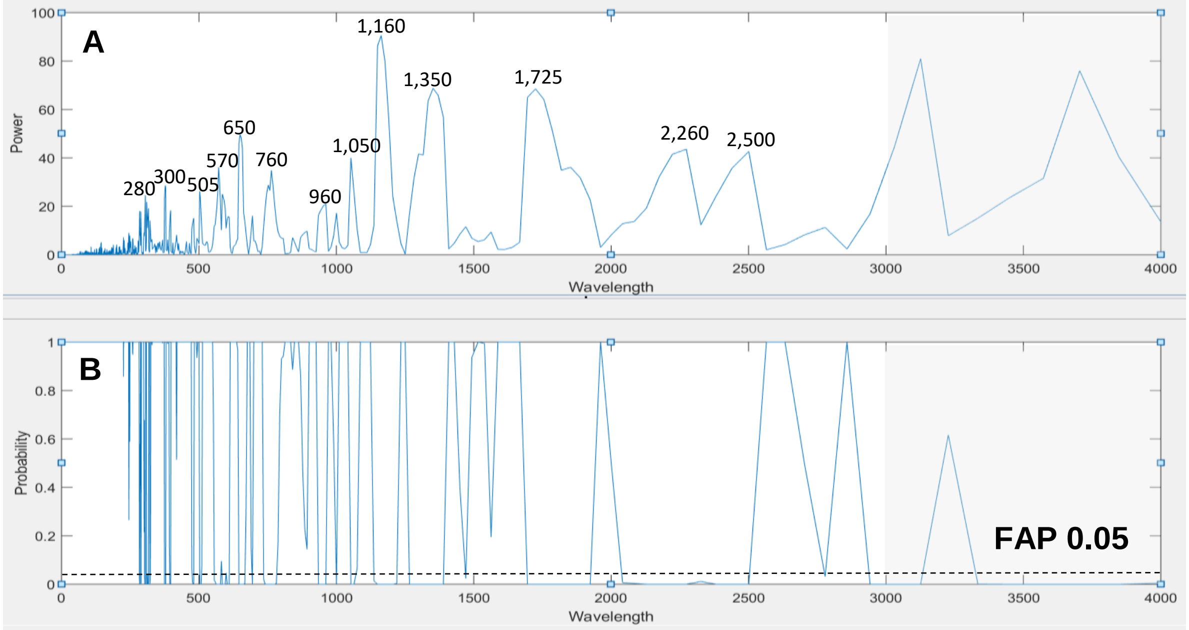


Figure 6.8 A: Power output for 'change after' sub-dataset constructed using the relative probability plot approach. Wavelengths greater than 3,000 years have been greyed out but included to show the 2 harmonic of the '1,500 year cycle'. B: Probability outputs for 'change after' data used to generate relative probability plots. FAP = α < 0.05 and are shown below the black dashed line.

Figure 6.7A identifies more significant cycles with shorter wavelengths than was identified in Figure 6.7A; there are 7 peaks in power that pass the FAP ($\alpha < 0.05$) that have a wavelength of under 1,000 years. Similar to Figure 6.6, there are two cycles present showing wavelengths of 3,100 and 3,500 years that pass the FAP but are also likely to be attributed to being ¼ length of the full dataset. Significant peaks occur in both datasets, which have wavelengths of around 780, 1,800, 2,200 and 2,600 years, and these are shown in Table 6.3.

The results presented in Figure 6.8 below show the occurrence of cyclicities within the residual $\Delta^{14}C$ dataset that is used to reflect the production of ^{14}C in the atmosphere to identify any correlations between the wavelengths identified in the fluvial data with the varying production rate of ^{14}C.

6.3.3.2 Residual $\Delta^{14}C$ dataset

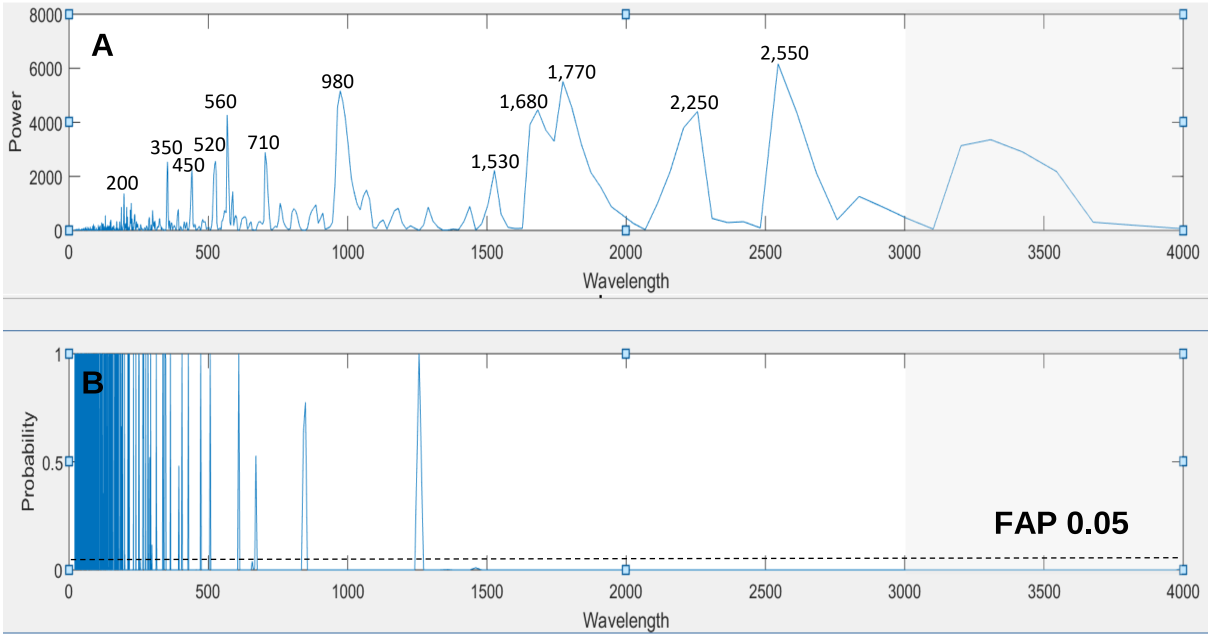


Figure 6.9 A Lomb-Scargle power output for the residual $\Delta^{14}C$ dataset. Wavelengths greater than 3,000 years have been greyed out but included to show the 2 harmonic of the '1,500 year cycle'. 3b Wavelengths that pass the false probability test ($\alpha < 0.05$) are shown below the black dashed line.

There is a noticeable difference in the power value within the residual $\Delta^{14}C$ data presented in Figure 6.8 and the 'change after' sub-dataset; the maximum power is ~6,000 and ~100 respectively. Figure 6.8 is characterised by 12 significant cycles (α = 200 – 2,550 years). Several of the significant peaks with shorter wavelengths (α = < 1,000 years) are characterised by sharp, narrow peaks which contrasts the peaks observed in the fluvial data. The significant cycle of around 3500 years in both of the fluvial frequency analyses can be recognised in Figure 6.8 by a very broad peak with wavelength of 3,300 years. The length of the residual $\Delta^{14}C$ data was included up to 16,000 years BP so it would be acceptable to also interpret this cycle to reflect the length of the dataset included in analysis. Another difference between the fluvial dataset and the residual $\Delta^{14}C$ data is that the FAP is passed for all wavelengths between 1,250 and 4,000 years.

6.3.3.3 Comparison of solar and palaeoflood data

The comparison of significant cyclicities in Table 6.3 shows four cycles occur in all three records (λ = ~750, ~1,750, ~2,250 and ~2,550 years). Six cycles present in 'change after' sub-dataset, used to construct summed PDFs, are also present in the residual $\Delta^{14}C$ record and nine cycles occur in the 'change after' dataset used to construct the RPP. This suggests that there is a correlation between the fluvial data and the residual $\Delta^{14}C$ record. However, it is surprising that there are more corresponding cycles in the RRP dataset because the methodology was designed to counteract the influences of the calibration curve, which should result in less corresponding cycles. There are two cycles that are only present within the fluvial datasets (λ = ~650 and ~1,350 years), which could suggest a periodicity that is unique to the fluvial dataset.

Table 6.3 Significant cycles identified in Lomb-Scargle frequency analysis using the frequency distributions of 'change after' data used to construct summed PDFs and RPP (Macklin et al., 2012), and in the residual Δ^{14}C record (Reimer et al., 2013). Significant peaks that occur in both statistical analyses of the fluvial data but not in the residual Δ^{14}C record are shown in bold green and significant cycles that are present in all three datasets are shown in red.

'Change after' dates summed PDF	**'Change after' dates RRP**	**Residual Δ^{14}C**
	280	200
	300	350
		450
	505	520
	570	560
680	650	
780	760	710
	960	980
	1,050	
	1,160	
1,370	1,350	
1,550		1,530
1,680		1,680
1,800	1,725	1,770
1,970		
2,200	2,260	2,250
2,600	2,500	2,550

Furthermore, one cycle occurs only in the 'change after' summed PDF sub-dataset (λ = ~1970 years) and two cycles occur only in the 'change after' RPP sub-dataset (λ = ~1,050 and ~1,160 years). Figure 6.9 below provides a comparison of the shape of the curves generated when the RPP of 'change after' dates is graphed with the residual Δ ^{14}C over time.

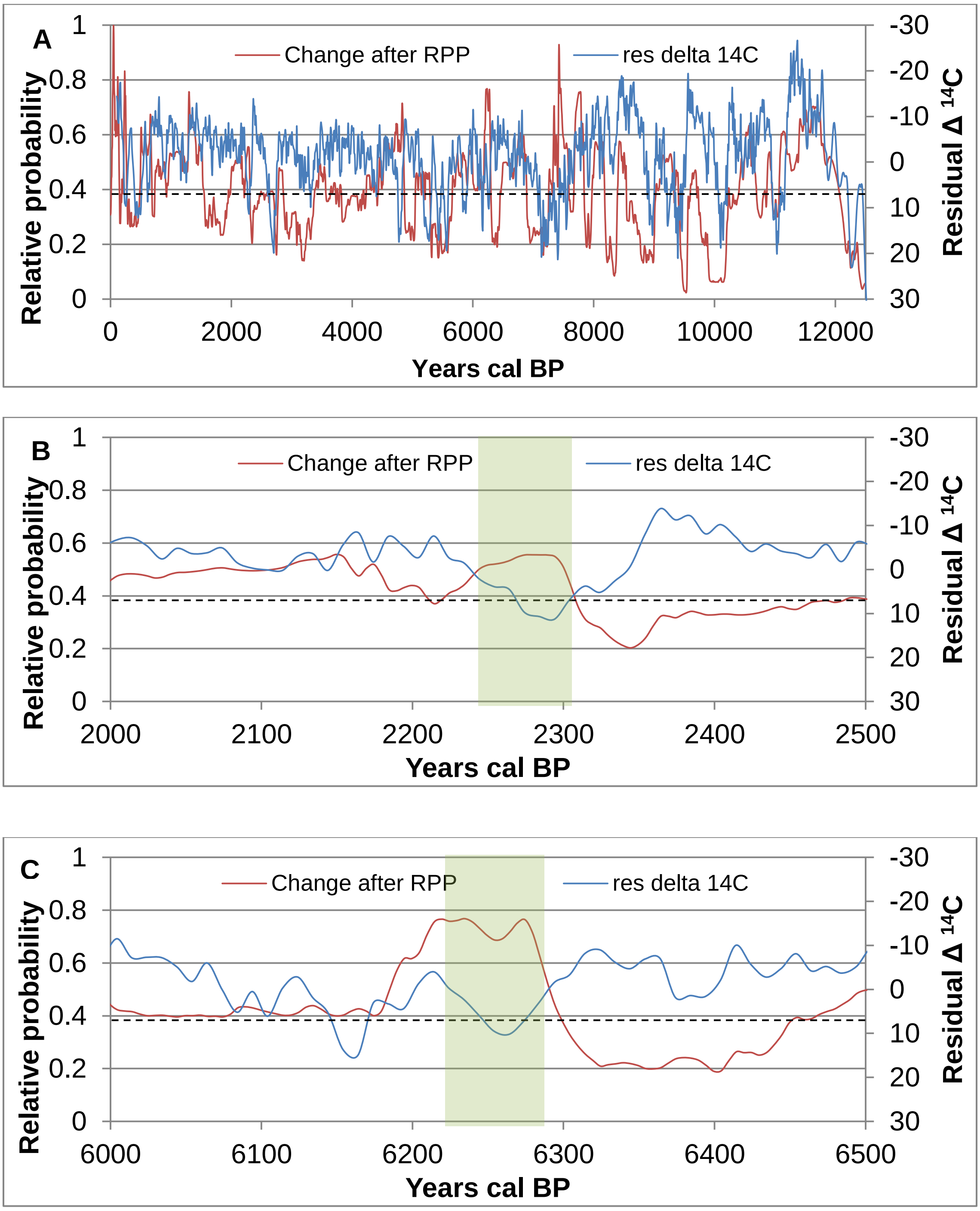


Figure 6.10 A: 'Change after' relative probability plot data plotted with residual Δ ^{14}C data for comparison. The dashed line represents the average relative probability (0.39) for the sub-dataset of the British database. B: same as A for 2,000 – 2,500 years cal BP, and C: for 6,000 – 6,500 years cal BP. Green shaded boxes highlight periods of increased probability of fluvial events that correspond with high residual Δ^{14}C.

Figure 6.10A-C shows the relationship between the residual $\Delta^{14}C$ record and the British database when represented as a RPP. The shape of both curves fluctuate over time and the general pattern that can be observed is that when probability reflected in the shape of the RPP is high, residual $\Delta^{14}C$ is low. Figure 6.10A highlights that during some time periods high probability within the British database corresponds with relatively high residual $\Delta^{14}C$ Figure 6.10B and C). The results presented in Figure 6.10A-C suggest that the data used to infer fluvial activity from the British database could correspond with low solar irradiance (high residual $\Delta^{14}C$).

6.4 Discussion

6.4.1 The viability of relative probability plots

The sensitivity of the shape of RPPs in relation to data selection was studied before the influence of the calibration curve was tested, because in previous chapter's data selection was identified as a key influencer in the shape of summed PDFs and quality control. It was identified in 6.3 that the shape of the curves in RPPs generated using different sub-datasets from the British database are sensitive to the data used to construct them, and this affects the location of peaks and hence the interpretation of timings of increased fluvial activity. These findings question the viability of RRPs as an effective and reliable method of analysis for groups of ^{14}C dates, especially when used to interpret the timings of events.

Previous studies have recommended that the shape of probability curves should be presented with the number of dates so that peaks can be interpreted in consideration with sample number availability (Williams, 2012; Benito *et al.*, 2008; Michczyńska and Pazdur, 2004). In 6.3.1, the relative probability identified significant peaks that exceeded the relative probability mean and related to at least three samples within a 200-year time interval. The threshold of a minimum of 3 dates per 200 years is consistent with the assumption that 200 dates are required to generate a summed probability curve with a reliable shape that has statistical fluctuations under 50% (Michczyńska and Pazdur 2004). However, if a higher threshold were applied based on the recommendation that 500 – 780 dates are required to generate a curve with less than 20% statistical fluctuations, this would equate to 7 – 11 dates per 200 years. There is currently not enough data available to be able to analyse the British database by sub-datasets to use the above method.

The findings presented in Table 6.2 (in 6.3.2) show how the shape of summed PDF curves generated using different quality control data selection criteria affect the location and presence of peaks. Four peaks, which were identified to be associated with steps in the radiocarbon calibration curve, occur in the sub-dataset of peat sample material but only one peak (at 2,570 years cal BP) is present in the dataset of wood sample material. This could potentially reflect the lack of preservation of wood sample material compared with the consistent nature of the development of peat rather than reflecting fluvial activity. Additionally, peaks that occur in the late-Holocene are not present in the sub-dataset defined by individual sites containing five or more dates but are present in the sub-dataset defined by individual sites containing less than five dates; and vice versa for peaks in the early-Holocene. The RPPs showing the dates with and without association to a potential fluvial event (‘change after’ dates and ‘non-change after’ dates) show conflicting evidence; the peak at 4,850 years cal BP is visible in both records and the peak at 11,220 years cal BP is not present in either record. In order to improve the viability of fluvial sedimentary records, more research is needed on the relationship between the calibration curve and the British database to enable more reliable interpretations of the shape of both summed PDFs and RPPs. Furthermore, the results from Table 6.2 clearly show that the presence of peaks is strongly influenced by sample material and data availability.

The process of defining a threshold whereby peaks in the RPP were only accepted as significant if they were associated with three or more dates reduced the number of RPP peaks interpreted as significant and hence periods of increased fluvial activity. To show the effect of using a 3-date minimum threshold, Table 6.4 compares with and without 3 date minimum threshold. Table 6.4 shows that the lack of dates in the Early-Holocene results in no significant periods being identified between 12,200 – 6,000 years cal BP, whereas in the analysis presented by Macklin et al. (2012), six significant periods are identified. This difference in interpretation could be attributed to there not being enough dates within the 'change after' sub-dataset during this time period to enable the classification of significant periods. However, there are several dates available that are not classified as 'change after' dates that could be causing peaks in the shape of the curve that result in the height of the peak exceeding the mean relative probability and hence being identified as significant when sample numbers are not integrated into the interpretation. Furthermore, Jones et al. (2015) interpreted the RPP generated using 'change after' dates with the addition of 68 new dates (16 are classified as 'change after' dates) and identified 2 additional significant periods (6,200 – 6,000 and 4,500 – 4,300 years cal BP) based only on the shape of the curve exceeding the relative mean probability; that demonstrates that the addition of new data changes the shape of the curve. Therefore, the shape of the curve will continue to change with the addition and/ or removal of data and we recommend that a formal sensitivity analysis of the effect of sample number size on the shape of RPPs is carried out using the approach used to test summed PDFs in this study.

Table 6.4 Comparison of the significant periods in the 'change after' sub-dataset identified by peaks above the mean probability that are associated with at least 3 dates per 200-year interval with periods identified only by exceeding the mean probability.

'Change after' analysis in this study based on a minimum of 3 samples per 200-year interval	**'Change after' date analysis by Macklin et al. (2012)**
11,400 – 11,200	11,800 – 11,100
	10,700 – 10,400
	9,400 – 9,100
	7,800 – 7,700
	7,600 – 7,300
	6,900 – 6,500
	6,300- 6,200
6,000 – 5,600	6,000 – 5,700
	5,300 – 5,100
4,900 – 4,500	4,900 – 4,500
	4,300 – 4,200
3,700 – 3,600	
3,500 – 3,400	3,600 – 3,400
2,900 – 2,800	2,900 – 2,800
2,300 – 2,200	2,300 – 2,200
2,100 – 2,000	
1,500 – 800, 600 - 500	1,500 - 500
150 - 0	300 - 0

Our findings demonstrate that compared to the unfiltered British database, data selection strongly influences the shape of RPPs. This finding is supported by the results presented in 4.3 and 5.3 that determined that sample number size strongly influences the shape of probability curves. This leads to the conclusion that the generation of RPPs *re-introduces* uncertainty in relation to data characteristics, which have been shown to influence the shape of probability curves. Furthermore, by statistically making the sub-datasets relative to the unfiltered record all of the uncertainties that we have hereto identified thus propagate into the shape of RPPs. Therefore, this study recommends that sub-datasets should be statistically analysed in isolation from the unfiltered record in order to achieve the most objective interpretation.

6.4.2 Identification of cycles within the fluvial record

As previously discussed, the interpretation of RPPs is based on the assumption that the shape of the curve is not influenced by the calibration curve and so peaks can be associated with fluvial events rather than methodological error (Jones *et al.*, 2015; Macklin *et al.*, 2012; 2010). However, the L-S analysis carried out in this study (6.3.3) shows some similar cycles within the fluvial and the residual $\Delta^{14}C$ records. Residual $\Delta^{14}C$ data is used as a proxy for solar activity with high residual $\Delta^{14}C$ reflecting low solar irradiance (Mudelsee, 2010), therefore we suggest many of the RPP peaks in the British database actually reflect solar activity rather than fluvial activity. The influence of solar activity on the occurrence of floods has been suggested in other studies for example, Wirth et al. (2013) suggest that the frequency of flooding in the Alps increases when solar irradiance is low, which was inferred to suggest that flooding is more likely to occur during cooler periods; and Vaquero (2004) proposed that flooding episode that were identified in the Spanish fluvial database (Benito *et al*., 2003) between 1,900 – 1,100 years AD could be reflecting solar activity.

We can also compare the cyclicities we identified in the British database to other studies. For example, Swindles et al. (2012) provides a comprehensive comparison of periodicities found in UK environments, including the 200-year sunspot cycle and the 1,500-year Bond cycle. Both of these well documented cycles are present within the residual $\Delta ^{14}C$ spectral analysis but are not present in spectral analysis of the British database. This suggests that certain spectra are not powerful enough within the data itself to power through because there is too much 'noise' in the British database to see anything meaningful, which again supports the need to apply a robust quality control methodology.

Importantly, not all significant cycles that are present within the spectral analysis of the British database correlate with significant cycles in the residual Δ ^{14}C spectral analysis. Interestingly, a cycle identified in peat in Northern England (λ = ~1,100 years; Swindles *et al.*, 2012) correlates with significant cycles within the British database (λ = 1,050 and 1,160 years). Figure 6.11 below plots the RPP of the 'change after' sub-dataset from the British database and highlights blocks that measure 1,100 years in length to test if the cycles identified in the L-S frequency analysis are also present in the shape of the RPP curve.

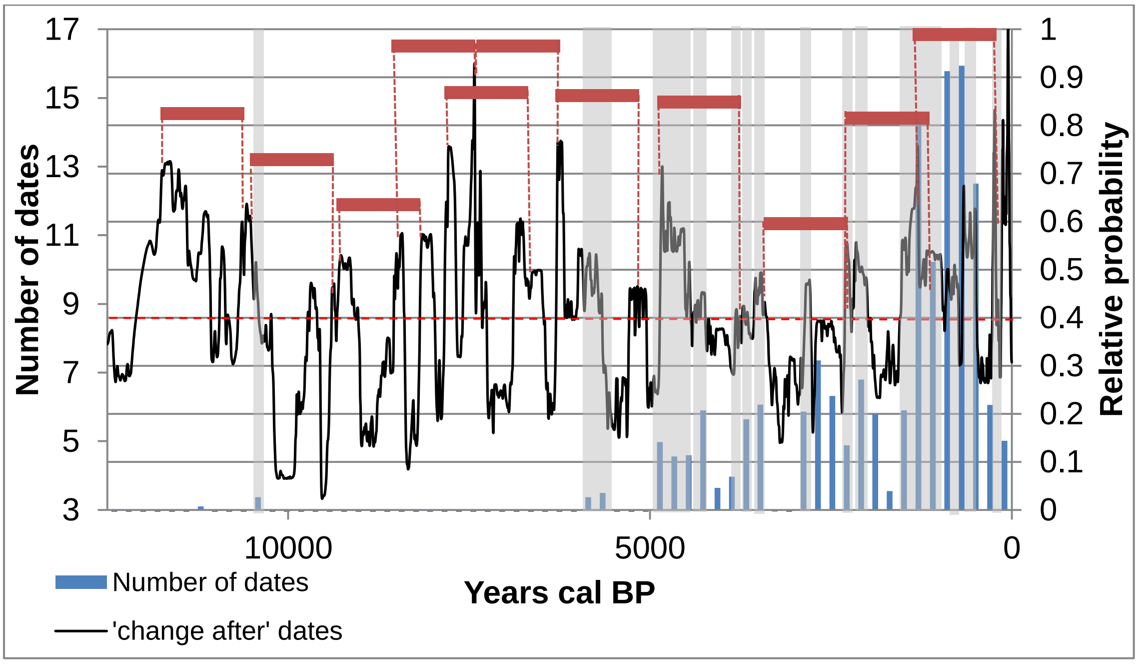


Figure 6.11 Relative probability plot of 'change after' dates showing significant periods (grey boxes) and peaks occurring every 1,100 years (red boxes). Relative mean probability is shown by the red dashed line and the number of dates per 200 years are shown by the blue histogram.

Figure 6.10 identifies several periods of time when peaks in the shape of the RPP curve occur around 1,100 years apart. Four of these periods overlap with periods of time identified as significant in this study based on the shape of the curve exceeding the mean relative probability during periods when there are three or more dates available, and by Macklin et al. (2012): 5,000 – 3,800, 3,600 – 2,400, 2,300 – 1,300, and 1,300 – 250 years cal BP. There are other overlaps that occur over 'non-significant' time periods and this leads back to the poor interpretation of probability curves when additional factors are not considered holistically. To the author's knowledge, this is the first direct comparison of cycles that have been identified in frequency analysis to probability curves and given the limitation of both the British database and the construction of probability curves, it is reasonable to conclude that much more research is needed into firstly identifying cyclicities in the shape of probability curves, and secondly, the relationship between solar irradiance and fluvial activity.

Frequency analysis of the British database suggests that the data is more sensitive at allowing shorter wavelengths to power through in the RPP dataset. However, there is also the possibility that cycles that occur with the fluvial dataset are a random occurrence because the collective use of single dates could introduce bias to the analysis (Carleton *et al.,* 2014). Carleton et al. (2014) apply a method to incorporate Bayesian analysis used to calibrate dates and to use *priori* information available to lake samples; however, this method is not directly transferable to the UK fluvial dataset because the database is dominated by single dates rather than collections of dates from single sites. Again, highlighting the limitations of the British database.

6.4.3 Summary

This chapter showed that data selection can drastically alter the shape of RPP's and therefore their interpretation as found for summed PDFs in 5.4. Data selection is affected by sample number, which also affects the shape of summed PDF curves. The sub-datasets characterised by wood and peat sample material, 'change after' dates, and the collective analysis of more than 5 dates per single site contained 219-250 dates, which were shown to reflect the frequency distribution of the unfiltered British database poorly. Whereas the sub-datasets characterised by less than 5 dates per single site and 'non-change' after dates contained 531 – 550 dates and showed a statistical better match to the unfiltered British database. These results suggest that to shape of RPPs are also influenced by the process of data selection and recommends that a comprehensive study is carried out based on sample number size and the shape of RPPs to put more confidence in the use of RPPs and to understand the mechanisms that affect it to better aid interpretations.

Furthermore, a comparison of steps and plateaus in the ^{14}C calibration curve can be matched to peaks in RPP's, for example the ^{14}C dates of 10,647 ± 900 years and 8,160 ± 560 years overlap with plateaus in the calibration curve; and previously reported steps in the calibration curve also occurred within some of the sub-datasets of the British database. Therefore, some peaks associated with steps and plateaus may have been falsely attributed to flooding periods.

Finally, use of the Lomb-Scargle frequency analysis for the first time on fluvial data reveals many RPP peaks are linked to the residual $\Delta^{14}C$ data, implying that they are a result of solar activity influencing the ^{14}C record, rather than increased periods of fluvial activity. However, some significant cycles, which have wavelengths of 650 - 680 and 1,350 – 1,370 years, are presently not attributable to any other source, so may well represent widespread changes in fluvial activity.

6.5 Conclusion

This chapter concludes that the fluvial record, as represented by relative probability plots, shows the occurrence of significant cycles that are also present in the residual $\Delta^{14}C$ data, which reflects the variation in radiocarbon production. Therefore, some peaks in the British fluvial database RPPs are directly influenced by the calibration curve. However, there are also significant cyclicities that occur in the fluvial dataset and not in the residual $\Delta^{14}C$ plot, which is likely to be a signal reflecting the something in the fluvial record. It is quite possible these peaks can reflect events, climatic periods or a different methodological error, potentially associated with data selection. Put bluntly, the British database does not necessarily capture signals of climatic and or environmental cycles as has been suggested before (Macklin *et al*., 2012; 2010). Further research on the use of Lomb-Scargle frequency analysis could represent an excellent way to develop the way that palaeoflood data are interpreted.

Chapter 7 Synthesis of thesis

7.1. Overview

The British database is a collection of 776 ^{14}C dated fluvial units that are currently available to extend the palaeoflood record in British catchments for the Holocene (Macklin *et al.,* 2012). Previous analysis of the British database has applied two pioneering techniques of data selection and data analysis to interpret the record. Firstly, the classification of 'change after' dates to filter the data that is more likely to reflect the timing of a potential fluvial event (Jones *et al*., 2015; Macklin *et al*., 2012). Secondly, the development of relative probability plots to combat the influences association with the calibration curve on the shape of summed probability distribution curves (Jones *et al*., 2015; Macklin *et al*., 2012). This study presented the reanalysis of the British database following on from critiques of these methods, (Williams, 2012; Chiverrell *et al*., 2011a, b), to test the sensitivity and viability of aggregated palaeoflood records and to examine whether they respond to environmental change – or something unrelated.

7.2 Response to research questions

Chapter 1 posed three research questions this Thesis aimed to answer:

1. How sensitive is the British database to the characteristics of large datasets, for example the type of sample material or the number of samples per single site?
2. How well are sub-datasets of the British database represented by frequency distributions of groups of radiocarbon dates?
3. Does the palaeoflood record show a response to past climatic changes?

Chapter 4 addressed question 1 directly looking at how important it could be to incorporate the characteristics of individual samples into analysis rather than treating the samples as only ^{14}C dates. Failing to do this could result in the loss of important additional sample context that could help the interpretation of a palaeoflood record (Chiverrell *et al.,* 2011). This study presented the impact of using several different methods of 'quality criteria', which included the filtering of 'change after' dates - that resulted in the loss of 70% of the ^{14}C dates available. This cropping of specific selection of data led to an expansion of the original question, asking if there enough data currently available in the British database to construct reliable results and make sound interpretations? Previous studies had indicated that 200 dates are required to generate summed PDFs that are reliable enough to be interpreted to reflect past environmental/ climatic changes (Michczyńska and Pazdur, 2004) and their assessment was used to justify the selection of 'change after' dates (Macklin *et al*., 2012). The results presented in this study determined that the shape of summed PDF curves is highly sensitive to the addition and removal of data, which questions the viability of previous analysis and subsequent interpretations of the British database.

For research question two, throughout this study, the statistical reliability of the shape of summed probability distribution functions has been tested. Influences on the shape of summed PDFs have been discussed in previous chapters. It is clear from the findings in Chapter 5 that a single ^{14}C age, which is used for ^{14}C dating, cannot be represented by a single date, but instead a range of ages. The range of dates that are produced during the process of ^{14}C calibration fluctuates, like the calibration curve itself. Steps and plateaus in the calibration curve affect the range of ages produced as shown earlier in this chapter. These findings concur with previous studies that calibrated ^{14}C dated material should not be represented by a single age, for example the mean or median, but should be presented with the full calibrated range of dates and associated ^{14}C error.

Additionally, the novel use of the Lomb-Scargle spectral analysis allowed us to address question 3 to determine whether (a) there was any cyclicity in the British database and (b) if this was linked to the ^{14}C calibration curve and dates calculation. By identifying common cyclicities in the British database and the residual Δ ^{14}C data (used to construct the IntCal13 calibration curve; Reimer *et al.*, 2013) it is evident that there is a strong correlation between cyclicities in both datasets. This is true for both statistical approaches: the summed probability distribution functions and especially for relative probability plots, where 9 cycles with similar wavelengths were observed. Many previously identified peaks in the fluvial record may be directly related to the influence of solar activity on the residual Δ ^{14}C data – in effect a relic in the method, not due to environmental drivers. The Lomb-Scargle analysis also identified a weakness of relative probability plots, that uncertainty is re-introduced during the production of these plots by making sub-datasets relative to dataset with high statistical variability. The interpretation of spectral analysis of fluvial data provided a novel alternative method of data interpretation – identifying a handful of peaks that may well indicate periods of increased flooding - but it is recommended that more research is needed on this topic to have confidence in its application to palaeoflood records.

Concerning question 3, this study suggested that a method of quality control should be applied to the British database in order to verify the ^{14}C ages that are used to interpret past events. This study suggests that when the total number of dates within a dataset is between 200 and 1,000, at least 80% of the dates need to be included in analysis to achieve the least statistical variability from the full dataset; and hence generate a summed PDF that has a curve of a similar shape. However, it has already been identified that the ‘true’ shape of the curve that could represent the occurrence of past flood events is unknown. Furthermore, if sub-datasets of the British database are studied relative to the unfiltered record then we shouldn’t necessarily try to achieve a close match to the unfiltered record because it may not be a true reflection of past events.

Chapter 4 also identified where variability occurs by classifying the dates per characteristic and the results from this chapter suggest that low statistical variability can be achieved by analysing the sub-dataset completely separate to the unfiltered British database. This has implications for the most recently method of analysis that creates a relative probability plot curve by dividing the probabilistic values of a sub-dataset (‘change after’ dates) by corresponding probabilistic values of the unfiltered British database (Jones *et al*., 2012) because this study suggests that variability is re-introduced into the dataset during this stage (Chapter 6). Therefore, this study demonstrates that the British database is very sensitive to the characteristics of data used *and* to the number of samples used.

Furthermore, palaeoflood records come in different types and there are many different ways to analyse and interpret them. It was recommended in 5.4 that sub-datasets should be analysed in isolation from the British database. The results presented in 6.3 provide further evidence to support this as relative probability plots can bring back in the (possibly erroneous) characteristics of the complete unfiltered British database, that can in turn affect the shape of the RPPs.

7.3 Relevance of research and contribution to science

The results of this study are highly relevant to researchers using summed probability distribution functions to interpret environmental and climatic data in any field. This study shows that summed probability distribution functions constructed using palaeoflood data are influenced by the number of samples per site, sample material, sample material context and sample material association. In particular, the recommendations made here regarding the application of a robust quality control criterion is the first of its kind and demonstrates that underlying methodological errors are capable of influencing the shape of summed probability distribution functions. The findings of this study could greatly increase the confidence in other analysis of databases of ^{14}C dates palaeoflood data, for example in Poland (Starkel *et al.,* 2006), Germany (Hoffmann *et al*., 2008) and Spain (Thorndycraft and Benito, 2006). This is critical if all of the databases want to be able to produce a record of palaeoflood frequency in relation to past climate. As Hoffmann et al. (2008) suggested, each database could be used to 'fill in the gaps' between different records.

The increasing use of summed probability distribution functions to represent ^{14}C dates has produced an increasing number of ^{14}C dated meta-datasets and this study is also relevant to users of ^{14}C dated samples. This research increases the knowledge base and understanding of factors associated with ^{14}C laboratory error that influence the shape of probability distribution curves. The methodology behind ^{14}C dating, including the calibration curve, is constantly developing so it is essential that the methodology used to analyse and interpret the data does as well.

For British Holocene fluvial geomorphology, it has been previously suggested that British fluvial deposits respond immediately to climate change and are sensitive hydroclimatic proxies (Macklin *et al.,* 2010). For example, the relationship of palaeoflood data to North Atlantic Ocean (NAO) oscillations during the Holocene (Macklin *et al.,* 2012). In other studies, linking fluvial events with Holocene climatic variation in Polish Rivers, peat bog data is compared with other environmental factors by 'wiggle-matching' with records, such as lake water level, speleothems and changes in vegetation (Starkel *et al.,* 2006). The Tagus River in Spain is correlated with NAO indexes that link periods of increased flood frequency with negative winter NAO during the last 3000 years (Benito *et al.,* 2008). A brief link was attempted to link historical flood frequency records of the Tagus River, Spain with solar variability during the last 1000 years (Vaquero, 2004). More recently, a Bayesian approach has been taken for Spanish River catchments using floodplain sediments, SWDs, stratigraphic profiles, palaeoflood discharge estimation, OSL dating and ^{14}C dating for bed-rock gorge sites for linking flood-climate interactions (Thorndycraft *et al.,* 2011). However, analysis of UK fluvial deposits relies solely on statistical analysis of ^{14}C dated organic material. This approach is based on assumptions to make interpretations, which has been extensively discussed and questioned throughout this study. The issues this thesis has identified should be accounted for in any future analysis or indeed prompt the re-analysis of past findings. Importantly, the novel application of Lomb-Scargle spectral analysis demonstrates a new tool to test the ability of palaeoflood data to reflect climatic changes.

7.4 Limitations

Fluvial archives, including sedimentary records and PSIs, are not capable of providing continuous records for many reasons, and so it is the task of researchers to piece together a record that is most likely to match the ‘true’ flood history. This is an ongoing problem for researchers to overcome and despite the investigation of the three research questions in this thesis, it is based on comparing one unknown to another unknown (a sub-dataset with the unfiltered record). The identification of sensitivity and uncertainty with the data selection process and statistical analysis of the British database has highlighted that a singular change can alter the subsequent frequency distributions. This leads to the question: how selective can we be with the data to reach an objective result without introducing bias? Without knowing the ‘true’ flood record this cannot be answered.

In relation to ^{14}C dating, the availability of dates has been identified as a source of bias. One solution to this could be to take samples from deeper depths and so increase the number of older samples, but this would also contribute to preferential sampling. Another question that remains unanswered is: if more dates were added to the database, would this strengthen the influence of the calibration curve and allow more variability to propagate into the shape of frequency distribution?

Frequency analysis has been applied to fluvial dataset before but to the authors knowledge Lomb-Scargle has not been used. This study applied Lomb-Scargle to ‘change after’ dates only to compare with the residual $\Delta^{14}C$ dataset; there is a clear opportunity to carry out Lomb-Scargle frequency analysis on the unfiltered British database as well as other sub-datasets.

7.5 Recommendations

7.5.1 Guidelines for reporting radiocarbon dates in palaeoflood studies

This study recommends the implementation of a standard protocol for the reporting of data in future studies that use radiocarbon dates. The recommendations provided below are specifically targeted at palaeoflood hydrology studies that are dominated by geomorphological data, to address a clear gap in knowledge; but the protocol could be adapted for other palaeoenvironmental or palaeoclimatic studies or across other disciplines. The protocol is characterised by three main sources of information: sample data, quality control, and data analysis. Some of the recommendations are already reported in some studies across different fields but not consistently and this list aims to collate examples for best practise, from studies including Rodríguez-Rey et al. (2015), Millard (2014), Williams (2012), English Heritage studies (Bayliss et al., 2008), and Michczyńska and Pazdur (2004), to create a working standard for the reporting of radiocarbon dated geomorphological data.

In terms of sample data, the following information should be reported and made available:

1) Location of sample:
    a) Latitude and longitude co-ordinates and/ or national grid reference;
    b) Depth of sample from ground level;
    c) Stratigraphic profile of 1 metre above and below the sample including photographs to allow a visual interpretation of context;
    d) If more than one sample is taken from one location, chronologically number the samples with 1 being closest to the surface.
2) Radiocarbon dating:

a) Sample material to determine if the sample was a short or long lived species, which will affect what ‘event’ the radiocarbon age represents (e.g. life or death of an organism);
b) Laboratory name;
c) Laboratory number;
d) Pre-treatment method applied;
e) Radiocarbon measurement technique used;
f) ^{14}C determination and associated uncertainty (in years BP or BC/AD);
g) Range of calibrated dates to 1σ (95.4%) and 2σ (68.2%) confidence intervals;
h) Offsets to test the statistical consistency of radiocarbon measurements when a collection of samples are measured;
i) Radiocarbon calibration software used;
j) Radiocarbon calibration curve used;
k) Stable isotope measurement ($\delta^{13}C$), which can be used to calculate reported ^{14}C ages after the correction procedure to account for fractionation.

3) Supporting information:
   a) Purpose and aim of sample collection for study to be used to support contextual interpretations of groups of radiocarbon dates;
   b) Project manager name and contact in case more information is required to carry out quality control or to find out why if any of the data is missing.

Once all of the above information is available, it will be possible to carry out a quality control procedure. This study recommends using the protocol identified by Rodriguez-Rey et al. (2015) (2.2.1.5) and any data excluded at this stage should also be reported (as an appendix). The more this technique is applied to new and existing collections of radiocarbon dates, more information will become available to help determine if the quality control criteria needs to be refined for the field of palaeoflood hydrology, and it will allow different quality control procedures to be tested on datasets. Currently, there is not enough data available to carry out the quality control criteria so further work is suggested in 7.5.2 to collect additional sample information to allow this.

Given the high number of characteristics of terrestrial geomorphic samples, and the evidence presented in this study of the effect of data selection on the shape of probability curves (4.3), there needs to be further study into the order that quality control criteria are applied. For example, if multiple quality control criteria are applied, such as sample material and sample association, which criteria should be applied first and how will this affect the shape and statistical reliability of the probability curve? These are crucial questions that need to be studied in order to work towards a more objective and practical analysis of palaeoflood data.

The third recommendation is the reporting on the method of data analysis. The method of data analysis should be fully explained and justified and if possible when analysing groups of radiocarbon dates, and it should be reported why another type of data analysis technique was not used. This will improve the knowledge base of analysing collections of radiocarbon dates. Sensitivity testing should be carried out to determine if there are any factors that significantly affect the shape of probability curves following the application of a roust quality control methodology. This extra step would contribute towards improving the precision and accuracy of palaeoflood data.

4) Guidelines for reporting include:

   a) Probabilistic approaches:

i) Present number of dates included in analysis with probability curve, preferably using the definition of reliable histograms given by Geyh (1980);

ii) Report relative uncertainties of radiocarbon ages as this is likely to affect the number of dates required to generate statistically reliable results;

iii) Present the probability curve for the data before quality control has been applied;

iv) Use sensitivity testing as applied in this study (4.3 and 5.3), in Williams (2012) and Michczyńska and Pazdur (2004) to comment on the most influential characteristics of a dataset, and to determine how the number of dates included in analysis statistically affects the reliability of a curve;

v) This study does not recommend the use of relative probability curves because sources of uncertainty are re-introduced at this point that would be excluded during the quality control procedure. In order to account of the influence of the radiocarbon calibration curve on the shape of probability curves, a randomly generated dataset should be generated and compared to identify common peaks and troughs.

b) Bayesian approaches:

i) Follow Thorndycraft et al. (2011) approach.

c) Spectral analysis:

i) Further research is needed to be able to reliably use spectral analysis to interpret palaeoflood data but as a starting point, this study recommends Lomb-Scargle frequency analysis to identify significant cyclicities in unevenly spread palaeoflood datasets using the standard script by Trauth (2010) or the script in Appendix A.

These recommendations aim to create a high standard of best practice on the reporting of radiocarbon dates used across the field of palaeoflood hydrology. This protocol is also intended to allow transparency of data and methodologies used, and to allow re-analysis of existing data when techniques used in radiocarbon dating are developed. This will ensure that data collected will be valuable in future research and can be adapted when quality control criteria's change because all of the relevant information would be available.

7.5.2 Future research

The following recommendations are given in the context of the British database, but they are completely applicable to other databases of ^{14}C dates palaeoflood data and it is the recommendation of this thesis that a formal standard of palaeoflood hydrology is defined and applied across this field.

1. Use the laboratory codes provided for the British database (Macklin *et al*., 2012) to gain the ^{14}C sample pre-treatment methods so that the first stage of quality control can be applied to validify the ^{14}C ages used.
2. Apply a robust quality control protocol, such as the method discussed in Chapter 4 to the British database to identify how many ^{14}C ages pass the initial quality control stage. Then quantify how many ^{14}C ages pass to the next stage to identify where more data is needed. It is hoped that by identifying a standard method of quality control, all future studies using ^{14}C dating will adhere to them to enable data to be used in the future, for example when the calibration curve is next updated.

3. When summed probability distribution functions or relative probability plots are used to represent frequency distributions, the number of *calibrated* dates per time interval should be presented, using the methodology discussed in 3.3.4. This will allow the reader to interpret the probability curve in relation to the availability of dates and to identify clusters of calibrated ^{14}C dates that could be a result of steps or plateaus in the calibration curve.
4. Study the use of Lomb-Scargle frequency analysis further to determine how applicable it is to groups of ^{14}C dates that can be used to interpret palaeoflood records.

7.6 Conclusion

Ultimately, this study recognises the usefulness of sedimentary palaeoflood data and recommends that future research is needed to improve the viability of the records, beginning with bringing the standard of quality control up the standard used in other ^{14}C studies. Once this has been achieved, the viability of the data itself can be used with confidence. This study demonstrated that there are many ways to analyse the British database, and each can lead to a different interpretation. Without the knowledge of the 'true record', more attention needs to be given to all of the factors identified in this study that have shown to influence the shape of probability curves to increase confidence in their ability to represent palaeoflood records.

References

Aalto, R. & Nittrouer, C. A. (2012) ^{210}Pb geochronology of flood events in large tropical river systems. *Philosophical Transactions of the Royal Society*, 370(1966), 2040-2074.

Ashmore, P., Brayshay, B. A., Edwards, K. J., Gilbertson, D. D., Grattan, J. P., Kent, M., Pratt, K. E. & Weaver, R. E. (2000) Allochthonous and autochthonous mire deposits, slope instability and palaeoenvironmental investigations in the Borve Valley, Barra, Outer Hebrides, Scotland. *The Holocene*, 10(1), 97-108.

Archer, D. & Fowler, H. (2018) Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain. *Journal of flood risk management,* 11(1), 121-133.

Archer, D., Parkin, G. & Fowler, H. (2016) Assessing long term flash flooding frequency using historical information. *Hydrology Research,* 48(1), 1-16.

Apel, H., Thieken, A. H., Merz, B. & Blöschl, G. (2006) A probabilistic modelling system for assessing flood risk. *Natural Hazards,* 38, 79-100.

Bachand, B. R. (2008) Bayesian refinement of a stratified sequence of radiometric dates from Punta De Chimino, Guatemala. *Radiocarbon,* 50(1), 19-51.

Baker, V. R. (1973) Erosional forms and processes for the catastrophic Pleistocene Missoula floods in eastern Washington. *Fluvial Geomorphology*, 123-148.

Baker, V. R. & Pickup, G. (2009) Palaeohydrologic analysis of Holocene flood slack- water sediments. *Modern and ancient fluvial systems*, 229-239.

Baker, V. R. (2016) *PAGES Floods Working Group in Grenoble, France* [personal communication]. 27 June 2016, 14:00.

Baker, V. R. (2006) Paleoflood hydrology in a global context. *Catena,* 66(1-2), 161-168.

Baker, V. R. (2008) Palaeoflood hydrology: Origin, progress, prospects. *Geomorphology,* 101, 1-13.

Ballesteros-Cánovas, J. A., Stoffel, M., St George, S., & Hirschboeck, K. (2015). A review of flood records from tree rings. Progress in Physical Geography, 39(6), 794-816.

Bartlein, P. J., Edwards, M. E., Shaffer, S. L. & Barker Jr, E. D. (1995) Calibration of radiocarbon ages and the interpretation of paleoenvironmental records. *Quaternary Research,* 44, 417-424.

Bayliss, A., Cook, G., Bronk Ramsey, C., van der Plicht & McCormac, G. (2008) *Radiocarbon dates from samples funded by English Heritage under the Aggregates Levy Sustainability Fund 2004-7*. (English Heritage, Swindon).

Bayliss, A., McCorman, G. & van der Plicht, H. (2004) An illustrated guide to measuring radiocarbon from archaeological samples. *Physics Education Special feature: Archaeology,* 39(2), 137-144.

Benito, G. & Thorndycraft, V. R. (2005) Palaeoflood hydrology and its role in applied hydrological sciences. *Journal of Hydrology*, 313(1-2), 3–15.

Benito, G., Thorndycraft, V. R., Rico, M., Sánchez-Moya, Y. & Sopeña, A. (2008) Palaeoflood and floodplain records from Spain: Evidence for long-term climate variability and environmental changes. *Geomorphology,* 101(1-2), 68-77.

Benito, G. & Díez-Herrero, A. (2015) "Palaeoflood hydrology: reconstructing rare events and extreme flood discharges". In *Hydro-Meteorological Hazards, Risks and Disasters* (pp. 65-104).

Benito, G., Lang, M., Barriendos, M., Llasat, C. M., Francés, F., Ouarda, T., Thorndycraft, V. R., Enzel, Y., Barbossy, A., Coeur, D. & Bobée, B. (2004) Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. Review of scientific methods. *Natural Hazards,* 31, 623-643.

Benito, G., Díez-Herrero, A., & De Villalta, M. F. (2003). Magnitude and frequency of flooding in the Tagus basin (Central Spain) over the last millennium. *Climatic Change*, 58(1-2), 171-192.

Beta Analytic. (2019a) *How does carbon dating work?* Available online: https://www.radiocarbon.com/about-carbon-dating.htm [accessed 9/1/2019].

Beta Analytic. (2019b) *Radiocarbon dating peat*. Available online: https://www.radiocarbon.com/ams-dating-peat.htm [accessed 9/1/2019].

Beta Analytic. (2019c) *Sample suitability: AMS or Radiometric Dating?* Available online: https://www.radiocarbon.com/radiometric-plus.htm [accessed 9/1/2019].

Blackwell, P. G. & Buck, C. E. (2008) Estimating radiocarbon calibration curves. *Bayesian Analysis*, 3(2), 225-248.

Blaauw, M., & Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian analysis*, 6(3), 457-474.

Blaauw, M., van der Plicht, J., & van Geel, B. (2004). Radiocarbon dating of bulk peat samples from raised bogs: non-existence of a previously reported 'reservoir effect'?. *Quaternary Science Reviews*, *23*(14-15), 1537-1542.

Bradwell, T. (2009) Lichenometric dating: a commentary, in the light of some recent statistical studies. *Geografiska Annaler: Series A, Physical Geography*, 91(2), 61-69.

Brazier, V., Whittington, G. & Ballantyne, C. K. (1988) Holocene debris cone evolution in Glen Etive, Western Grampian Highlands, Scotland. *Earth Surface Processes and Landforms*, 13(6), 525–531.

Briant, R. M. & Bateman, M. D. (2009) Luminescence dating indicates radiocarbon age underestimation in late Pleistocene fluvial deposits from eastern England. *Journal of Quaternary Science: Published for the Quaternary Research Association*, 24(8), 916-927.

Brock, F., Higham, T., Ditchfield, P. & Ramsey, C. B. (2010) Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon*, 52(1), 103-112.

Bronk Ramsey, C. (2009) Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, 51(1), 337–360.

Bronk Ramsey, C. (2008) Radiocarbon dating: Revolutions in understanding. *Archaeometry*, 50(2), 249–275.

Bronk Ramsey, C. (2016) Query regarding sum() [Email]. Message sent to J Moloney (J.L.Moloney@2014.hull.ac.uk). 15 Jun 2016, 17:26.

Buck, C. E., Christen, J. A., Kenworthy, J. B. & Litton, C. D. (1994) Estimating the duration of archaeological activity using ^{14}C determinations. *Oxford Journal of Archaeology,* 13(2), 229-240.

Centre for Ecology & Hydrology (2018) *National River Flow Archive*. Available online: https://nrfa.ceh.ac.uk/ [accessed 19/12/2018].

Chiverrell, R. C., Sear, D. A., Warburton, J., Macdonald, N., Schillereff, D. N., Dearing, J. A., Croudace, I. W., Brown, J. & Bradley, J. (2019) Using lake sediment archives to improve understanding of flood magnitude and frequency: Recent extreme flooding in northwest UK. *Earth Surface Processes and Landforms.*

Chiverrell, R. C., Thorndycraft, V. R. & Hoffman, V. R. (2011a) Cumulative probability functions and their role in evaluating the chronology of geomorphological events during the Holocene. *Journal of Quaternary Science,* 26, 76-85.

Chiverrell, R. C., Thorndycraft, V. R. & Hoffmann, T. O. (2011b) Reply to comment: Cumulative probability functions and their role in evaluating the chronology of geomorphological events during the Holocene. Richard C. Chiverrell, Varyl, R. Thorndycraft & Thomas, O. Hoffmann, Journal of Quaternary Science 26: 76–85. *Journal of Quaternary Science*, 26(2), 241-244.

Chiu, T. C., Fairbanks, R. G., Cao, L., & Mortlock, R. A. (2007) Analysis of the atmospheric 14C record spanning the past 50,000 years derived from high-precision 230Th/234U/238U, 231Pa/235U and 14C dates on fossil corals. *Quaternary Science Reviews*, *26*(1-2), 18-36.

Corella, J.P., Benito, G., Rodriguez-Lloveras, X., Brauer, A. and Valero-Garcés, B.L., (2014) Annually-resolved lake record of extreme hydro-meteorological events since AD 1347 in NE Iberian Peninsula. *Quaternary Science Reviews*, 93, 77-90.

Croft, R. (2006) *Brant Broughton Gauging Station.* Available online: https://www.geograph.org.uk/photo/166904 [accessed 30 March 2019].

Curry, J. P. & Schmidt, O. (2007) The feeding ecology of earthworms–a review. *Pedobiologia*, 50(6), 463-477.

Dotterweich, M. (2008) The history of soil erosion and fluvial deposits in small catchments of central Europe: deciphering the long-term interaction between humans and the environment—a review. *Geomorphology*, 101(1-2), 192-208.

Environment Agency (2014) *Flood and coastal erosion risk management, Long-term investment scenarios (LTIS) 2014.* Available online: https://www.gov.uk/government/publications/flood-and-coastal-risk-management-in-england-long-term-investment [accessed 8/2/2019].

Feathers, J. K. (2003) Single-grain OSL dating of sediments from the Southern High Plains, USA. *Quaternary Science Reviews,* 22, 1035-1042.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014) *Bayesian Data Analysis - Third Edition*. Boca Raton: CRC Press.

Geyh, M. A. (1980) Holocene sea-level history: case study of the statistical evaluation of ^{14}C dates. *Radiocarbon,* 22(3), 695-704.

Greenwood, M. T., Wood, P. J. & Monk, W. A. (2006) The use of fossil caddisfly assemblages in the reconstruction of flow environments from floodplain paleochannels of the River Trent, England. *Journal of Paleolimnology*, 35(4), 747–761.

Gregory, J. M., Jones, P. D. and Wigley, T. M. L. (1991) Precipitation in Britain: an analysis of area-average data up to 1989. *International Journal of Climatology,* 11, 331-345.

Godwin, H. (1962). Half-life of radiocarbon. Nature, 195(4845), 984.

Guilderson, T. P., Reimer, P. J. & Brown, T. A. (2005) The Boon and Bane of Radiocarbon Dating. *Science,* 307(5708), 362-364.

Gurnell, A. M., Piégay, H., Swanson, F. J. & Gregory, S. V. (2002) Large wood and fluvial processes. *Freshwater Biology,* 47, 601-619.

Haesaerts, P., Borziac, I., Chekha, V. P., Chirica, V., Drozdov, N. I., Koulakovska, L., Orlova, L. A., van der Plicht, Johannes & Damblon, F. (2010). Charcoal and wood remains for radiocarbon dating Upper Pleistocene loess sequences in Eastern Europe and Central Siberia. *Palaeogeography, Palaeoclimatology, Palaeoecology,* 291(1-2), 106-127.

Helsel, D. R. & Hirsch, R. M. (2002) *Statistical methods in water resources* (Vol. 323). Reston, VA: US Geological survey.

Hocke, K. & Kämpfer, N. (2009) Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram. *Atmospheric Chemistry and Physics*, 9(12), 4197-4206.

Hoffmann, T., Lang, A. & Dikau, R. (2008) Holocene river activity: analysing ^{14}C-dated fluvial and colluvial sediments from Germany. *Quaternary Science Reviews,* 27, 2031-2040.

Hooke, J. M. (2006) Human impacts on fluvial systems in the Mediterranean region. *Geomorphology,* 79, 311-335.

Joerin, U. E., Nicolussi, K., Fischer, A., Stocker, T. F. & Schlüchter, C. (2008) Holocene optimum events inferred from subglacial sediments at Tschierva Glacier, Eastern Swiss Alps. *Quaternary Science Reviews*, 27(3-4), 337-350.

Johnstone, E., Macklin, M. G. & Lewin, J. (2006) The development and application of a data base of radiocarbon-dated Holocene fluvial deposits in Great Britain. *Catena,* 66, 14-23.

Jones, A. F., Lewin, J. & Macklin, M. G. (2010) Flood series data for the later Holocene: available approaches, potential and limitations from UK alluvial sediments. *The Holocene*, 20(7), 1123-1135.

Jones, A. F., Macklin, M. G. & Benito, G. (2015) Meta-analysis of Holocene fluvial sedimentary archives: a methodological primer. *Catena,* 130, 3-12.

Knighton A. D. (1998) *Fluvial Forms and Processes: A New Perspective*. London: Arnold.

Knox, J. C. (2000) Sensitivity of modern and Holocene floods to climate change. *Quaternary Science Reviews,* 19(1-5), 439-457.

Kochel, R. C. & Baker, V. (1988) Paleoflood analysis using slackwater deposits. *Flood geomorphology,* 357-376.

Kochel, R. C. & Baker, V. R. (1982) Paleoflood Hydrology. *Science*, 215(4531), 353–361.

Kondrashov, D., Feliks, Y. & Ghil, M. (2005) Oscillatory modes of extended Nile River records (AD 622–1922). *Geophysical research letters*, 32(10).

Lewin, J., Macklin, M. G. & Johnstone, E. (2005) Interpreting alluvial archives: Sedimentological factors in the British Holocene fluvial record. *Quaternary Science Reviews*, 24(16-17), 1873–1889.

Lewis, C. J., McDonald, E. V., Sancho, C., Peña,J. L. & Rhodes, E. J. (2009) Climatic implications of correlated Upper Pleistocene glacial and fluvial deposits on the Cinca and Gállego Rivers (NE Spain) based on OSL dating and soil stratigraphy. *Global and Planetary Change,* 67, 141-152.

Li, Z., Yang, D., Gao, B., Jiao, Y., Hong, Y. & Xu, T. (2015) Multiscale hydrologic applications of the latest satellite precipitation products in the Yangtze River Basin using a distributed hydrologic model. *Journal of Hydrometeorology*, *16*(1), 407-426.

Libby, W. F., Anderson, E. C. & Arnold, J. R. (1949) Age Determination by Radiocarbon Content: World-Wide Assay of Natural Radiocarbon. *Science (New York, N.Y.)*, 109(2827), 227–228.

Lowe, J. J. & Walker, M. J. (2000) Radiocarbon Dating the Last Glacial-Interglacial Transition (Ca. 14–9 14 C Ka Bp) in Terrestrial and Marine Records: The Need for New Quality Assurance Protocols 1. *Radiocarbon*, 42(1), 53-68.

Macdonald, N. (2006) An underutilized resource: historical flood chronologies a valuable resource in determining periods of hydro-geomorphic change. *Sediment Dynamics and the Hydromorphology of Fluvial Systems,* 306, 120-127.

Macdonald, N. & Black, A. R. (2009) Reassessment of flood frequency using historical information for the River Ouse at York, UK (1200-2000). *Hydrological Sciences Journal,* 55(7), 1152-1162.

Macklin, M. G., Rumsby, B. T. & Heap, T. (1992) Flood alluviation and entrenchment: Holocene valley-floor development and transformation in the British uplands. *Geological Society of America Bulletin*, 104(6), 631-643.

Macklin, M. G. & Lewin, J. (2003) River sediments, great floods and centennial-scale Holocene climate change. *Journal of Quaternary Science*, 18, 101-105.

Macklin, M. G. & Lewin, J. (1993) Holocene river alluviation in Britain. *Zeitschrift für Geomorphologie Supplement-Band,* 88, 109-122.

Macklin, M. G., Fuller, I. C., Jones, A. F. & Bebbington, M. (2012) New Zealand and UK Holocene flooding demonstrates interhemispheric climate asynchrony. *Geology,* 40(9), 775-778.

Macklin, M. G., Jones, A. F. & Lewin, J. (2010) River response to rapid Holocene environmental change: evidence and explanation in British catchments. *Quaternary Science Reviews,* 29, 1555-1576.

Macklin, M. G., Johnstone, E. & Lewin, J. (2005) Pervasive and long-term forcing of Holocene river instability and flooding in Great Britain by centennial-scale climate change. *The Holocene*, 15(7), 937–943.

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

Met Office (2018) *UKCP18 – Climate impacts narratives*. Available online: https://www.metoffice.gov.uk/research/collaboration/ukcp/applied-products [accessed 19/12/2018].

Michczyńska D. J. & Pazdur A. (2004) Shape analysis of cumulative probability density function of radiocarbon dates set in the study of climate change in Late Glacial and Holocene. *Radiocarbon,* 46(2), 733–44.

Michczyński, A. & Michczyńska, D. J. (2006) The effect of PDF peaks' height increase during calibration of radiocarbon date sets. *Geochronometria,* 25, 1-4.

Michczyńska, D. J, Michczyński, A. & Padzur, A. (2007) Frequency distribution of radiocarbon dates as a tool for reconstructing environmental changes. *Radiocarbon,* 49(2), 799-806.

Milan, D. J. (2012) Geomorphic impact and system recovery following an extreme flood in an upland stream: Thinhope Burn, northern England, UK. *Geomorphology*, *138*(1), 319-328.

Millard, A. R. (2014) Conventions for reporting radiocarbon determinations. *Radiocarbon*. 56 (2), 555-559.

Mudelsee. M. (2010) *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*. Second Edition. Springer, Cham.

Murray, A. S., Roberts, R. G. & Wintle, A. G. (1997) Equivalent dose measurement using a single aliquot of quartz. *Radiation Measurements*, *27*(2), 171-184.

Murray, A. S. & Wintle, A.G. (2003) The single aliquot regenerative dose protocol: potential for improvements in reliability. *Radiation Measurements,* 37, 377-381.

National River Flow Archive (2019) *About NRFA*. Available online: https://nrfa.ceh.ac.uk/about-nrfa [accessed 6/3/19].

Nesje, A., Dahl, S.O., Matthews, J.A. and Berrisford, M.S., (2001) A ~ 4500 yr record of river floods obtained from a sediment core in Lake Atnsjøen, eastern Norway. *Journal of Paleolimnology*, 25(3), 329-342.

O'Connell, D. R. H. (2005) Non-parametric Bayesian flood frequency estimation. *Journal of Hydrology,* 313, 79-96.

Press, W. H. & Rybicki, G. B. (1989) Fast algorithm for spectral analysis of unevenly sampled data. *The Astrophysical Journal,* 338, 277-280.

Patton, P. C., Baker, V. R. & Kochel, R. C. (1982) Slack-water deposits: a geomorphic technique for the interpretation of fluvial paleohydrology. In Rhodes, D. D. & Williams, G. P (eds) *Adjustments of the fluvial system. Proc. 10th geomorphology symposium, Binghamton, 1979*. Allen & Unwin, 225-254.

Past Global Changes (2018) *Floods Working Group*. Available online: http://www.pastglobalchanges.org/index.php/ini/wg/floods/intro [accessed 2/12/2019].

Panin, A. & Matlakhova, E. (2014) Fluvial chronology in the East European Plain over the last 20 ka and its palaeohydrological implications. *Catena,* 130, 46-61.

Perez, L., Crisci, C., Hanebuth, T. J. J., Lantzsch, H., Perera, G., Rodríguez, M., Pérez, A., Fornaro, L. & García-Rodríguez, F. (2018) Climatic oscillations modulating the late Holocene fluvial discharge and terrigenous material supply from the Río De La Plata into the southwestern Atlantic Ocean. *Journal of Sedimentary Environments,* 3(4), 205-219.

Pettitt, P. B., Davies, W., Gamble, C. S. & Richards, M. B. (2003) Palaeolithic radiocarbon chronology: quantifying our confidence beyond two half-lives. *Journal of Archaeological Science*, 30(12), 1685-1693.

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M. Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M., van der Plicht, J. (2013) IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon*, *55*(4), 1869-1887.

Rhodes, E. J., Bronk Ramsey, C., Outram, Z., Batt, C., Willis, L., Dockrill, S. & Bond, J. (2003) Bayesian methods applied to the interpretation of multiple OSL dates: high precision sediment ages from Old Scatness Broch excavations, Shetland Isles. *Quaternary Science Reviews,* 22, 1231-1244.

Richards, K. S. (1981) Evidence of flandrian valley alluviation in Staindale, North York Moors. *Earth Surface Processes and Landforms*, 6(2), 183–186.

Rittenour, T. M. (2008) Luminescence dating of fluvial deposits: applications to geomorphic, palaeoseismic and archaeological research. *Boreas,* 37, 613-635.

Rodríguez-Rey, M., Herrando-Pérez, S., Gillespie, R., Jacobs, Z., Saltré, F., Brook, B. W. Prideaux, G. J., Roberts, R. G., Cooper, A., Alroy, J., Miller, G. H., Bird, M. I., Johnson, C. N., Beeton, N., Turney, C. S. M. & Bradshaw, C. J. A. (2015) Criteria for assessing the quality of Middle Pleistocene to Holocene vertebrate fossil ages. *Quaternary Geochronology,* 30, 69-79.

Ruf, T. (1999) The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. *Biological Rhythm Research*, 30(2), 178-201.

Saint-Laurent, D. (2004) Palaeoflood hydrology: an emerging science. *Progress in Physical Geography*, *28*(4), 531-543.

Schillereff, D. N., Chiverrell, R. C., Macdonald, N., Hooke, J. M., Welsh, K. E., Piliposian, G. & Croudance, I.W. (2019) Convergent human and climate forcing of late-Holocene flooding in Northwest England. *Global and Planetary Change,* 182, 102998.

Schillereff, D. N., Chiverrell, R. C., Macdonald, N. & Hooke, J. M. (2014) Flood stratigraphies in lake sediments: A review. *Earth-Science Reviews,* 135, 17-37.

Schimmel, M. (2001) Emphasizing difficulties in the detection of rhythms with Lomb-Scargle Periodograms. *Biological Rhythm Research,* 32(3), 341-345.

Schulz, M. & Stattegger, K. (1997). Spectrum: Spectral analysis of unevenly spaced paleoclimatic time series. *Computers & Geosciences,* 23(9), 929 – 945.

Scott, E. M., Cook, G. T. & Naysmith, P. (2007) Error and uncertainty in radiocarbon measurements. *Radiocarbon*, *49*(2), 427-440.

Siegenthaler, U., Heimann, M. & Oeschger, H. (1980) ^{14}C variations caused by changes in the global carbon cycle. *Radiocarbon,* 22(2), 177-191.

Stuiver, M. and Polach, H.A. (1977) Discussion reporting of ^{14}C data. Radiocarbon, 19(3), 355-363.

Starkel, L., Soja, R. & Michczyńska, D. J. (2006) Past hydrological events reflected in Holocene history of Polish rivers. *Catena*, *66*(1-2), 24-33.

Stolk, A., Törnqvist, T. E., Hekhuis, K. P. V., Berendsen, H. J. A., & Plicht, J. V. D. (1994) Calibration of ^{14}C Histograms: A Comparison of Methods. *Radiocarbon*, 36(1), 1-10.

Stuiver, M., & Suess, H. E. (1966). On the relationship between radiocarbon dates and true sample ages. Radiocarbon, 8, 534-540.

Stuiver, M., Reimer, P. J., Bard, E., Beck, J. W., Burr, G. S., Hughen, K. A., Kromer, B., McCormac, G., Van Der Plicht, J. & Spurk, M. (1998) INTCAL98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon*, 40(3), 1041-1083.

Swindles, G. T., Patterson, R. T., Roe, H. M. & Galloway, J. M. (2012) Evaluating periodicities in peat-based climate proxy records. *Quaternary Science Reviews,* 41, 94-103.

Taylor, R. (1990) Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1), 35-39.

Taylor, M. P. & Lewin, J. (1997) Non-synchronous response of adjacent floodplain systems to Holocene environmental change. *Geomorphology*, 18(3-4), 251-264.

Telford, R. J., Heegaard, E. & Birks, H. J. B. (2004) The intercept is a poor estimate of a calibrated radiocarbon age. *The Holocene,* 14(2), 296-298.

Thorndycraft, V. R. & Benito, G. (2006) The Holocene fluvial chronology of Spain: evidence from a newly compiled radiocarbon database. *Quaternary Science Reviews*, 25(3-4), 223-234.

Thorndycraft, V. R. & Benito, G. (2006) Late Holocene fluvial chronology of Spain: The role of climatic variability and human impact. *Catena,* 1-2, 34-41.

Thorndycraft, V. R., Benito, G., Rico, M., Sopeña, A., Sánchez-Moya, Y. & Casas, A. (2005) A long-term flood discharge record derived from slackwater flood deposits of the Llobregat River, NE Spain. *Journal of Hydrology,* 313(1-2), 16-31.

Thorndycraft, V. R., Benito, G., Sánchez-Moya, Y. & Sopeña, A. (2011) Bayesian age modelling applied to palaeoflood geochronologies and the investigation of Holocene flood magnitude and frequency. *The Holocene,* 1-10.

Tipping, R. (1998) The chronology of late Quaternary fluvial activity in part of the Milfield Basin, northeast England. *Earth Surface Processes and Landforms*, 23(9), 845–856.

Trauth, M. H. (2010) Time-Series Analysis. *MATLAB® Recipes for Earth Sciences,* 2nd Edition. New York: Springer, 107-159

Vaquero, J. M. (2004). Solar signal in the number of floods recorded for the Tagus river basin over the last millennium. *Climatic Change*, 66, 23-26.

Van De Wiel, M. J. & Coulthard, T. J. (2010) Self-organized criticality in river basins: Challenging sedimentary records of environmental change. *Geology.* 38(1), 87-90.

Van der Plicht, J. & Hogg, A. (2006) A note on reporting radiocarbon. *Quaternary Geochronology,* 1, 237 – 240.

VanderPlas, J. T. (2018). Understanding the Lomb–Scargle Periodogram. *The Astrophysical Journal Supplement Series*, 236(1), 16.

Vig, J. R. (1992). Introduction to quartz frequency standards. Oscilent Corporation – Technical References. Army Research Laboratory, Electronics and Power Sources Directorate.

Wallinga, J. (2002) Optically stimulated luminescence dating of fluvial deposits: a review. *Boreas,* 31, 303-322.

Wallinga, J., Murray, A. & Wintle, A. (2000) The single-aliquot regenerative-dose (SAR) protocol applied to coarse-grain feldspar. *Radiation Measurements,* 32, 529-533.

Wang, Y., Amundson, R. & Trumbore, S. (1995) Radiocarbon Dating of Soil Organic Matter. *Quaternary Research,* 45, 282-288.

Weijian, Z., Xuefeng, L., Zhengkun, W., Lin, D., Jull, A. J. T., Donahue, D. & Beck, W. (2002) Peat record reflecting Holocene climatic change in the Zoigê Plateau and AMS radiocarbon dating. *Chinese Science Bulletin,* 47(1), 66-70.

Wheater, H. S. (2006) Flood hazard and management: a UK perspective. *Philosophical transactions of the Royal Society,* 364, 2135-2145.

Wilby, R. L., Beven, K. J. & Reynard, N.S. (2008) Climate change and fluvial flood risk in the UK: more of the same? *Hydrological processes,* 22, 2511-2523.

Wilhelm, B., Arnaud, F., Sabatier, P., Magand, O., Chapron, E., Courp, T., Tachikawa, K., Fanget, B., Malet, E., Pignol, C. and Bard, E., (2013) Palaeoflood activity and climate change over the last 1400 years recorded by lake sediments in the north-west European Alps. Journal of Quaternary Science, 28(2), 189-199.

Wilhelm, B., Ballesteros Cánovas, J. A., Macdonald, N., Toonen, W. H., Baker, V., Barriendos, M., Benito, G., Brauer, A., Corella, J. P., Denniston, R., Glaser, R., Ionita, M., Kahle, M., Liu, T., Luetscher, M., Macklin, M., Mudelsee, M., Munoz, S., Schulte, L., St. George, S., Stoffel, M. &Wetter, O. (2019) Interpreting historical, botanical, and geological evidence to aid preparations for future floods. *Wiley Interdisciplinary Reviews: Water*, 6(1), Available online: http://wires.wiley.com/WileyCDA/WiresArticle/wisId-WAT21318.html [accessed 11/2/2018].

Wilk, J., Kniveton, D., Andersson, L., Layberry, R., Todd, M. C., Hughes, D., Ringrose, S. & Vanderpost, C. (2006). Estimating rainfall and water balance over the Okavango River Basin for hydrological applications. *Journal of Hydrology*, *331*(1-2), 18-29.

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

Wirth, S. B., Glur, L., Gilli, A., & Anselmetti, F. S. (2013). Holocene flood frequency across the Central Alps–solar forcing and evidence for variations in North Atlantic atmospheric circulation. *Quaternary Science Reviews*, 80, 112-128.

Wohl, E. E. (1992) Bedrock benches and boulder bars: floods in the Burdekin Gorge of Australia. *Geological Society of America Bulletin*, 104(6), 770–778.

Wood, R. (2015) From revolution to convention: the past, present and future of radiocarbon dating. *Journal of Archaeological Science*, *56*, 61-72.

Yang, D., Koike, T. & Tanizawa, H. (2004) Application of a distributed hydrological model and weather radar observations for flood management in the upper Tone River of Japan. *Hydrological Processes*, 18(16), 3119-3132.

Zare, R. N. (2010) Ultrasensitive radiocarbon detection. *Nature*, 482, 312-313

Zielonka, T., Holeksa, J., & Ciapała, S. (2008) A reconstruction of flood events using scarred trees in the Tatra Mountains, Poland. *Dendrochronologia*, 26(3), 173-183.

Appendix

Appendix A Script used to apply Lomb-Scargle frequency analysis to British database and residual Δ¹⁴C data (after Trauth, 2010)

```
clear

% Load data

series3 = load('southeasteng.txt');
t = series3(:,1);
x = series3(:,2);

% bracket frequencies

int = mean(diff(t));
ofac = 4; hifac = 1;
f = ((2*int)^(-1))/(length(x)*ofac) : ((2*int)^(-1))/(length(x)*ofac) : hifac*(2*int)^(-1);

% Normalise the data

x = (x - mean(x));

% Do analysis

for k = 1:length(f)
   wrun = 2*pi*f(k);
   px(k) = 1/(2*var(x))*((sum(x.*cos(wrun*t-atan2(sum(sin(2*wrun*t)),sum(cos(2*wrun*t)))/2))).^2)/(sum((cos(wrun*t-atan2(sum(sin(2*wrun*t)),sum(cos(2*wrun*t)))/2)).^2))+((sum(x.*sin(wrun*t-atan2(sum(sin(2*wrun*t)),sum(cos(2*wrun*t)))/2))).^2)/(sum((sin(wrun*t-atan2(sum(sin(2*wrun*t)),sum(cos(2*wrun*t)))/2)).^2));
```

```
end

% calculate probability

prob = 1-(1-exp(-px)).^length(x);

wav = power(f,-1);

%plot

plot(wav,px)
xlabel('Wavelength')
ylabel('Power')
title('')

figure
plot(wav,prob)
xlabel('Wavelength')
ylabel('Probability')
title('')
```