1. Introduction

On 5–6 December 2015 Storm Desmond swept across northern Britain, leaving record rainfall and widespread flooding in its wake (Priestly, 2016). One of the worst affected places was Carlisle, where thousands of homes were flooded after newly completed flood defences were overtopped by rising flood waters (BBC, 2015a). While ministers insisted the defences could not be expected to cope with the “completely unprecedented and unpredicted levels of rainfall” (ITV News, 2015) from Storm Desmond, local residents wondered why a £38 million defence scheme had failed barely five years after it was built (BBC, 2010).

The Carlisle case illustrates many of the central challenges of quantifying the risk of flooding. The design of individual schemes, like the wider system for allocating resources for flood defence in England, requires estimates of the probability of flooding to support risk-based management (Lane et al., 2011). Quantitative assessments of flood frequency are also central to planning regulation (Porter and Demeritt, 2012), insurability standards and pricing (Krieger and Demeritt, 2015), and to the flood risk maps and management plans mandated by the EU Floods Directive (2007/60/EC). These and other flood risk management policy instruments all depend on so-called ‘design floods’, whose magnitude is defined in terms of nominal return-periods, like the 1 in 100 year flood. However estimating flood frequency is necessarily uncertain, and uncertainty in the design floods used for flood risk assessment can have major implications for multi-million pound decisions about whether and where new developments are permitted and what, if any, standard of protection will be provided to defend them from flooding.

One of the major sources of uncertainty in flood frequency analysis is the paucity of instrumental records. Although the 0.01 Annual Exceedance Probability (AEP) flood event has become something of an international default for design floods, mandated both by the EU Floods Directive and by the US National Flood Insurance Program, only a tiny portion of gauges provide continuous data for that long. Hydrologists have developed a number of approaches for dealing with the uncertainties arising from this deficit of instrumental records, but they still leave wide confidence
intervals, particularly for extreme flood events, whose frequency we are most concerned with estimating correctly.

To address that challenge, this paper develops a Bayesian model for supplementing instrumental readings of flood discharge from a river flow gauge with estimates derived from documentary records of historical floods, using the River Eden at Carlisle, UK as a case study. For this case study, we use the term ‘historic data’ to refer to estimates of flood discharge from any time prior to the introduction of the river flow gauge in 1967; whereas discharge readings from the systematic period since 1967 are referred to as ‘gauge data’. Compared to the conventional frequentist approaches to estimating flood return periods, our Bayesian approach allows us to update our initial estimates of flood frequencies by incorporating other increasingly uncertain kinds of data and to quantify the total uncertainty involved in combining them through Monte Carlo methods of sensitivity analysis. Our data, analytical methods, and statistical models are described in Section 3. Then in Section 4 we contrast the results of our Bayesian model to the flood frequency estimation performed using the WINFAP-FEH software (WHS, 2014), which the Environment Agency (2012) recommends for use in official flood appraisals in the England and Wales. In Section 4 we also estimate the AEP for the recent extreme flood in Carlisle and use this example to evaluate the effect that an additional data point can make to the uncertainty in the flood frequency curve for sites of interest. The paper concludes with a discussion of the wider implications of its findings for flood risk management.

2. Uncertainty in flood frequency analysis

Flood frequency analysis involves both epistemic uncertainties, which arise from imperfect knowledge of the system being modelled, and aleatory or stochastic uncertainties, which, for the purposes of flood frequency analysis, can be thought of as truly random (Beven, 2008). Environmental modelling has advanced by minimising epistemic uncertainty through improved representation of natural processes whilst characterising aleatory uncertainties probabilistically (Merz and Thieken, 2009). In the context of flood frequency analysis, the aleatory uncertainties associated with the chaotic aspects of long term weather forecasting limit the goals of the analyst to that of establishing, through a flood frequency curve, the probabilities of floods of certain magnitudes rather than making predictions of when flood will actually occur. Consequently, an idealised flood frequency curve for a location would be an accurate representation of the flood probabilities in that location. In practice it is inconceivable that all epistemic uncertainty could be removed, even then, there would remain some sources of aleatory uncertainty (such as the non-stationarity of the system due to climate change) which would introduce error.

In this section we briefly review some of the sources of epistemic uncertainty that undermine the conventional statistical approaches for deriving probability distribution functions used to represent stochastic processes. We then introduce the concept of Bayesian approaches to analysis that allow the incorporation of heterogeneous sources of data into the statistical analysis with the overall aim of reducing the uncertainty in the probability distribution of flood frequencies.

2.1. Statistical approaches to flood frequency analysis

Fisher and Tippett (1928) developed the first frequency curves for estimating the probability of extreme events from time series data, and the field of extreme value statistics is now well developed (e.g. Castillo et al., 2005; Coles et al., 2001; De Haan and Ferreira, 2006; Reiss and Thomas, 2007). For hydrological purposes, time series are typically divided into periods of 1 year\(^1\) with the maximum discharge figure recorded in each year termed the annual maximum (AMAX). Strategies for selecting the most appropriate extreme value distribution (EVD) and fitting the distribution parameters to the data can be reviewed in texts such as Hosking and Wallis (1997), Beven and Hall (2014), and Parkes (2015).

In the UK the Flood Studies Report (FSR) (NERC, 1975) first prescribed standardised ways to estimate flood frequencies from the limited river and rainfall gauge data that was available. Subsequently, the Flood Estimation Handbook (FEH), the successor to the FSR (IH, 1999), described several ways of estimating flood frequency curves which, together with the related WINFAP-FEH software from WHS (2014), are heavily recommended by the Environment Agency (2012). The WINFAP-FEH software makes it relatively simple to derive a flood frequency curve for a river location, especially if there is an operational river flow gauge nearby (see Fig. 6 for an example).

Often, the dominant source of uncertainty in flood frequency analysis is the sampling error due to the shortness of the available time series (Apel et al., 2008; Kjeldsen et al., 2014a). The FEH advises that return period estimates from flood frequency curves derived from a single gauge should not be used to estimate flood return periods greater than half the length of the gauged record (WHS, 2009b). This ‘rule of thumb’ advice has serious consequences for practitioners: most gauge records in Britain are less than 40 years old (Kjeldsen et al., 2008) so the use of conventional statistical methods for flood frequency analysis is severely restricted. A widely used method for overcoming this limitation is to ‘trade space for time’ (Van Gelder et al., 2000) by combining the limited systematic gauge data at one site with data from ‘hydrologically similar’ sites elsewhere. This method is known as ‘regional frequency analysis’, and it is comprehensively assessed by Hosking and Wallis (1997). The WINFAP-FEH software supports regional frequency analysis with its ‘pooled analysis’ feature but, at present, the WINFAP-FEH software provides no method for estimating confidence limits for pooled analyses. Furthermore, concerns with the earlier methods of site pooling have led to alterations in the pooling method, referred to as ‘enhanced single site analysis’, such that greater weighting is now given to the gauged records at the site (WHS, 2009b). This suggests that issues of inter-site correlation and heterogeneity endemic to site pooling methods are not yet fully understood and will undermine any attempt at uncertainty estimation. Consequently, pooled analysis is not considered further in this paper. For further information see Hosking and Wallis (1997) or Kjeldsen and Jones (2006).

Uncertainties also arise from the different methods of measuring discharge, whose accuracy varies considerably with the type of gauge, the geomorphology of the channel, and the level of water. Furthermore discharge measurement uncertainties increase during floods (Di Baldassarre and Montanari, 2010; IH, 1999; Neppel et al., 2010; Rosso, 1985). For the majority of stations in the UK, the water level is recorded every 15 min and the discharge is calculated indirectly from the stage-discharge relation using a rating curve (CEH, 2014). Estimates of the general errors in discharge estimates range from 3% to 5% of the discharge estimate (Cong and Xu, 1987) up to as much as 30% for extreme flows (Kuczera, 1996; Potter and Walker, 1981). Other sources give typical values in the range of 4–8% with estimates tending to cluster around 6% (see for example Leonard et al., 2000; Pappenberger et al., 2006 and sources therein). Measurement errors will contribute to the overall uncertainty of the flood frequency curve, but flood frequency curves derived from river flow gauges often do not take

---

\(^1\) The ‘year’ for hydrological purposes in the UK is defined as the ‘water year’ which begins on 1st October, deemed to be the time when groundwater storages are most usually low.
account for this source of uncertainty (for example MacDonald et al., 2013; Miller et al., 2012). The model described in this paper explicitly considers the uncertainty in the gauged data (see Section 3.6).

2.2. Information from historic flood records

Where information about historic floods is available, researchers have sometimes tried to incorporate it into their analyses. While the original flood studies report (NERC, 1975) mentioned the practice was unusual, with just a few isolated examples (Potter (1978), Sutcliffe (1978), McEwen (1987, 1990) and Acreman (1989)). But subsequently, flexible data fitting methods such as the Expected Moments Algorithm (EMA) (Cohn et al., 1997, 2001) that can accommodate forms of historical data sets are becoming standard practice in the US. In the UK, a report published by the Centre for Ecology and Hydrology in 2001 provided guidance for, and underlined the benefits of, incorporating historical data in flood frequency estimates (Bayliss and Reed, 2001). Soon thereafter the British Hydrological Society produced a database of historical hydrological records (Black and Law, 2004), which have been used in several studies (see, for example Black and Burns, 2002; MacDonald, 2013; MacDonald and Black, 2010; Macdonald et al., 2006; Williams and Archer, 2002). But concerns remain about historical data introducing biases that can degrade the accuracy of flood frequency analysis (Hosking and Wallis, 1986).

A recent review by Kjeldsen et al. (2014b) highlights the unrealised potential to reduce the uncertainty in flood risk estimates through careful use of historic data. Europe provides a wealth of archival resources for studying historical floods; Brázdil et al. (2006) cite numerous European studies spanning no fewer than 10 countries. In China too there is a long history of recording high water marks, which have been exploited by historical flood studies (see for example Li et al., 2013). Other historic information sources include articles or photographs from newspaper archives (MacDonald, 2014), high water marks carved onto bridges or other buildings (Macdonald, 2007), dated stones placed at high water marks of out of bank flows (Condie and Lee, 1982), diary or journal entries and even literary works (The Bristol Post, 2013).

An epigraphic mark or a description in a journal can give an idea of the maximum flood depth at a certain point, but this must then be converted into peak discharge estimate, typically by using a rating curve (see Section 2.1), though geomorphological changes to the channel may mean the current rating curve equation is less applicable to historic floods (see for example Archer et al., 2007). Issues of non-stationarity such as this are discussed further in Section 5.

Such historic flood discharge estimates are likely to contain considerable errors that must be accounted for. First, regardless of the source, all historic data must be subjectively evaluated to assess its reliability and the value it might add to existing flood frequency estimates. For example a recording of a single flood event from a journal of 1000 years ago may provide valuable insights about the potential impact of a similar flood on contemporary society (Brázdil et al., 2006) and assist in the estimation of the Probable Maximum Flood (see for example Acreman, 1989; Kirby and Moss, 1987) for a location. But that same source would be of little value in improving estimation of modern day flood frequencies if it doesn’t cover a reasonably continuous time period. If, however, the historical flood record provides some quantifiable information about the likely flood magnitude and covers a continuous multidecade period such that the researcher can be reasonably confident that all significant floods in that period at that location are recorded, then the data can often be combined with the recent, instrumental time series to improve the estimation of flood frequencies.

As well as using historical data to provide estimates of peak flood discharge, historical data may also provide information on historical changes to the catchment, channel or hydrology that undermine assumptions of stationarity and affect flood magnitude and inundation extent (see for example Naulet et al. (2005), Razavi et al. (2015)). This aspect is discussed for the Carlisle case study in Section 3.3.

One important step in working with historic flood records is identifying the smallest flood event still significant enough to have been captured in the records. The estimated discharge associated with that event becomes the ‘perception threshold’ ($Q_{0}$) for the historical time series which can then be treated as a ‘censored sample’ for statistical purposes (Reis and Stedinger, 2005). Furthermore, even if it is not possible to estimate the precise magnitude of floods above the threshold of perception, the historical time series can still be incorporated as a ‘binomial-censored’ sample whereby it is known that in the historical record of $n$ years, $Q_{0}$ was exceeded $k$ times (Stedinger and Cohn, 1986). Where it is possible to estimate the magnitude of historical floods, it is likely that the uncertainty in those peak discharge estimates would be greater than the instrumental record, and this must also be considered when estimating the uncertainty of the flood frequency curve (Viglione et al., 2013). Examples of studies where MCMC methods have been used in Bayesian models that incorporate ‘censored samples’ of historic floods include: O’Connell (2005), Parent and Bernier (2003), Reis and Stedinger (2005); and Viglione et al. (2013).

2.3. Bayes theorem

The Reverend Thomas Bayes’ (1702–1761) famous theorem relating conditional probabilities was published posthumously by Richard Price, a mathematician friend of Bayes in an essay that appeared in the Philosophical Transactions of the Royal Society (Bayes and Price, 1763; Bertsch-McGrayne, 2011). Bayes’ rule can be considered very simplistically as a statement that “by updating our initial belief with objective new information, we get a new and improved belief” (Bertsch-McGrayne, 2011). In mathematical form it is written:

$$p(\theta|D) = \frac{\ell(D|\theta)p(\theta)}{\int \ell(D|\theta)p(\theta)d\theta}$$

(1)

For any probability distribution with parameters $\theta$, such that $p(\theta)$ is the prior distribution (or initial belief) and $p(D|\theta)$ is the posterior distribution (new and improved belief) after having observed the data, $D$ (new information). $\ell(D|\theta)$ is the probability distribution or likelihood of the data ($D$) conditional on the parameters ($\theta$). The integral denominator in Eq. (1) is a normalisation constant ensuring the area under $p(\theta|D)$ is unity. Consequently Bayes’ theorem can be expressed more simply as:

$$p(\theta|D) \propto \ell(D|\theta)p(\theta)$$

(2)

Notwithstanding the simplicity of Eqs. (1) and (2), the integral denominator in Eq. (1) often has no analytical solution (Viglione et al., 2013). Furthermore as more than one source of uncertain observed data is used to infer multi-parameter distributions, the expression of the joint posterior distribution also tends to become intractable (Lunn et al., 2013). The steps required to address these problems are far from straightforward, which is perhaps why Bayesian approaches to flood frequency analysis are uncommon outside academic research (Kjeldsen et al., 2014b).

But advances in computing power now make it possible to run Markov Chain Monte Carlo (MCMC) simulations that produce a sample of the desired posterior distribution after a large number
of steps. This brings Bayesian solutions within the reach of researchers without extensive specialist statistical expertise (Reis and Stedinger, 2005; Robert and Casella, 2004), and there are many examples where researchers have proposed ways to combine several data sources for flood frequency analysis and manage the uncertainty within those data sources (see for example Chandler, 2013; Fernandes et al., 2010; Fill and Stedinger, 1998; Gaume et al., 2010; Haddad and Rahman, 2012; Haddad et al., 2012; Kuczera, 1999; Neppel et al., 2010; O’Connell et al., 2002; Parent and Bernier, 2003; Payrastre et al., 2011, 2013; Reis and Stedinger, 2005; Reis et al., 2005; Renard et al., 2006; Ribatet et al., 2007; Seidou et al., 2006; Viglione et al., 2013; Wood and Rodríguez-Iturbe, 1975).

3. Data and methods

Carlisle is a city in Cumbria, North West England with a population of approximately 100,000 and a long history as an important regional centre pre-dating the Roman occupation. Carlisle is located on the confluence of the rivers Eden, Caldew, and Petteril, all of which have flooded in recent decades. The fact that the city has been established for many centuries means there are substantial local data archives charting the long history of flooding in the area. A river flow gauge was installed on the River Eden at Sheepmount (see Fig. 1 for location) in 1967. The gauge has been in operation for over 40 years recording both flow and water level data every 15 min (CEH, 2014). At the time of writing the most recent published gauge data is from 2012, giving a 46 year AMAX time series.

On 8 January 2005 Carlisle experienced its worst flood for over 200 years (Environment Agency, 2006), inundating much of the city (see Fig. 1). This event has been heavily analysed thanks to a unique set of accurate observations of peak flood depth and extent collected immediately after the flood (see for example Fewtrell et al., 2011; Horritt et al., 2010; Neal et al., 2012, 2009; Parkes et al., 2013). As a result of the January 2005 event, an improved flood defence scheme was developed and completed in 2010 (BBC, 2010), protecting the city against 0.005 AEP floods (return period of 200 years) (Environment Agency, 2009).

3.1. River gauge data at Carlisle

The Flood Studies report (CEH, 2014) lists the gauge on the River Eden at Sheepmount as ‘grade B’, while a Flow Quality Comment document describes the gaugings there as showing less than 10% deviation from the rating curve. It characterises the high flow record as “good to valid” although substantial bypass flow is also noted when the level is above bank (Environment Agency, 2011c). In a paper modelling the January, 2005 Carlisle flood, Horritt et al. (2010) set up a simple finite volume model (Horritt, 2004) in order to assess the performance of the Sheepmount gauge and the Cumberdale gauge on the River Caldew. They find the EA rating curve to be a good fit to both the within bank stage-discharge measurements (up to 1000 m³ s⁻¹) and for the January, 2005 event, although they recommend revising the peak flow from 1520 down to 1490 m³ s⁻¹. This evidence suggests that gauge readings at Sheepmount can be relied upon to be within approximately ±10% even for out-of-bank events. The Bayesian model described in Section 3.5 takes account of this source of uncertainty and the sensitivity of the model to it can be seen in Fig. 5.

3.2. Historical data at Carlisle

The January 2005 flood was described at the time as the worst to affect Carlisle in over 200 years (Environment Agency, 2006), only to be exceeded by the December 2015 flood (Marsh, 2016). But these are by no means the only destructive floods to have affected Carlisle; flood marks have been recorded on the Eden Bridge since its construction in 1815, and these can be cross-referenced with newspaper reports and a staff gauge maintained by Carlisle City Council for nearly 100 years from 1850 (Environment Agency, 2006). Smith and Tobin (1979) collated records of historic flood events in Carlisle between 1800 and 1970, which the Environment Agency (2006) updated as part of the Cumbria Floods Technical Report produced in response to the 2005 floods. Table 1 summarises data for the 212 years from 1800 to 2012.

Using the current rating equation (CEH, 2014):

\[
Q = 56.612 \times (h - 2.980)^{1.699}
\]

where \( h \) is the measured (or estimated) stage, which for this site is AOD – 6.9 m.

Notes:

1. Level at Eden Bride derived from recurrence interval in Smith and Tobin (1979). For historical records, the water level at the Sheeemount gauging station is calculated as approximately 1.1 m lower than the estimated level at the Eden Bridge.

2. Level at Eden Bridge from Cumbria Floods Technical Report (Environment Agency, 2006). For historical records, the water level at the Sheepmount gauging station is estimated to be approximately 1.1 m lower than the estimated level at the Eden Bridge. This figure comes from hydraulic flood models of the January 2005 flood (see Parkes, 2015; Parkes et al., 2013). 3. From Sheepmount river gauge records (CEH, 2014).

4. These records are not used in the historical data set because they are not the maximum flood for the ‘water year’ which starts on 1st October.

The historical (pre-gauge) flood data available for Carlisle is in the form of a “censored” time series where discharge estimates for all floods above a certain “perception threshold” are assumed to be known (Stedinger and Cohn, 1986). Combining the gauged data from the Sheepmount gauge with the historical estimates (Table 1) produces the time series shown in Fig. 2. Fig. 2(a) displays the gauged and historical flood peaks as known values, whereas in reality both the gauged AMAX readings and the historical estimates are uncertain (although the gauge readings are likely to be more accurate than the historical estimates). The ‘perception threshold’ is inferred from the lowest historical flood peak. In this case the perception threshold is likely to remain fairly constant over the historical period because, in some sense, it is anchored by the presence of the Eden Bridge, which was constructed in 1815, and has provided a benchmark against which the height of subsequent floods have been marked by local residents (see Section 3.3). The smallest flood in the historical record is estimated to be 893 m³ s⁻¹, suggesting a perception threshold of approximately 890 m³ s⁻¹. Since the perception threshold is inferred from the highly uncertain historical flood record, its value is also uncertain. Fig. 3(b) shows the uncertainty associated with the gauge readings, historical flood estimates, and the perception threshold. Methods for deriving flood frequency curves should ideally incorporate all these sources of uncertainty when generating confidence limits for the EVDF parameters. Fortunately a Bayesian analysis integrating the measurement errors and uncertainties can provide a full posterior distribution of the EVDF parameters (Reis and Stedinger, 2005).

3.3. Land use and channel change

When using historic flood data to estimate contemporary flood frequency curves, it is often necessary to consider physical changes
to the catchment, river channel or floodplain that may have affected the extent of historic flooding and thus the rating curve for the river at the study site.

The relationship between land-use change in the catchment and fluvial flood risk is not yet fully understood, but is thought to vary with the scale of the catchment under consideration (Pattison and Lane, 2012). Historical maps of the catchment show minimal land use change over the historical period; 200 years ago the uplands were devoted to sheep grazing with only a little tillage, much as they are today (see Vision Of Britain (2016) for historical maps and details on land use). Furthermore, for larger catchments, such as the River Eden at Carlisle, there is no evidence that local scale effects of land-use change aggregate downstream (O'Connell et al., 2007).

However, natural or man-made changes to the river channel in or near the study area are likely to have a significant impact on local flood levels (Neppel et al., 2010; Shaw et al., 2010, p. 128). For example, the construction of levees will reduce the available flood plain storage resulting in a higher water level at the channel for a given discharge. Similarly, bridges often narrow the channel and restrict flow. Changes to the river bed, from natural processes such as erosion and deposition or anthropomorphic activities from dredging, are also likely to affect the stage-discharge relationship over time (Jalbert et al., 2011).

Quantifying the full effects of this non-stationarity in Carlisle is beyond the scope of this paper. However, a brief review of the literature, including historic maps and images (e.g. Fig. 4) of the area, provides some indicative guidance.

Fig. 3(a) shows that before the historical period of this study, the River Eden at Carlisle comprised two branches separated by an area called 'The Sands'. By 1821 (Fig. 3(b)) the Eden no longer had two branches and 'The Sands' was not an island as the southern branch of the Eden had been filled in (Cooke, 1829). The two bridges shown in the 1805 map (Fig. 3a) were replaced by the 'Eden Bridge', completed in 1815 and still standing (Eden Bridge, 1932). It is clear that the geomorphology of the Eden changed considerably between 1805 and 1821 and construction of the Eden Bridge in that period may also have altered the river's hydraulics.

From the evidence above it can be deduced that the infilling of the southern branch of the Eden and the removal of the earlier bridges over it occurred at some point between 1805 and 1815, when the new Eden Bridge was completed. The impact of these changes on flood levels and discharge is likely to be substantial and largely unquantifiable. Consequently the estimated discharge of 1025 m3 s\(^{-1}\) for the 1809 flood is less certain than the discharge estimates for the subsequent historic floods. For this reason, the discharge estimate for the larger 1809 flood is not used directly; instead, the flood is given broad upper and lower bounds (see Section 3.5).

Although the bridge built in 1815 still stands, its form has changed. There was originally a second new bridge completed contemporaneously to the south of the five arches in existence today. Both bridges can be seen in Fig. 4. The second bridge spanned the earlier, southern branch of the River Eden that was now dry, but still available for channelling floodwater. The arches forming the second bridge were replaced by an embankment in 1969 to support a new road scheme (Cumberland News, 2010). The removal of the arches will have further restricted the flow of floodwater past the Eden Bridge in times of flood.

Similarly flood defences on the floodplain in Carlisle have been built and improved in several stages during the historical period. Smith and Tobin (1979) document the construction of embankments in the early nineteenth and twentieth centuries with further improvements in 1932. More significantly, major improvements to the flood defences were implemented following the 1968 flood, which affected more than 400 properties (Atkins, 2011; CEH, 2014). The removal of the additional bridge arches and the flood defence works carried out over the past century were probably designed to keep the flows in the river. This hypothesis is backed up by expert consultation with a local Flood Risk Management Advisor for the Environment Agency (Environment Agency FRMA, 2014 interview), who explained:

"I suppose over the years we have done a lot of work which will keep the flows in the rivers so flow that was a meter below in 1856 might have had the same flow as we’ve had now, but we don’t know"
There is no systematic information available about within-bank channel changes over the historical period, but it is likely that the practice of dredging to improve river navigation and, to a lesser extent, flood management would have been more common during the historical period than in recent decades, as is suggested by references to the practice in Carlisle in 1948 and 1963 (Carlisle History, 2014). This has been the trend across Britain as cost, environmental concerns, and the shift from river to road transportation have reduced dredging activities (Environment Agency, 2010).

Overall the effect of channel changes is likely to mean that discharge figures calculated for historical floods using the current rating curve formula for the Sheepmount river gauge will be low when compared to the AMAX time series from the gauge data. This would manifest itself as a systematic error/bias in the historic data.

In order to estimate the scale of this bias, simulations of the January 2005 flood were run using a digital elevation model (DEM) that had been altered to remove the post-1968 defences and incorporate the additional Eden Bridge arches. The simulations were run using the 2005 DEM. By increasing the estimated flood levels for the historic floods in Table 1 and recalculating the discharges using the rating Eq. (3), the discharge estimates increase by between 93 and 101 m$^3$s$^{-1}$. Even with this adjustment, none of the adjusted historical floods exceed the peak discharge of the January 2005 flood. This agrees with local expert judgment that the 2005 flood exceeded all floods in the historic period (Environment Agency FRMA, 2014).

### 3.4. Selection of extreme value distribution (EVD)

Several different distributions have been proposed and used to characterise flood frequency curves (Cunnane, 1989). Official river agencies sometimes recommend or even mandate the use of particular distributions they consider most appropriate. In the United States, Bulletin 17B from the United States Water Resources Committee (Kirby and Moss, 1987; USWRC, 1982) recommends using the Log-Pearson Type III (LP3) distribution (Bobée and Ashkar, 1991; Reis and Stedinger, 2005; Vogel and McMartin, 1991). Conversely, in the UK the Flood Estimation Handbook (vol. 3) recommends the generalised logistic (GL or GLO) distribution (IH, 1999), which, together with the generalised extreme (GEV) distribution, is thought to be a satisfactory fit for almost all of the 602 gauged locations in the UK (Kjeldsen et al., 2008).

There is no purely logical reason for preferring any particular extreme value distribution (EVD) for modelling a time series of AMAX river gauge readings. Fisher and Tippett (1928) assert that for a sequence of independent, identically distributed random variables, the asymptotic distribution for the largest value in a block of $n$ values will be of a GEV type as $n$ tends to infinity. However, in the real world, the AMAX time series is not bound to eventually converge on a GEV distribution or any other mathematical function, especially when one considers the many factors that violate the ‘independent, identically distributed’ assumption. Most commonly used distributions have 3 parameters (often describing their location, scale and shape), giving them the flexibility to approximate

### Table 1

| Date       | Estimated level at Eden Bridge (m AOD) | Estimated level at Sheepmount gauging station (m AOD) | Estimated discharge at Sheepmount gauging station (m$^3$s$^{-1}$) | Approx. rank | Notes |
|------------|----------------------------------------|--------------------------------------------------------|---------------------------------------------------------------|-------------|------|
| 2/1869     | 13.6–14.0                              | 12.7                                                   | 1025                                                          | 7–11        | 1    |
| 9/1869     | 13.6                                   | 12.5                                                   | 963                                                           | 12–18       | 1,4  |
| 11/1815    | 13.6                                   | 12.5                                                   | 905                                                           | 25–31       | 2    |
| 2/2/1822   | 14.00                                  | 12.9                                                   | 1089                                                          | 7–11        | 2    |
| 1/1/1851   | 13.41                                  | 12.3                                                   | 905                                                           | 25–31       | 2    |
| 10/2/1852  | 13.53                                  | 12.5                                                   | 941                                                           | 7–11        | 2    |
| 13/12/1852 | 13.83                                  | 12.7                                                   | 1015                                                          | 19–24       | 2    |
| 8/12/1856  | 14.19                                  | 13.1                                                   | 1152                                                          | 2–6         | 2    |
| 11/1857    | 13.45                                  | 12.4                                                   | 947                                                           | 19–24       | 2    |
| 11/1857    | 13.35                                  | 12.5                                                   | 963                                                           | 12–18       | 2    |
| 1/2/1868   | 13.60                                  | 12.5                                                   | 1015                                                          | 7–11        | 2    |
| 7/10/1874  | 13.83                                  | 12.7                                                   | 908                                                           | 25–31       | 2    |
| 1891       | 13.42                                  | 12.3                                                   | 963                                                           | 12–18       | 1    |
| 1899       | 13.6                                   | 12.5                                                   | 1019                                                          | 7–11        | 2    |
| 27/1/1903  | 13.50                                  | 12.4                                                   | 932                                                           | 19–24       | 2    |
| 2/1/1916   | 13.37                                  | 12.3                                                   | 932                                                           | 25–31       | 2    |
| 27/12/1924 | 13.63                                  | 12.5                                                   | 1019                                                          | 12–18       | 2,4  |
| 30/12/1924 | 13.78                                  | 12.7                                                   | 1125                                                          | 7–11        | 2    |
| 2/1/1925   | 14.11                                  | 13.0                                                   | 917                                                           | 25–31       | 2    |
| 21/8/1928  | 13.45                                  | 12.35                                                  | 932                                                           | 19–24       | 2    |
| 29/12/1929 | 13.50                                  | 12.4                                                   | 1116                                                          | 2–6         | 2    |
| 4/11/1931  | 14.08                                  | 13.0                                                   | 950                                                           | 19–24       | 2    |
| 18/10/1954 | 13.56                                  | 12.8                                                   | 902                                                           | 12–18       | 2    |
| 9/12/1964  | 13.40                                  | 12.3                                                   | 902                                                           | 25–31       | 2    |
| 17/10/1967 | 13.28                                  | 12.6                                                   | 1200                                                          | 2–6         | 3    |
| 23/3/1968  | 13.16                                  | 12.4                                                   | 957                                                           | 12–18       | 3    |
| 4/1/1982   | 12.50                                  | 12.3                                                   | 936                                                           | 19–24       | 3    |
| 21/12/1985 | 12.43                                  | 12.4                                                   | 953                                                           | 12–18       | 3    |
| 1/2/1995   | 12.49                                  | 12.35                                                  | 1516                                                          | 1           | 3    |
| 8/1/2005   | 14.15                                  | 1516                                                   | 1029                                                          | 7–11        | 3    |

* The discharge (Q) at the Sheepmount gauging station is calculated from the stage (water level).
a wide range of ‘real’ distributions which are obviously not bound to adhere asymptotically to a mathematical formula. This uncertainty about the relationship of the chosen EVD to the natural processes it is representing, is referred to as model uncertainty (Wood and Rodríguez-Iturbe, 1975).

One approach to managing model uncertainty is to create a composite distribution of several different EVDs, each allocated a weighting based on how well it fits the available data (Apel et al., 2008, 2004; Wood and Rodríguez-Iturbe, 1975). However the goodness of fit tests employed to establish the weightings for candidate EVDs will always be limited because they can only operate within the range of observed data (Keef, 2014, p. 199). Furthermore, as Fig. 6 suggests, the sampling uncertainty often masks the model uncertainty (Cunnane, 1987) such that it is not possible to directly estimate the effect of the model error (Kjeldsen et al., 2014a). For these reasons, we make no attempt to represent model uncertainty using a composite of different distribution types. Instead we selected a single distribution using the ‘goodness of fit’ test proposed by Hosking and Wallis (1997) and described in detail by Kjeldsen et al. (2008).

Table 2 shows the results for the Sheepmount river gauge of the ‘goodness of fit test’ implemented using the WINFAP-FEH software. The test was run both for a single site analysis and a pooled analysis. Under the goodness of fit test, the distribution is considered an acceptable fit to the data if the absolute value of Z is less than 1.64 (Hosking and Wallis, 1997; Kjeldsen et al., 2008), which roughly corresponds to acceptance of the hypothesised distribution at a 90% confidence level (Hosking and Wallis, 1993). Only the GEV and Pearson Type III distributions have an acceptable fit for both single site and pooled analysis. Overall the GEV distribution gives the lowest absolute Z value so we selected it for use in our Bayesian analysis of flood frequency at the Sheepmount river gauge. Within extreme value theory, the GEV distribution is considered a sensible choice of distribution for modelling sequences of independent random variables (Chandler et al., 2014; Embrechts et al., 1997; Jenkinson, 1955) and is commonly used for flood frequency analysis (see for example Gaume et al., 2010; Haberlandt and Radtke, 2014; Kuczera, 1996; Naulet et al., 2005; Neppel et al., 2010).

3.5. Model design

To conduct the Bayesian analysis for this paper, we used the OpenBUGS software (Lunn et al., 2009). BUGS stands for Bayesian inference Using Gibbs Sampling, so the software uses an implementation of the Gibbs sampler (Geman and Geman, 1984), which is a special case of the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953).

If just the gauged AMAX data were available, the likelihood function for gauged data, \( \ell_g(D|\theta) \), in Eq. (1) would be:

\[
\ell_g(D|\theta) = \prod_{i=1}^{s} f(x_i)
\]

In this case \( D \) is the AMAX time series (\( x_1, x_2, x_3, \ldots, x_s \)) of gauged data of length \( s \) years and \( f(x) \) is the probability distribution function (pdf) for \( X \). Here the GEV distribution is used, which has parameters, \( \theta = (\mu, \sigma, \gamma) \), and pdf given by:

\[
f(x; \mu, \sigma, \gamma) = \frac{\Gamma(1+\frac{\gamma}{\sigma}(X-\mu))^{-1}}{\sigma \Gamma(1+\frac{1}{\gamma})} \exp\left(-\frac{1+\gamma(X-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}
\]

The notation for the location, scale, and shape parameters of the GEV distribution is not consistent across disciplines. For example Kuczera (1996) uses \( m, a, \) and \( k \) respectively; Viglione et al. (2013) use \( \theta_1, \theta_2 \) and \( \theta_3 \), and reverse the sign of the shape parameter (\( \theta_3 \)); a common practice in hydrology (Wilks, 2006, p. 105).
Fig. 3. Historic maps of the River Eden at Carlisle. (a–c) from Cumbria Image Bank (Cumbria County Council, 2014). (d) from Carlisle Library © Ordnance Survey (Ordnance Survey, 2011).
For non-zero \( \eta \) and, when \( \eta = 0 \):

\[
f(x : \mu, \sigma, \eta) = \frac{1}{\sigma} \exp\left(-\frac{x - \mu}{ \eta \sigma } \right) \exp\left(-\exp\left(-\frac{x - \mu}{ \eta \sigma } \right) \right)
\]

(6)

When incorporating historic data, it is necessary to take account of the incomplete nature of the dataset. The concept of the ‘perception threshold’ described in Section 3.2 can be defined as ‘the level above which all flood events were recorded in the historic period’. This can be expressed in terms of probabilities as:

\[
Pr(\text{recorded}|X) = 1 \text{ for } X > X_p
\]

Pr(\text{recorded}|X) = 0 \text{ for } X < X_p
\]

(7)

where \( X_p \) is the perception threshold and \( Pr(\text{recorded}|X) \) is the probability of there being a recording of a historical flood in any year given the existence of the ‘true’ annual maximum flood, \( X \). The assumption that no large flood events were missed during the historic period is justifiable in the case of Carlisle due to the relatively extensive local sources listed in Table 3. However, setting a value for the perception threshold isn’t simply a case of choosing a value just below the smallest historical flood event, because of the uncertainty about the discharge estimates for those historic floods.

Fig. 4. 1832 engraving of the Eden Bridge completed in 1815, showing the main five arch span over the Eden and the second bridge over the dry branch of the Eden. From Cumbria Image Bank (Cumbria County Council, 2014).

Reis and Stedinger’s (2005) methods for representing censored data and the method of Viglione et al. (2013) for representing historic floods with upper and lower bounds, the likelihood function for the historical data, \( \ell_h(D|\theta) \) is:

\[
\ell_h(D|\theta) = \left[ \frac{h}{k} \right] F_X(X_p)^{(h-k)} \prod_{j=1}^{k'} \left( 1 - \left[ 1 + \eta \frac{X_j - \mu}{\sigma} \right]^{-1/\eta} \right)
\]

(8)

This contains terms for the \( h-k \) historical floods below the perception threshold, the \( k' \) floods for which an estimate of peak discharge is available, and the \( k'' \) historical floods known to lie between upper and lower bounds. \[ \frac{h}{k} = \frac{\text{histo} \text{. known \. years}}{\text{histo} \text{. known \. years}} \] is the binomial coefficient, and \( F_X(X) \) is the cumulative density function of the GEV distribution given by:

\[
F(x : \mu, \sigma, \eta) = \exp\left(-\left[ 1 + \eta \frac{x - \mu}{\sigma} \right]^{-1/\eta} \right)
\]

(9)

Combining Eqs. (4) and (8) gives the overall likelihood function \( \ell(D|\theta) \):

\[
\ell(D|\theta) = \left[ \frac{h}{k} \right] \prod_{j=1}^{k'} F_X(y_j)^{(h-k)} \prod_{m=1}^{k''} \left( F_X(y_{m+1}) - F_X(y_m) \right) \prod_{i=1}^{s} F_X(x_i)
\]

(10)

Error terms for historical \( y_j \) and gauged \( x_i \) discharge estimates are not shown.

3.6. Uncertainty in discharge estimates

Our Bayesian model for producing the probability distribution functions for the location, scale, and shape parameters of the
GEV distribution needs to take account not only of the sampling uncertainty due to the limited data available but also any random and possibly systematic data errors in the gauged and historical flood estimates. Fortunately the MCMC can easily incorporate data uncertainty without much impact on execution time (Reis and Stedinger, 2005). Various approaches to representing the error in gauged and historical readings have been employed (see for example Neppel et al., 2010), such as a log-normally distributed error for gauged data (Kuczera, 1996) and a uniformly distributed error with upper and lower limits for historical flood estimates (Viglione et al., 2013). The approach for random errors taken here is to assume both gauged and historical discharge estimates have a normally distributed error term associated with them, whose standard deviation represents our confidence in the accuracy of the estimate. Error in the gauge readings is likely to increase with the magnitude of the flood (Di Baldassarre and Montanari, 2010; IH, 1999; Neppel et al., 2010; Rosso, 1985). By using a standard deviation for the error term which is a percentage of the gauge reading, error will scale in proportion to the estimated magnitude of the flood.

Historical estimates of floods are not necessarily affected by this bias. Instead, the uncertainty of historical estimates is affected by other factors such as the time since the event and the richness of the information source. We believe that modelling these factors explicitly would add unnecessary complexity to the model. Consequently the model uses the same standard deviation for the uncertainty in all historical discharge estimates rather than a percentage of the estimate (see Fig. 2b). Additionally the likelihood of a systematic bias in the historical discharge estimates should be considered and is discussed in the following section.

### 3.7. Model parameters and prior distributions

The OpenBUGS configuration settings and parameters used in our Bayesian model are described in Table 3. The first entries in Table 3 are the three prior distributions for the GEV model parameters \( \theta = \{ \mu, \sigma, \eta \} \). Kuczera (1999) recommends using informative priors based on a regional flood frequency analysis. Another approach is to assume prior independence of the GEV parameters using broad, uniform distributions for the location (\( \mu \)) and scale.
parameters ($\sigma$), and then giving some sort of geophysically representative prior for the shape ($\xi$) parameter: a beta distribution in the case of Martins and Stedinger (2000) or a Gaussian distribution in the case of Neppel et al. (2010). For the UK there is no self-evidently representative prior distribution, so we used uniform (non-informative) priors for all three GEV parameters (see Robert and Casella, 2004 for a discussion on choice of priors).

The default value for the gauged data uncertainty is based on the reviewed literature (Environment Agency, 2011c; Horritt et al., 2010), where an error of $\pm 10\%$ is roughly equivalent to a standard deviation of 5%. The default value for the historical adjustment factor is based on the historical flood modelling described in Section 3.1. The uncertainty in the historical floods and the perception threshold are more subjective. We chose confidence limits thought to be wide enough to cover most uncertainties in the historical flood estimates.

### 3.8 Sensitivity analysis

In order to establish how the model behaviour changes as the parameters are altered, we performed a sensitivity analysis. The direct model output consists of the posterior estimates for the mean and standard deviation of the three GEV parameters. The flood frequency curve is then derived with confidence limits around the mean and standard deviation of the three GEV parameters. Altering the model parameters will alter the output, and the extent of this change can be measured.

The flood frequency curve is then derived with confidence limits around the mean and standard deviation of the three GEV parameters. Altering the model parameters will alter the output, and the extent of this change can be measured.
Table 3
Bayes model parameters and prior distributions and BUGS MCMC settings.

| Parameter               | Description                                      | Default value(s) | Range used in sensitivity analysis |
|-------------------------|--------------------------------------------------|------------------|------------------------------------|
| Prior distribution for GEV | µ Location                                        | 0–1600           |                                    |
| distribution parameters | σ Scale                                           | 0–500            |                                    |
|                         | η Shape                                           | −0.5 to 0.5      |                                    |
| Gauged data uncertainty | Uncertainty (standard deviation of normally distributed perturbation) applied to gauged AMAX readings. Percentage of readings. | 5%               | 0–20%                             |
| Historical uncertainty  | Uncertainty (standard deviation of normally distributed perturbation) applied to historical flood estimates. Absolute value. | 50 m³ s⁻¹       | 0–316 m³ s⁻¹                      |
| Historical adjustment factor | Factor applied to estimates of historical flood discharge to take account of construction of flood defences and reduction in dredging activities. | 97 m³ s⁻¹ | 200 to 300 m³ s⁻¹ |
| Perception threshold    | Minimum value of floods in the historic time period included in the historic flood record. | 900 m³ s⁻¹ | 740–1100 m³ s⁻¹ |
| Perception threshold uncertainty | Uncertainty (standard deviation of normally distributed perturbation) applied to perception threshold. Absolute value. | 50 m³ s⁻¹ | 0–140 m³ s⁻¹ |
| Chain length            | Number of MCMC steps per model run                | 20,000           | 20,000–100,000                     |
| Burn-in steps           | Number of MCMC steps disregarded as burn-in       | 5000             | 5000–20,000                       |

4. Results

4.1. Frequency analysis using gauged data only (single site analysis)

Initially we present a ‘single site’ flood frequency analysis performed with the WINFAP-FEH software using just the 46 years of annual maximum discharge time series data (Section 3.1) from the Sheepmount river gauge. The resulting flood frequency curve (referred to as ‘single site analysis’) is shown in Fig. 6 where the GEV, G1, and LP3 distributions have been fitted to the data using the method of L-moments (Hosking, 1990). The 95% confidence intervals for the GEV distribution were calculated by the WINFAP-FEH software using a balanced re-sampling method (Gleason, 1988; WHS, 2009a), which is a commonly used technique of ‘bootstrap’ procedure (Kjeldsen et al., 2014a). The confidence intervals represent the uncertainty of the curve fitting due to the length of the time series: the longer the time series, the narrower the confidence interval would be. Unfortunately, long and reliable time series of AMAX records from river gauges are uncommon in the UK; the 46 years recorded at Sheepmount is comparatively long, the average record length being roughly 35 years (Kjeldsen et al., 2008). The estimated AEP for the peak discharge (1516 m³ s⁻¹) measured during the January, 2005 Carlisle flood is 0.0053, or a return period of 180 years, with 95% confidence intervals of 0.0386 to less than 0.0001, or return periods from 34 to greater than 10,000 years (see Table 6).

Even the enormous width of these confidence limits may be unrealistically narrow; the ‘bootstrap’ method is often found to give lower uncertainty estimates than other, comparable methods using the same data (Hall et al., 2004; Kjeldsen et al., 2014a). This raises questions about how much protection is actually provided by flood defence designs. After the 2005 flood, defences in Carlisle were upgraded to protect against a similar scale event. However, in producing Fig. 9, we have not applied the historical adjustment factor to the December 2015 flood estimate. Also it was necessary to extend the censored period to account for the years 2013, 2014 and 2015 which have not been published yet as part of the official instrumental record. The updated flood frequency curve at the Sheepmount gauging station is shown in Fig. 10. The effect of adding such a large flood to the dataset driving the model has a significant effect on the upper (low frequency event) end of the curve. Table 6 shows how the output of the Bayesian model changes as the new data is added and Fig. 11 compares the flood frequency curves from the Bayesian model both without and including the December 2015 flood.

4.2. Bayesian model results

Running the model with the default parameters gives the posterior distributions for the GEV model parameters described in Table 4. Fig. 7 shows the resulting flood frequency curve with confidence limits plotted using output from the Bayesian model. Fig. 8 transposes the flood frequency curves for the single site analysis using the WINFAP-FEH software (GEV distribution fitted using L-Moments) with the flood frequency curve generated using the Bayesian model. The advantages of the Bayesian model over the single site analysis presented in Section 4.1 can be seen in its narrower confidence limits, which are also evident in the figures for the confidence limits for the design floods shown in Table 5. The Bayesian model gives an AEP for the January, 2005 flood of 0.00382 (return period 262 years) with 95% confidence limits of 0.014 to 0.000246 (return periods: 70 years to 4060 years).

4.3. Incorporating the December 2015 flood in the Bayesian model

The results presented so far have only included data up to 2012, which at the time of writing was the extent of the AMAX time series available from CEH (2014). But the controversy over the recent Storm Desmond flooding of Carlisle in December 2015 has raised important questions about just how unprecedented that event was. The time of writing two estimates for the peak discharge at the Sheepmount gauging station are available: 1680 m³ s⁻¹ (Marsh, 2016) and 1700 m³ s⁻¹ (Parry et al., 2016). These preliminary estimates for the 2015 flood can also be incorporated into our Bayesian model simply by treating the recent flood as an uncertain estimate of an historic event. This enables us to update Fig. 2(b) with Fig. 9. Note that, in producing Fig. 9, we have not applied the historical adjustment factor to the December 2015 flood estimate.

5. Discussion

The magnitude of the flood flows affecting Carlisle in December 2015 was unprecedented, with an AEP and return period we estimate to be 0.00311 and 420 years respectively (95% confidence limits of 0.0119–0.000255 for AEP, 42 to greater than 10,000 years.
for return period). Although the city’s flood defences had been upgraded after the 2005 floods and succeeded in protecting the city against the floods in 2009 (BBC, 2009), they were completely overwhelmed by the scale of the flooding in 2015. Residents whose houses have been flooded twice in just over a decade are now demanding to know why a £38 million investment in new flood defence failed to protect them (BBC, 2015b). Completed in 2010, the new defence scheme had been designed to protect the city against peak flood flows with an estimated AEP of 0.005 (return period of 200 years) (Environment Agency, 2011a). This promised to protect against floods at least as big as the January 2005 event, which was estimated at the time, to have an AEP between 0.0057 and 0.005 (return period of 175–200 years) (Environment Agency, 2006).

But there were substantial uncertainties about those return period estimates, and the December 2015 flood event raises questions about their validity. First, AEP estimates are sensitive to the inclusion of such extreme events in the sample used to calculate the frequency curve. Beyond mere ‘hindsight bias’ (Tversky and Kahneman, 1974), our analysis shows how accounting for the 2015 flood suggests that very large floods of this kind are likely to occur with greater frequency than previously thought. For example, we show that a flood in Carlisle of 1341 m$^3$ s$^{-1}$ which had been expected to occur on average once every 100 years according to our Bayesian model (Table 5) is now expected on average once every 75 years when the model is run to include the December 2015 flood data (Fig. 11).

Second, in the wake of yet another winter of severe flooding in Britain, many people are asking if climate change is now causing such extreme flood events to occur more frequently (Met Office, 2015). Indeed the government has commissioned a review to see if its flood frequency estimates are taking appropriate account of climate change uncertainties (Defra, 2016). There is a substantial body of research into the impacts of long term climate change on river discharge (e.g. Arnell et al., 2014; Fowler and Kilsby, 2007; Hannaford, 2015; Ramsbottom et al., 2012; Reynard et al., 2009). Several studies have sought to attribute specific flood events in the UK to climate change (Kay et al., 2011; Pall et al., 2011), and similar attribution studies of Storm Desmond are under way (van Oldenborgh et al., 2015). The censored time series presented in Fig. 9 is certainly consistent with the hypothesis that extreme floods are becoming more frequent. However, the large uncertainties in the flood frequency curve derived for this paper emphasise the difficulties of attributing individual flood events to global-scale climate changes. This problem of detecting a climate change signal is further confounded by other sources of non-stationarity. Land-use and channel change were discussed in Section 3.3, and there is also evidence of non-stationarity caused by climate variability. For example Pattison and Lane (2011) find evidence that multi-decadal cycles in the North Atlantic climate system are correlated with variations in flood frequencies in Cumbria.

Regardless of the cause, non-stationarity seems to have been accepted as a fact of life (Milly et al., 2008). To assist in understanding this, the use of proxy records from before the historical period can provide valuable information on non-stationarities. Paleo-flood analysis typically involves the use of sedimentology or geomorphology techniques to search for evidence of deposits or erosion from previous events (see for example Baker, 2006; Benito et al., 2004; Li et al., 2013), but other proxy records such as tree ring chronology can also inform researchers on issues of climate and hydrologic stationarity (Razavi et al., 2015). In theory it would also be possible to extend our Bayesian model to take account of one or more paleoflood estimates derived from sediment analysis in the study area. This would be done by adding a ‘prehistoric’ period to join the ‘gauged’ and ‘historic’ periods. The ‘prehistoric’ period would consist of uncertain estimates for the paleofloods and a
second prehistoric perception threshold derived from those paleo-flood estimates. In practice, however, the model might require some redesign to deal with the length of the prehistoric period.

Uncertainties about the impacts of climate change and non-stationarity in the catchment simply add to and compound the many other epistemic uncertainties in flood frequency estimation, which frustrate policymakers needing to prioritise scarce resources for flood risk management and communicate the standards of protection being provided to the public (Kuklicke and Demeritt, 2016). Chief amongst these uncertainties is the sampling error caused by the limited availability of data; symptomatic of this is the sensitivity of the high return period end of the flood frequency curve to the addition of a new extreme event, as demonstrated in Section 4.3. Our analysis shows that this sampling error can be reduced by using historic flood estimates. They are particularly important because they consist of estimates of previous extreme floods and those are the ones we are most interested in because they disproportionately influence the upper end of the flood frequency curve.

In this paper we have shown that these uncertainties can be substantially reduced by incorporating historical flood records into flood frequency analysis. The Bayesian model described in Section 3.5 shows the value of incorporating estimates of historic floods, bringing a reduction in 95% confidence limits of roughly 50% when compared to analysis that uses only the local gauged data (Section 4.2). In addition the model employed here can also incorporate subjective adjustment factors resulting from consultation with stakeholders with local hydrological knowledge. The ability to add possibly subjective, expert hydrological information to the Bayesian model (also see Viglione et al., 2013) can perhaps assuage the concern that flood frequency analysis is dominated by statistical analysis rather than hydrology (Singh and Strupczewski, 2002). This point was borne out during discussions with practitioners responsible for making flood risk estimates in areas they have worked in for many years (Environment Agency FRMA, 2014; Flood Risk Manager, 2014). The ability to incorporate their local expertise could potentially make the resulting estimates

---

**Table 5**

| AEP (return period, years) | 0.5 (2) | 0.0133 (75) | 0.01 (100) | 0.001 (1000) |
|---------------------------|---------|-------------|------------|--------------|
| Mean Discharge (m$^3$s$^{-1}$) (95% confidence intervals) | | | | |
| Single site analysis | 617 (558–678) | 1342 (1036–1853) | 1400 (1056–1983) | 1876 (1123–3295) |
| Bayesian model output | 621 (574–670) | 1289 (1144–1538) | 1341 (1177–1631) | 1766 (1405–2588) |
| Reduction in confidence interval$^a$ | 20% | 52% | 51% | 46% |
| Bayesian model (using gauge data only) | 623 (562–690) | 1408 (1142–2037) | 1471 (1178–2200) | 2000 (1421–4020) |

$^a$ This is the reduction in the width of the 95% confidence interval by using the Bayesian model over the single site analysis.
more ‘physically-based’ in the sense that they take account of the physical events and changes to the channel known to have occurred but otherwise difficult to quantify. This can be done within the Bayesian framework as long as it is communicated transparently how the subjective factors are incorporated in the model and how the model responds. As such flood risk estimates and their associated uncertainty ought to be more plausible insofar as they incorporate, in a systematic and rigorous way, more of the available information about the historical frequency and nature of flooding in the catchment. But the software recommended by the Environment Agency does not support the use of unsystematic (non-gauged) data, so flood frequency estimates making direct use of historic data are currently confined to the academia. Discussions with practitioners responsible for making flood risk estimates in areas they have worked in for many years suggest flood risk managers do indeed take account of local expertise when making flood risk estimates (Environment Agency FRMA, 2014; Flood Risk Manager, 2014; Lane et al., 2011; Odoni and Lane, 2010), but the subjective process is unlikely to be done consistently or transparently if it is not part of the recommended protocol.

Bayesian approaches can overcome both these problems. They enable you to use historic data and to do so transparently so that the resulting impacts on the uncertainty are clear. The results presented in Section 4.2 show a reduction of over 50% in the confidence intervals for the 0.01 AEP flood when our Bayesian model is run using historic data combined with gauged data. If we wanted to mainstream this approach, what would be the next steps? At the moment the model employed for this project has only been used for the data from Carlisle. But there is no reason...
why the model would not work at other sites where there is a mixture of gauged and historic data either in the form of uncertain estimates of historic floods (censored historic time series), or just knowledge of large floods in the past with no estimate of the peak discharge (binomial censored time series) (Reis and Stedinger, 2005). If the model proved adaptable to other sites, it could be packaged up with a ‘user friendly’ interface that would allow more widespread use among flood risk management community.

However, substantial changes are often required to successfully adapt academic software for commercial distribution (The Economist, 2016). As well as improving the user interface, error handling, and support infrastructure, one particular hindrance is that, when using Bayesian MCMC analysis, it can be unclear when the simulation is not converging on the posterior distribution (Lunn et al., 2013, p. 72). For all the model runs in this project, 20,000 iterations with 5000 taken as burn-in was sufficient to achieve convergence for the GEV parameter distributions. But this may not be the case for other locations with different sets of data. There are techniques for formally identifying convergence, for example Plummer et al. (2006), but none are completely reliable (Lunn et al., 2013, p. 75). Even though commercial statistical packages are starting to incorporate Bayesian analysis modules, (for example Stata, 2015), it is still difficult to envisage a software package for estimating flood frequencies using Bayesian analysis that could be used by practitioners with no knowledge of the underlying techniques.

Furthermore, for the Carlisle case study, our job as researchers was facilitated by the fact that the historical flood data had already been collated from local archives and published. In order to mainstream our Bayesian model, a practitioner would need to gather the following inputs:

1. A historic catalogue of flood events. A useful starting point may be the national database of historical records produced by the British Hydrological Society (Black and Law, 2004), but the practitioner may be required to supplement that data with records from local archives.
2. Historic changes to the catchment and channel that may have affected the intensity of floods and the stage-discharge relationship. This too may require searching local archives then testing the sensitivity of hydrological or hydraulic models to the changes.

| Flood event | Peak Discharge (m$^3$ s$^{-1}$) | January 2005 | December 2015 |
|-------------|--------------------------------|--------------|---------------|
|             | 1516                           | 1690         |
| AEP (95% CL)| 0.0055 (0.0386–0.0001)         | 0.00238 (0.0185–0.0001) |
| Return period (95% CL) | 180 (34–<10,000) | 420 (54–<10,000) |

Table 6

Estimates of AEP and return periods (in years) for the January 2005 and December 2015 floods at Carlisle. The data show how the flood frequency curve is altered as new flood data becomes available.

Fig. 11. River Eden at Sheepmount flood frequency curves with 95% confidence limits showing comparison of curves derived using the Bayesian model both including and without the December 2015 flood.
3. Estimates of the peak discharge associated with historic flood events. The task of creating a censored time series of historical flood estimates would be daunting for a practitioner starting from scratch, who was unlikely to have been trained in the methods required for scouring local archives for records of flooding.

4. Perception threshold. The sensitivity analysis of our model (Section 3.8) shows that the identification of a realistic perception threshold is important when building a censored historical flood record. The estimation of the perception threshold was facilitated in the case of Carlisle due to the presence of the Eden Bridge which provided a benchmark for much of the historical period.

In the UK it may be feasible to identify a number of sites that do satisfy these criteria. A cursory review of previous studies in the UK suggest locations on the Rivers Tyne (Archer et al., 2007), Trent (Macdonald, 2013), Ouze in Sussex (MacDonald et al., 2013) and Yorkshire (Macdonald and Black, 2010) and Tay in Scotland (Macdonald et al., 2006) would be suitable for the use of our Bayesian model. Further research should utilise a framework such as that employed by Renard et al. (2013) to compare, across several case studies, this method of frequency analysis with others that can incorporate historical data, e.g. the EMA (Cohn et al., 2001).

6. Conclusion

While uncertainty is endemic to all estimates of flood risk, the Bayesian approach developed in this paper offers a method both for reducing its scale and managing it more transparently. In the case of Carlisle, incorporating historic data into our Bayesian model reduced the size of 95% confidence intervals by roughly 50% for annual exceedance probabilities of less than 0.0133 (return periods over 75 years) compared to standard flood frequency estimation methods using solely gauge data. Several other locations in the UK have been identified as possible additional case studies, but even if these localities do not show similar reductions in uncertainty from incorporating historic data, using a formal Bayesian framework still provides greater transparency in flood frequency estimation.

The two recent floods in Carlisle have highlighted the danger in not taking explicit account of uncertainty when planning defences and communicating their limits to residents. Although there is clearly great promise in using models that can incorporate all available data sources, more work is still required to mainstream the Bayesian approach tested here and to develop decision-support systems for helping making it available to users with no specialist knowledge of the statistical techniques behind the model.

Acknowledgements

Brandon Parkes was funded by the Economic and Social Research Council (ESRC) (reference: ES/I004297/1ES/I004297/1) as a joint interdisciplinary ESRC/NERC studentship. Funding support is also acknowledged from the European Commission's KULTURisk project (http://www.kulturisk.eu) and the Natural Environment Research Council (NE/K008595/1): DEMERITT D, Malamud D (2013–2017) Susceptibility of catchments to INTense Rainfall and flooding (SNITRA).

We would like to thank Dr Paul Smith and Dr Florian Pappenberger of the ECMWF, and Professor Hannah Cloke of Reading University for reviewing an earlier version of this paper and critiquing our Bayesian statistical model.

References

Acreman, M.C. 1989. Extreme historical UK floods and maximum flood estimation. J. Inst. Water Environ. Manage. 3 (4), 404–412.
Apel, H., Merz, B., Thieken, A.H. 2008. Quantification of uncertainties in flood risk assessments. Int. J. River Basin Manage. 6 (2), 149–162.
Apel, H., Thieken, A.H., Merz, B., Blöschl, G. 2004. Flood risk assessment and associated uncertainty. Copernicus Publications on behalf of the European Geosciences Union. Nat. Hazards Earth Syst. Sci. 4 (2), 295–308.
Archer, D.R., Leech, F., Harwood, K. 2007. Assessment of severity of the extreme River Tyne flood in January 2005 using gauged and historical information. Hydrolog. Sci. J. 52 (5), 992–1003.
Arnell, N.W., Charlton, M.B., Lowe, J.A. 2014. The effect of climate policy on the impacts of climate change on river flows in the UK. J. Hydrol. 510, 424–435. http://dx.doi.org/10.1016/j.jhydrol.2013.12.046.
Atkins, N. 2011. Carlisle SFRA Final Report.
Baker, V.R. 2006. Palaeoflood hydrology in a global context. Catena 66 (1), 161–168.
Bayes, T., Price, R., 1763. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F.R.S. Philos. Trans. Royal Soc. 53, 370–418.
Beven, A.C., Reed, D.W., 2001. The use of historical data in flood frequency estimation. Report to MAFF, Centre for Ecology and Hydrology, Wallingford.
BBC, 2009. More than 200 People Rescued in Floods in Cumbria Town. More than 200 People Rescued in Floods in Cumbria Town. BBC News.
BBC, 2010. Carlisle flood defence scheme completed Available at <http://www.bbc.co.uk/news/2013-06-24> (accessed 26 April 2016).
BBC, 2015a. Storm Desmond: Defences Against Indefensible Floods Available at <http://www.bbc.com/news/science-environment-35028180> (accessed 26 April 2016).
BBC, 2015b. Storm Desmond: Thousands of people flooded out of homes Available at <http://www.bbc.com/news/uk-35023558> (accessed 26 April 2016).
Benito, G. et al., 2004. Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. Review of scientific methods. Natural Hazards 31 (3), 623–643.
Bertsch-McGrayne, S., 2011. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Russian Submarines & Emerged Triumphant From Two Centuries of Controversy. Yale University Press, New Haven.
Beven, K., 2008. Environmental Modelling: An Uncertain Future? Taylor & Francis.
Beven, K., Hall, J., 2014. Applied Uncertainty Analysis for Flood Risk Management. Imperial College Press, London.
Black, A.R., Burns, J.C., 2002. Re-assessing the flood risk in Scotland. Sci. Total Environ. 294 (1), 169–184.
Black, A.R., Law, F.M., 2004. Development and utilization of a national web-based chronology of hydrological events/Développement et utilisation sur internet d’une chronologie nationale d’événements hydrologiques. Hydrolog. Sci. J. 49 (2).
Bobée, B., Ashkar, F., 1991. The Gamma Family and Derived Distributions Applied in Hydrology. Water Resources Publications, Littleton ^eColorado Colorado.
Brisz, R., Zbigniew, W.K., Benito, C., 2006. Historical hydrology for studying flood risk in Europe. Hydrolog. Sci. J. 51 (5), 739–764.
Carlisle History, 2014. Carlisle Encyclopaedia. <www.carlislehistory.co.uk> (accessed 12 November 2014).
Cartillo, E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M., 2005. Extreme Value and Related Models With Applications in Engineering and Science. Wiley Chichester.
CEH, 2014. Centre for Ecology and Hydrology, National River Flow Archive. Eden Sheepeamton Gauge Details. <http://www.ceh.ac.uk/data/mrf/data/station.html?stid=76007> (accessed 29 September 2014).
Chandler, R.E., 2013. Exploiting strength, discounting weakness: combining information from multiple climate simulators. Philos. Trans. Royal Soc. A – Math. Phys. Eng. Sci. 371 (1991). http://dx.doi.org/10.1098/rsta.2012.0388.
Chandler, R.E. et al., 2014. Uncertainty in rainfall inputs. In: Beven, K., Hall, J. (Eds.), Applied Uncertainty Analysis for Flood Risk Management. Imperial College Press, London.
Cohn, T., Lane, W., Baier, W., 1997. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. Water Resour. Res. 33 (9), 2083–2096.
Cohn, T.A., Lane, W.L., Stedinger, J.R., 2001. Confidence intervals for expected moments algorithm flood quantile estimates. Water Resour. Res. 37 (6), 1695–1706.
Coles, S., Bava, J., Trenner, L., Durazio, P., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.
Condle, R., Lee, K.A., 1982. Flood frequency analysis with historic information. J. Hydrol. 58 (1), 47–61.
Cong, S.Z., Xu, Y.B., 1987. The effect of discharge measurement error in flood frequency-analysis. J. Hydrol. 96 (1–4), 237–254. http://dx.doi.org/10.1016/0022-1694(87)90156-9.
Cooke, G.A., 1829. Topography of Great Britain or, British Traveller’s Pocket Dictionary, Containing Westmorland and Cumberland, vol. XXII. Sherwood, Neely and Jones, Paternoster Row, London.
Cumberland News, 2010. A bridge too far for Carlisle folk. <http://www.cumberlandnews.co.uk/a-bridge-too-far-for-carlisle-folk-1.7028187> (accessed 12 November 2014).
Cumberland News.
Cumbria County Council, 2014. Libraries and Archive Service – Cumbria Image Bank cumbriaimagebank.org.uk (accessed 12 November 2014).

Cumberland News, 2010. A bridge too far for Carlisle folk. <http://www.cumberlandnews.co.uk/a-bridge-too-far-for-carlisle-folk-1.7028187> (accessed 12 November 2014).
Cumbria Image Bank cumbriaimagebank.org.uk (accessed 12 November 2014).

Cumberland News.
Cumbria County Council, 2014. Libraries and Archive Service – Cumbria Image Bank cumbriaimagebank.org.uk (accessed 12 November 2014).
Cumbria Image Bank cumbriaimagebank.org.uk (accessed 12 November 2014).
Cumbria Image Bank...
MacDonald, N., Kjeldsen, T.R., Proedocomi, I., Sangster, H., 2013. Reassessing flood frequency For the Sussex Ouse, Lewes: the Inclusion of historical flood information since AD 1650. Nat. Hazards Earth Syst. Sci. Discuss. 1, 7615–7646.

MacDonald, N., Werrity, A., Black, A., McEwen, L., 2009. Historical and pooled flood frequency analysis for the River Tay at Perth, Scotland. Area 38 (1), 34–46.

Marsh, T., 2016. UK Hydrological Bulletin: November 2015 – January 2015. The late autumn and early winter floods of 2015/2016. Circula.t – Newsletter Br. Hydrol. Soc. 128 (1), 17–27.

Martins, E.S., Stedinger, J.R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resour. Res. 36 (3), 514–524. http://dx.doi.org/10.1029/1999wr900303.

McEwen, L.J., 1987. Sources for establishing a Historic Flood Chronology (Pre-1940) within Scottish River Catchments. Scottish Geogr. Mag. 103 (1), 132–140.

McEwen, L.J., 1990. The establishment of a historical flood chronology for the River Tweed through combined documentary and flood level information. Hydrol. Sci. J. 35 (1), 37–48.

Merz, B., Thieken, A.H., 2009. Flood risk curves and uncertainty bounds. Nat. Hazards 51 (3), 437–458.

Met Office, 2015. Met Office News Blog. Did Climate Change have an Impact on Storm Desmond? Available at <http://blog.metoffice.gov.uk/2015/12/07/did-climate-change-have-an-impact-on-storm-desmond/> (accessed 26 April 2016).

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 (6), 1087–1092. http://dx.doi.org/10.1063/1.1699114.

Miller, J., Kjeldsen, T., Hannafor, J., Morris, D., 2012. A Hydrological Assessment of the November 2009 Floods in Cumbria, UK.

Milly, P.C.D., 2014. Stationarity is dead: whither water management? Science 319, 573–574. http://dx.doi.org/10.1126/science.1200270.

Nauel, R., Lang, M., Ouarda, T., Coeur, D., Bobée, B., Recking, A., Moussay, D., 2005. Flood frequency analysis on the Archebre river using French documentary sources from the last two centuries. Hydrol. Sci. J. 51 (1), 58–78.

Neal, J.C., de Fre, C., Bates, B.C., Beven, K., Leedal, D., 2012. Probabilistic flood risk mapping including spatial dependence. Hydrolog. Process. 27, 1349–1363. http://dx.doi.org/10.1002/hyp.7952.

Neppel, L., et al., 2010. Flood frequency analysis using historical data: accounting for random and systematic errors. Hydrol. Sci. J. – J. Des Sci. Hydrol. 55 (2), 192–204. http://dx.doi.org/10.1002/hyp.6909.

NERC, 1975. (National Environment Research Council) Flood Study Committee. O’Connell, D.R., Oostena, D.A., Levi, D.R., Klinger, R.E., 2002. Bayesian flood frequency analysis with paleohydrologic bound data. Water Resour. Res. 38 (5), 1263–1271. http://dx.doi.org/10.1029/2001wr000916.

O’Connell, D.R., Oostena, D.A., Levi, D.R., Klinger, R.E., 2002. Bayesian flood frequency analysis with paleohydrologic bound data. Water Resour. Res. 38 (5), 1263–1271. http://dx.doi.org/10.1029/2001wr000916.

O’Connell, D.R., Stedinger, J.R., 2000. Nonparametric Bayesian flood frequency estimation. J. Hydrol. 231 (1–2), 79–96. http://dx.doi.org/10.1016/j.jhydrol.2000.05.005.

O’Connell, P., Ewe, J., O’Connell, G., Quinn, P., 2007. Is there a link between agricultural land-use management and flooding? Hydrol. Earth Syst. Sci. 11 (1), 96–107.

Odom, N.A., Lane, S.N., 2010. Knowledge-theoretic models in hydrology. Prog. Phys. Geogr. 34 (2), 117–137. http://dx.doi.org/10.1177/0309133310370215.

Ordnance Survey, 2011. Ordnance Survey MasterMap Database Available online at <http://www.ordnancesurvey.co.uk/oswebsite/products/os-mastermap/index.html> (accessed 26 April 2016).

Parkes, B.L., Cloke, H.L., Pappenberger, F., Neal, J., Demeritt, D., 2013. Reducing uncertainty in Flood Risk and its Implications for Management: An Offshoot of a Climate-flooding feedback. Hydrol. Process. 16 (18), 3737–3740. http://dx.doi.org/10.1002/hyp.5083.

Parkes, B.L., 2015. Uncertainty in Flood Risk and its Implications for Management and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. Water Resour. Res. 42 (12). http://dx.doi.org/10.1002/2015wr015696.

Parkes, B.L., Cloke, H.L., Pappenberger, F., Neal, J., Demeritt, D., 2013. Reducing uncertainty in Flood Risk and its Implications for Management: An Offshoot of a Climate-flooding feedback. Hydrol. Process. 16 (18), 3737–3740. http://dx.doi.org/10.1002/hyp.5083.

Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods, second ed. Springer Science & Business Media.

Rosso, R., 1985. A Linear-approach to the influence of discharge measurement error on flood estimates. Hydrol. Sci. J. – J. Des Sci. Hydrol. 55 (2), 192–204. http://dx.doi.org/10.1002/hyp.6909.

Russell, D., 1985. A Linear-approach to the influence of discharge measurement error on flood estimates. Hydrol. Sci. J. – J. Des Sci. Hydrol. 55 (2), 192–204. http://dx.doi.org/10.1002/hyp.6909.

Seibert, B., Kharin, V., 1999. Extreme-value quantile estimators for hydrologic data. Water Resour. Res. 36 (5), 1263–1271. http://dx.doi.org/10.1029/1999wr900397.

Seo, S., Ouarda, T.B.M.J., Bruneau, P., Bobee, B., 2006. A parametric Bayesian combination of local and regional information in flood frequency analysis. Water Resour. Res. 42 (11). http://dx.doi.org/10.1029/2005wr004397.

Shaw, E., Beven, K., Chappell, N.A., Lamb, R., 2010. Hydrology in Practice. Taylor & Francis.

Singh, V.P., Strupczewski, W.G., 2002. On the status of flood frequency analysis. Hydrol. Process. 16 (18), 3737–3740. http://dx.doi.org/10.1002/hyp.5083.

Smith, K., Tobin, G., 1979. Human Adjustment to the Flood Hazard. Longman, New York.

Stata, 2015. Stata: A Bayesian analysis of flood frequency. Stoch. Env. Res. Risk Assess. 29 (10), 1430–1449. http://dx.doi.org/10.1007/s00477-014-0957-2.

Tatur, A., Khanna, M., 1974. Judgment under uncertainty – heuristics and biases. Science 185 (4157), 1241–1248. http://dx.doi.org/10.1126/science.185.4157.1241.

Tufte, E.J., 1983. Charting Data. Of More Than Academic Interest, The Economist. Pearson, London. Available online at<http://www.economist.com/_.

Vogel, R., McMartin, D.E., 1991. Probability plot goodness-of-fit and skewness estimation for hydrologic data. Water Resour. Res. 27 (8), 1051–1060. http://dx.doi.org/10.1029/91wr015696.

WHS, 2009a. Wallingfor HydroSolutions Ltd. WINFAP-FEH 3 software.
Williams, A., Archer, D., 2002. The use of historical flood information in the English Midlands to improve risk assessment. Hydrol. Sci. J. 47 (1), 57–76.

Wood, E.F., Rodríguez-Iturbe, I., 1975. A Bayesian approach to analyzing uncertainty among flood frequency models. Water Resour. Res. 11 (6), 839–843.