Natural flood management potential of leaky dams in upland catchments

Zora Ruth van Leeuwen

School of Geography

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

31st October 2021

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to express my gratitude to my academic supervisors Professor Lee Brown, Dr Megan Klaar and Dr Mark Smith who helped me finalise my thesis. This research would not have been possible without the support of Professor Rob Lamb (JBA Trust) whose insights helped guide me along the way. During their time at the Yorkshire Dales Rivers Trust, Dan Turner and Jack Hirst gave me a practitioner's perspective, and helped me to build leaky dams in my study site. Thanks also go to Alex Scott and Steve Rose (JBA Trust & JBA Consulting) for inception of the project and to the Mawle family for continued support of research on their land.

I will be forever grateful to the many friends and colleagues who helped me to set up the monitoring network in Coverdale, conduct surveys, build leaky dams and especially collect data during storms. Conversations with friends, family and fellow researchers at the University of Leeds helped to shape this research, particularly I'm grateful to Dr Eleanor Pearson for conversations about NFM, Dr Matthew Gaddes & Jess Boyd for patiently explaining statistics and R to me, David Smith for opening the door to time series modelling and Dr Michael Watson for casting a critical eye over my analysis.

For their support I'd like to thank Rosie Hadfield, Dr Lisa Watson, Libby Jones, Dr Becca Chandler-Bostock, Fiona Lines and her girls and Cordula Wittekind. For their care, advice and all the encouragement I'm grateful to my family. I can't thank my partner, James Malloch, enough for keeping me going and helping me in every way that he could.

Abstract

Natural Flood Management (NFM) is an increasingly popular approach to flood risk management which aims to slow and store flood water in the landscape by restoring natural hydrological and geomorphological processes.

Due to difficulties in obtaining precise hydrological data, a lack of robust, empirical evidence describing the effectiveness of NFM measures at reducing downstream flood risk currently undermines confidence in its efficacy and limits its adoption. Furthermore, uncertainty about the risks associated with NFM features can present a barrier to implementation.

This research quantifies the benefits and risks of installing engineered leaky dams for the purpose of NFM in an upland, headwater catchment in North Yorkshire, England. To overcome difficulties associated with the empirical quantification of leaky dam impacts on flood peak magnitude, hydrological data from a three year Before-After-Control-Impact style monitoring campaign was combined with a data-based time series modelling approach.

The results quantify, for the first time, the impact of a series of eight channel-spanning leaky dams on flood peak magnitude for a range of events. The leaky dams reduced flood peak magnitude of events with a return period up to one year by 10% on average, but their impacts were highly variable. In order for the benefits provided by leaky dams to be evaluated against their potential hazards the study included assessment of the failure probability of leaky dams based on observations of leaky dam failures from five UK sites. The empirical

fragility analysis showed that leaky dam failure rates were lower than naturally occurring large wood mobility rates reported in literature.

The novel application of both the data-based timeseries modelling approach and empirical fragility analysis to assess the potential of leaky dams for NFM demonstrates the opportunities that these techniques offer to quantify leaky dam impacts on downstream flood risk.

Abbreviations

ACF	Auto Correlation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
AMAX	Annual Maximum
AOD	Above Ordnance Datum
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroscedasticity
ARMA	Autoregressive Moving Average
BACI	Before After Control Impact
BIC	Bayesian Information Criterion
CaBA	Catchment Based Approach
CAP	Common Agricultural Policy
CBFM	Catchment-Based Flood Managament
CCF	Cross Correlation Function
CIRIA	Construction Industry Research and Information Association
DBM	Data Based Mechanistic
DEFRA	Department for Envrionment, Food & Rural Affairs
EA	Environment Agency
EFRA	Environment, Food and Rural Affairs Committee
ELMS	Environmental Land Management Scheme
EU	European Union
FCERM	Flood & Coastal Erosion Risk Management
FRA	Flood Risk Assessment
FRM	Flood Risk Management
FSIC	Finite Sample Information Criterion
GARCH	Generalised Autoregressive Conditional Heteroscedasticity
GEV	Generalised Extreme Value
IQR	Inter Qaurtile Range
ISO	International Organisation for Standardisation
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
KS	Kolmogorov-Smirnov
LLFA	Lead Local Flood Authority

LOESS	locally weighted smoothing
LW	Large Wood
LWD	Large Woody Debris
MA	Moving Average
MAICE	Minimum Akaike Information Criterion Estimation
MLE	Maximum Likelihood Estimation
NBS	Nature Based Solutions
NFM	Natural Flood Management
NFU	National Farmers Union
NGR	National Grid Reference
NSE	Nash-Sutcliffe Efficiency
OS	Ordnance Survey
PACF	Partial Autocorrelation Function
PDF	Probability Density Function
PE	Peak Error
PEP	Percentage Error in Peak
POT	Peak Over Threshold
PRAGMO	Practical River Restoration Appraisal Guidance for Monitoring Options
PTE	Peak Timing Error
QQ	Quantile-Quantile
RMSE	Root Mean Square Error
RRC	River Restoration Centre
RSE	Relative Standard Error
SDP	State Dependent Parameter
SE	Standard Error
SEPA	Scottish Environment Protection Agency
TFN	Transfer Function Noise
UK	United Kingdom
US	United States
USA	United States of America
USDA	United States Department of Agriculture
WFD	Water Framework Directive
WMO	World Meteorological Organisation
WwNP	Working with Natural Processes

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Drivers of increased flood risk in the UK

Increased autumn and winter rainfall has led to increased flooding across north-west Europe (Blöschl et al., 2019). Although it is difficult to obtain definitive conclusions attributing increased flood risk to climate change (Hannaford, 2015; Kay et al., 2011; Schaller et al., 2016), it is generally accepted that flood risk is increasing globally due to the effects of climate change (Hirabayashi et al., 2013) and socio-economic drivers such as land-use change (Winsemius et al., 2016). In the UK, climate change projections predict increases in the frequency of extreme rainfall events with increases in both short, high intensity rainfall and prolonged rainfall events (Defra, 2018; Defra et al., 2018; Thompson et al., 2017) which lead to pluvial, fluvial and groundwater flooding (Committee on Climate Change, 2017). The Committee on Climate Change has listed increases in flood risk as one of the greatest climate change related threats facing the UK (Committee on Climate Change, 2017).

Non-climatic factors affecting flood risk in the UK are two-fold; firstly, it is widely accepted that increased urbanisation and development have increased the risk and economic impacts of floods worldwide (Jackson et al., 2008; Mudelsee et al., 2003; Wheater, 2006). Secondly, agricultural intensification and land management change are perceived to increase flooding (Evrard et al., 2007; Jackson et al., 2008; O'Connell et al., 2007; Pinter et al., 2006; Wheater, 2006). Such

practices included removing hedgerows to create larger fields, cultivation practices which compact soils, land drains and ditches which increase drainage, channelising rivers and removal of their riparian buffer zones (O'Connell et al., 2007). These major land use and management changes were supported by agricultural subsidies following a drive for self sufficiency in food production after the second World War (O'Connell et al., 2007). As a result, large parts of the UK may not generate and route runoff, or hydrologically function, as natural landscapes (O'Connell et al., 2007).

1.2 Flood risk management policy in the UK

Although it is widely accepted that the drivers of increased flood risk are socio-environmental, flood risk is most commonly managed in the UK by physically altering rivers and floodplains (Cook et al., 2016; Lane et al., 2011; Purseglove, 2015; Wescoat and White, 2003), particularly favouring hard-engineered defences (Brown and Damery, 2002; Wilby et al., 2008). However, catchment scale approaches, which allow the effects of land use and management to be taken into account, are increasingly being considered in the UK's approach to flood risk management (Wilby et al., 2008).

In 2007 the wettest summer on record was recorded in England (Pitt, 2008). The flooding resulted in 13 fatalities, 55,000 flooded properties and over £3 billion in pay out by the insurance industry (Pitt, 2008). The devastation was so widespread that the floods were professed to be the country's *"largest peacetime emergency since World War II"* (Pitt, 2008, p. vii). In response to this, the Government requested a comprehensive, independent review to be carried out into the 2007 floods. The highly influential policy review, *'Learning lessons from the 2007 floods'*, was completed by Sir Michael Pitt in 2008. The review emphasised the potential of flood defences which work with natural processes to store water away from urban areas to slow down and reduce runoff. Reducing flood risk by restoring or emulating natural floodplain, river and coastal processes is termed Working with Natural Processes (WwNP) to reduce flood risk (Burgess-Gamble et al., 2017). One of the recommendations of the report was that the UK Government's Department for Environment and Rural Affairs (Defra), the Environment

Agency (EA) and Natural England should work with their partners to establish a programme to deliver greater working with natural processes to reduce flood risk. (Pitt, 2008).

Many of the recommendations made in the review have been written into flood management policy. The Flood Water and Management Act (England and Wales) 2010 advocates working with natural processes and in Scotland the Flood Risk Management Act (Scotland) 2009 made it a requirement to consider working with natural processes in flood management schemes. In 2009, Defra funded three multi-objective flood management demonstration projects with the aim to generate quantitative evidence to demonstrate the effectiveness of implementing a range of measures which work with natural processes to reduce flood risk, alongside delivering a range of environmental and social benefits for communities (National Trust, 2015).

More recently, the National Flood Resilience Review was composed in response to extensive flooding in December 2015, which caused half a billion pounds of direct damages (Calderdale Council, 2016). It drew attention to the importance of better management of rainfall in the natural environment to relieve both flood risk and water stress (Government, 2016). The review stated that changes would be made to the way funding was allocated for flood risk management so that Natural Flood Management (NFM) would be able to compete against engineered defences (Government, 2016). NFM is a term which has been coined to describe working with natural processes to reduce flood risk. Natural flood management measures aim to manage the sources and pathways of flood waters by working with and restoring the natural hydrological and morphological processes of catchments (Forbes et al., 2015).

The Environment, Food and Rural Affairs Committee (EFRA), commissioned by the House of Commons, called for an overhaul of the Government's approach to flood risk management by proposing a new governance model (House of Commons Environment Food and Rural Affairs Committee, 2016). One of the three main flood management problems identified in the report was the need for greater adoption of catchment management measures. Specifically, EFRA called on Defra to launch a catchment wide study into Natural Flood Management (NFM)

measures and to work with the National Farmers Union (NFU) to develop a system by which farmers could be incentivised to store flood water on their land (House of Commons Environment Food and Rural Affairs Committee, 2016).

The latter point was, in part, addressed by an update of countryside stewardship agreements. In 2011 Defra completed an upland policy review which identified the multiple benefits, including flood risk management, which can be provided by the uplands. Defra made a commitment to supporting hill farmers to diversify *"as managers of the natural resources and ecosystems of the uplands"* (Defra, 2011b, p. 4). As a result new Countryside Stewardship agreements, active since 2016, included a range of options for natural flood management.

The 2000 Water Framework Directive (WFD) and 2007 EU Floods Directive already required the coordinated production and implementation of river basin management plans and flood risk management plans respectively. Over the past two decades the WFD, which requires that all waterbodies meet good ecological status or potential by 2027, has been the key legislative driver for river restoration in the UK (England et al., 2008; Skinner and Bruce-Burgess, 2005; Smith et al., 2014). The WFD and EU Floods Directive, together with Defra's vision set out in the 2011 White Paper, to value the services and resources that nature provides (Defra, 2011a), led to a more integrated, catchment based approach to managing the water environment. The Catchment Based Approach (CaBA) aimed to bridge the gap between the national scale river basin management plans, and local scale restoration opportunities by establishing a clear, evidence based agenda at the catchment scale, which is implemented through a variety of projects at smaller geographical scales (Defra, 2013). The CaBA was thereby designed to change the way river restoration was initiated from opportunistic, as observed by Skinner and Bruce-Burgess (2005) in the early stages of the WFD, to a more co-ordinated, concept driven approach.

The Woodland for Water project is an example of such a co-ordinated approach; it mapped opportunities for the reduction of pollution and/or flood risk through tree planting at a nation wide scale. In Yorkshire and North East England over 3000 km^2 of priority areas for planting for flood risk reduction were identified, including in upland areas in the Yorkshire Dales and Cheviot Hills (Broadmeadow and Nisbet, 2013).

Following the UK's exit from the European Union (EU), the Environment Bill, House of Commons (57, 2019-21 & 2021-22), and the Agriculture Act (2020) aim to legislate the intention to take a holistic approach to land and water management (Klaar et al., 2020). The Government have proposed to replace the EU's Common Agricultural Policy (CAP) with the Environmental Land Management Scheme (ELMS) which aims to pay landowners directly for the provision of public goods, including flood risk management (Defra, 2020). ELMS aims to fulfil the requirements of the Government's 25 year Environment plan to reduce the risk of harm from flooding whilst using natural resources more sustainably, mitigating climate change and protecting biodiversity and wildlife (Defra, 2018). Deriving such multiple benefits from ecosystem functions, termed 'ecosystem services' (Costanza et al., 1997), is encouraged by the proposed 'public money for public goods' approach of the ELMS (Klaar et al., 2020). Hence, whilst NFM has been evident in UK policy since 2005 (Defra, 2005) it is likely to become more mainstream if the UK is able to successfully implement its "public money for public goods" approach set out in the Agriculture Act.

1.3 Natural Flood Management

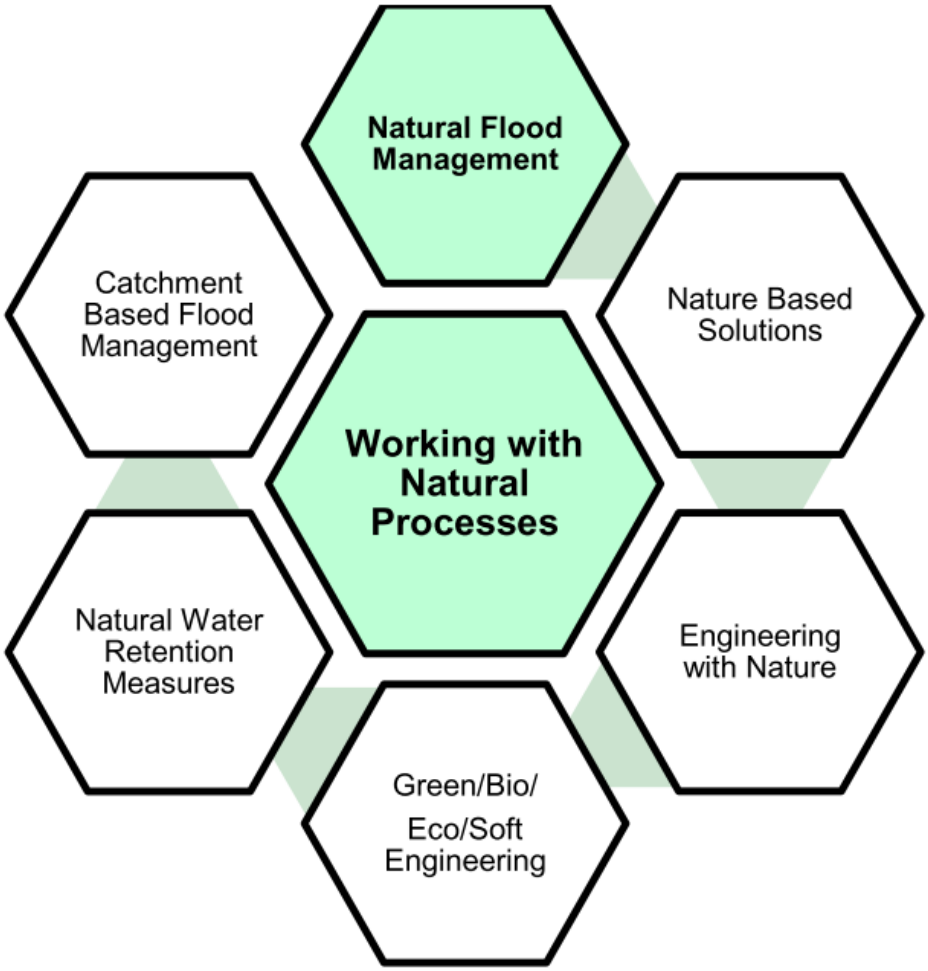


Figure 1.1: Terminology used to denote methods which work with natural processes to manage flood risk, from Burgess-Gamble et al. (2017, p. 4)

Working with natural processes (WwNP) is a broad term for work which protects, restores and emulates natural processes (Burgess-Gamble et al., 2017). Nature Based Solution (NBS) is a subset of WwNP relating to works which deliver environmental benefits, which can include flood risk management. Natural Flood Management (NFM) is a NBS which refers specifically to interventions which intend to reduce flood risk. A consensus on the definition of the term

NBS is lacking, and therefore the framework within which it is used can differ. Nesshöver et al. (2016) argue it is emerging as the catch-all term for actions which use or are inspired by nature to address environmental, social and economic issues, which includes WwNP and NFM. These terms, amongst others (Figure 1.1), are often used interchangeably to refer to NFM (Dadson et al., 2017).

NFM is underpinned by the principles of Catchment-Based Flood Management (CBFM) in which changes are made to the wider catchment to manage downstream flood risk, rather than managing flooding locally where it occurs (Dadson et al., 2017). NFM aims to restore or emulate the natural functioning of river catchments to increase infiltration, slow flows and store water (Forbes et al., 2015). Alternative definitions include "the alteration, restoration, or use of landscape features for the purposes of reducing flood risk" (Parliamentary Office of Science and Technology, 2011, p. 1), or as a subset of CBFM which "seeks to restore or enhances catchment processes that have been affected by human intervention" (Dadson et al., 2017, p. 2). NFM measures can be grouped into three main groups: Woodland creation, land management, and river and floodplain restoration (Forbes et al., 2015). These measures, illustrated in Figure 1.2, reduce the rapid generation of runoff from hillslopes, store water during high flows and 'slow the flow' by decreasing the connectivity between runoff sources and areas susceptible to flooding (Lane, 2017).

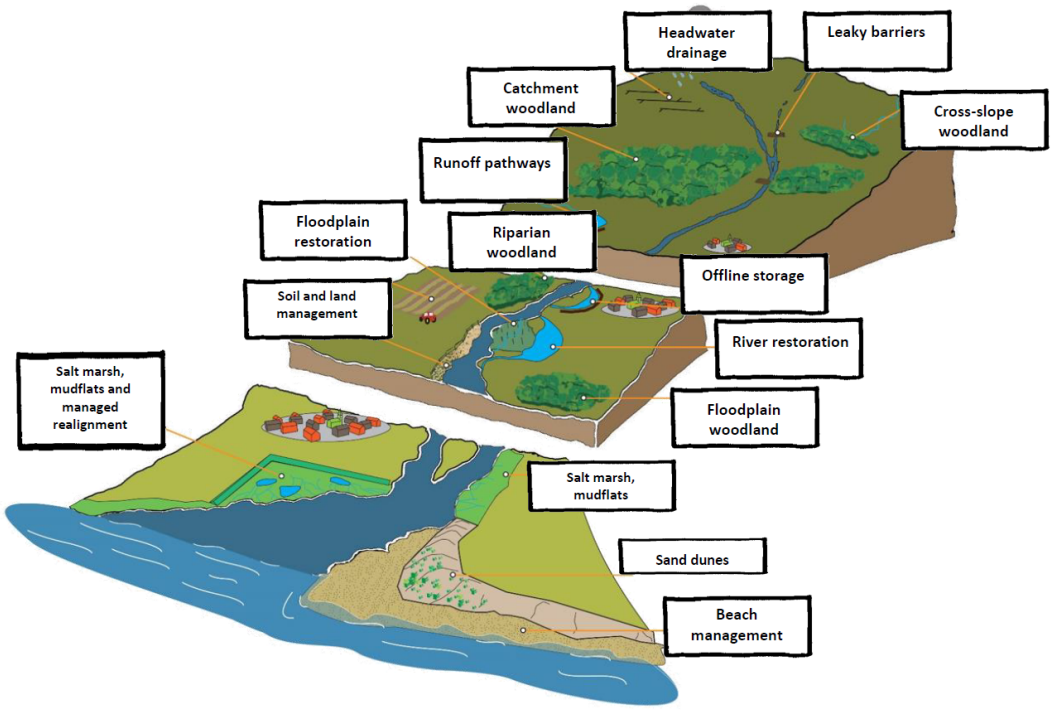


Figure 1.2: Measures which work with natural processes to reduce flood risk, from Burgess-Gamble et al. (2017, p. 3)

Woodland creation includes catchment, floodplain and riparian tree planting. The effectiveness of tree planting as a flood risk management measure has been studied (Alila et al., 2009; Archer, 2007; Carrick et al., 2019; Lacombe et al., 2016; Robinson, 1986; Robinson et al., 2003; Robinson and Dupeyrat, 2005), including in the UK where planting of trees in buffer strips, along with reduced livestock density in Pontbren, Wales, reduced surface runoff by 78% (Marshall et al., 2014; Wynne-Jones, 2016). Although overall, the impact of woodland on downstream flood risk is mixed, particularly in more extreme events (Bradshaw et al., 2007; Soulsby et al., 2017; Stratford et al., 2017) and therefore their effectiveness remains uncertain.

Land management measures aim to improve the hydrological function of the land (Hess et al., 2010) and include measures such as blocking field drains and ditches, improving the water holding capacity of soils (e.g. by increasing soil organic matter, aeration or avoiding compaction) and installing non-floodplain

wetlands. Whilst soil improvements have led to mixed results, modification of headwater ditches and restoration of bare peatland have been shown to reduce storm discharge (Alderson et al., 2019; Grand-Clement et al., 2013; Holden et al., 2006; Shuttleworth et al., 2019).

River and floodplain restoration includes measures such as river re-meandering, river bank restoration and installing instream structures, which increase floodplain connectivity and the time taken for flood water to travel downstream (Gregory et al., 1985; Keys et al., 2018; Sholtes and Doyle, 2011).

Alongside aiming to deliver flood risk management, NFM measures may have multiple benefits ranging from environmental impacts such as water quality improvement, habitat provision and climate regulation to social and cultural impacts such as air quality improvement, health benefits and enhancement of recreational activities (Burgess-Gamble et al., 2017). The range of ecosystem services delivered by NFM measures makes them particularly suitable for implementation under the UK Government's proposed environmental land management scheme (Klaar et al., 2020).

1.3.1 Opportunities for NFM in upland catchments

Although there is no official definition of upland areas (Defra, 2011b) the term refers to areas which are typically over 250–400 m above sea level, and are above the limit of enclosed agriculture (Defra, 2015). Typically low intensity farming, sparse human population, and flood prone communities which do not qualify for traditional flood defence schemes present extensive opportunities for the implementation of NFM schemes in upland catchments. Uplands cover 17% of the total land area of England (2.2 Million ha), which is equivalent to the land area of Wales (Defra, 2011b). Large areas of the uplands consist of estates which are owned by a single landowner; simplifying delivery of large schemes (Wingfield et al., 2019). The impact NFM measures could have on the flood hydrograph in upland catchments is particularly critical because of the opportunities and risks offered by the potential to synchronise or de-synchronise tributary flows (Dixon et al., 2016). Pattison et al. (2014) showed that delaying tributary flows to be asyn-

chronous from the main channel flood peak reduced downstream flood magnitude and could, if designed well, play an important role in flood risk management.

1.4 Large wood in rivers

A widely implemented NFM measure in upland catchments is the use of engineered leaky dams which emulate the hydrological functions of natural large wood in rivers (Abbe et al., 2003; Grabowski et al., 2019; Kail et al., 2007). Engineered leaky dams consist of large wood pieces which are installed across the the channel and extend onto the floodplain. Leaky dams aim to restore the natural functions of large wood in rivers by locally increasing hydraulic roughness (Buffington and Montgomery, 1999; David et al., 2011), hydraulic resistance (in the form of spills) (Curran and Wohl, 2003; Dust and Wohl, 2012; Kitts, 2010) and floodplain connectivity (Keys et al., 2018).

1.4.1 Naturally occurring instream wood

Large wood, defined as having length >1 m and diameter >0.1 m (Piegay and Gurnell, 1997), also referred to as large woody debris, log jams or snags are naturally occurring features of watercourses in forested ecoregions which influence a large number of stream ecosystem processes across a range of spatial scales (Kail et al., 2007; Wohl et al., 2016) (Figure 1.3). Large wood affects a watercourse's geomorphology, hydrology, channel hydraulics and ecosystems (Abbe et al., 2003).

Large wood enters a stream through processes such as bank erosion, fallen trees, and transport of wood from upstream (Gurnell et al., 2002; Keller and Swanson, 1979). It can accumulate to form 'active', 'complete' or 'partial' dams depending on their position in the stream (Gregory et al., 1985). Active dams, also known as 'overflow accumulations' (Wallerstein et al., 1997), form a barrier to flows and sediment movement and create a step in the channel profile. Complete dams form a barrier to flow and sediment movement without the accompanying step in the channel profile. Partial dams do not fully span the channel and are known as 'deflector accumulations', as they deflect flows. Additionally, wood can

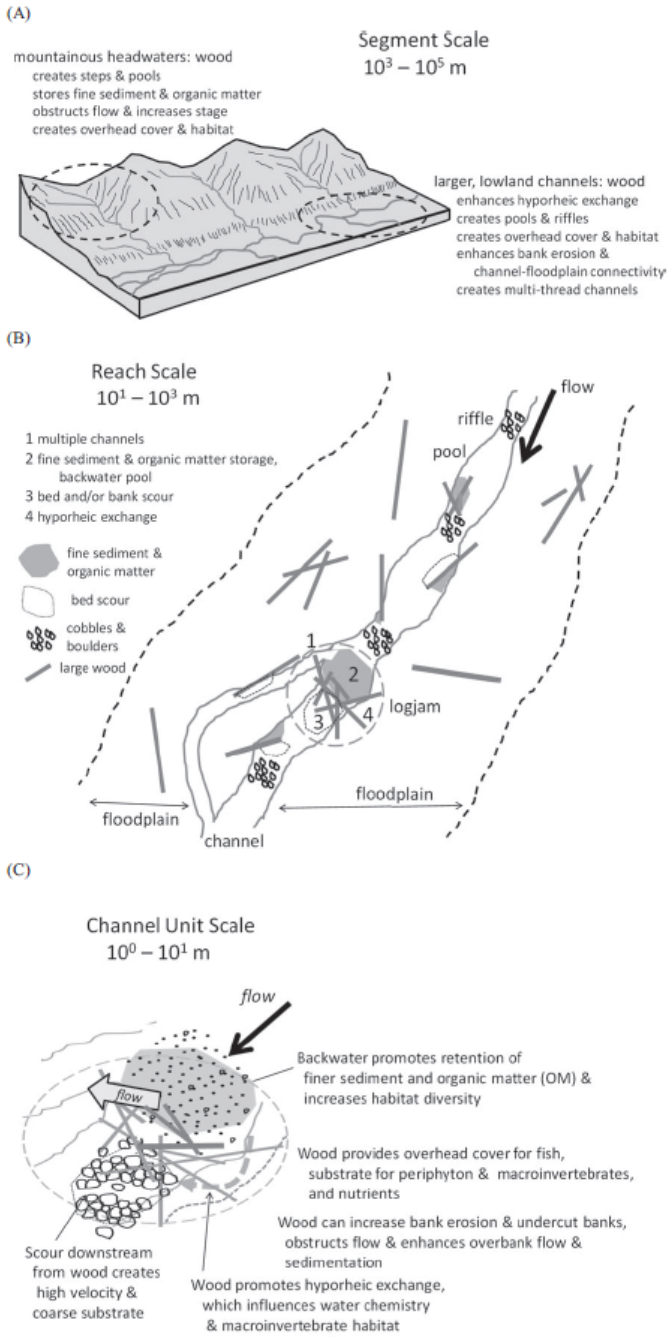


Figure 1.3: Schematic of the benefits of instream wood at a range of spatial scales, from Wohl et al. (2016, p. 318)

form 'underflow accumulations' which span the channel but only affect the flow during high flows (Wallerstein et al., 1997).

Ecological benefits of large wood rely on it to be stable for long enough for its interaction with hydrological processes to alter the local geomorphology (Millington and Sear, 2007). Large wood can leave the system by decay, transport downstream or anthropogenic removal (Gurnell et al., 2002). The majority of large wood in streams is highly mobile (Dixon and Sear, 2014; Gregory et al., 1985; Gurnell et al., 2002), but pieces greater than 2.5 times the channel width have been observed to be functionally immobile (Dixon and Sear, 2014). Although wood accumulations are often unstable, active dams are relatively stable and can form more permanent features of watercourses (Gregory et al., 1985).

Instream wood is known to locally increase hydraulic roughness, decrease flow velocities and increase water levels (Curran and Wohl, 2003; Shields and Gippel, 1995) which has been shown to increase flood wave travel time (Black et al., 2021; Gregory et al., 1985; Kitts, 2010) and floodplain connectivity (Keys et al., 2018; Sear et al., 2010). By locally increasing water levels upstream of instream wood, overbank flows occur during lower discharges and earlier in flood events (Sear et al., 2010). Instream wood increases roughness in the channel by presenting an obstruction to flow which exerts drag forces (Buffington and Montgomery, 1999; David et al., 2011) but the largest increases in hydraulic resistance have been observed for active dams which cause a drop in the channel bed causing a spill (Curran and Wohl, 2003; Kitts, 2010), which effectively acts as a broadcrested weir (Dust and Wohl, 2012).

The influences on the local hydraulics exerted by instream wood can lead to beneficial changes in patterns of deposition and erosion. Instream wood features have long been shown to be linked with the formation of scour pools and bars in the Pacific Northwest and Alaska (Montgomery et al., 2003), and in central European streams (Gregory et al., 1985; Kail, 2003; Piegay and Gurnell, 1997). The backwater effect caused by instream wood slows flows sufficiently to decrease the transport competence of the channel which results in the deposition of sediment in bars (Buffington et al., 2004). Conversely, scour pools are formed where instream wood increases erosion by forcing flows towards the river bed (Beschta,

1983; Hogan, 1986; Robison and Beschta, 1990; Schalko et al., 2019; Wallerstein and Thorne, 2004; Wood-Smith and Buffington, 1996).

Scour pools, riffles and bars increase habitat diversity, providing habitat for both juvenile and adult fish (Linstead, 2001). Instream wood has been observed to increase fish (Montgomery et al., 2003; Peters et al., 1998) and invertebrate density (Coe et al., 2006; Deane et al., 2021) by dramatically increasing organic matter availability (Gurnell et al., 2002; Piegay and Gurnell, 1997; Trotter, 1990).

Overland flows induced by instream wood can scour the river bank (Wohl, 2013), introduce secondary channels (Jeffries et al., 2003; Sear et al., 2010) and deposit fine sediment on the floodplain (Collins and Montgomery, 2002). Such floodplain features can provide highly biodiverse habitats in features such as temporary pools (Davis et al., 2007) and wet woodlands (Braccia and Batzer, 2008). Given these ecological benefits, in Europe, the main aim of the reintroduction of instream wood in river restoration works has been to increase structural complexity of watercourses (Kail et al., 2007).

1.4.2 Reintroducing large wood in rivers

Widespread deforestation and past management practices which recommended the removal of wood from rivers has depleted naturally occurring instream wood in the UK (Linstead and Gurnell, 1999). Traditionally, instream wood was routinely removed from watercourses to allow for navigation, prevent bank erosion, and ensure the hydraulic capacity of channels was maintained (Gippel et al., 1996). However, since the early 1980s research on the wider environmental benefits has changed attitudes towards the management of instream wood (Gippel et al., 1996). The Environment Agency now recommends that woody debris should not be removed from watercourses unless it increases flood risk, increases bank erosion or is a hazard to navigation (Environment Agency, 2010b). In Scotland this is extended to a recommendation that its reintroduction should be considered in watercourses which have been depleted of instream wood (SEPA, 2009). Indeed, instream wood has been reintroduced to restore the ecological and hydromorphological status of rivers in the UK (River Restoration Centre, 2017), Central Europe (Kail et al., 2007) and America (Bernhardt et al., 2005).

Increasingly, instream wood is being installed in UK rivers in the form of engineered 'leaky dams' to restore hydrological processes for the purpose of flood risk management (Dodd et al., 2016; Hester et al., 2016; National Trust, 2015; Nisbet et al., 2015b; Uttley and Skinner, 2017).

Engineered leaky dams consist of wood placed in the river channel and on the river banks to mimic the function of natural accumulations of large wood in rivers. Large wood has been widely used in river restoration to protect banks from erosion, control sediment movement, increase habitat for aquatic species, and accelerate hydrogeomorphological processes (Abbe et al., 2003; Addy and Wilkinson, 2016; Bernhardt et al., 2005; Kail et al., 2007). It is only in recent years that its use as a NFM measures has been trialled (Grabowski et al., 2019) and its design has been adapted for the primary purpose of delaying and reducing the magnitude of flood events.

For river restoration applications large wood is usually installed on the river bed or bank, doesn't span the channel width, and often has the rootwad attached (Abbe and Brooks, 2011) (Figure 1.4). Leaky dams installed for the purpose of NFM, on the other hand, consist of channel-spanning logs which may extend onto the floodplain and are placed perpendicular to the flow, often raised above the channel bed to allow for fish passage (Dodd et al., 2016; Hester et al., 2016; National Trust, 2015; Nisbet et al., 2015b; Uttley and Skinner, 2017) (Figure 1.5). Some leaky dam designs aim to replicate natural wood accumulations whilst others take a more uniform, engineered approach. Research has shown that engineered instream wood features which mimic natural features (e.g. not fixed so that types and sizes of dams can form where they naturally would) have greater environmental benefits than more engineered structures (Kail et al., 2007).

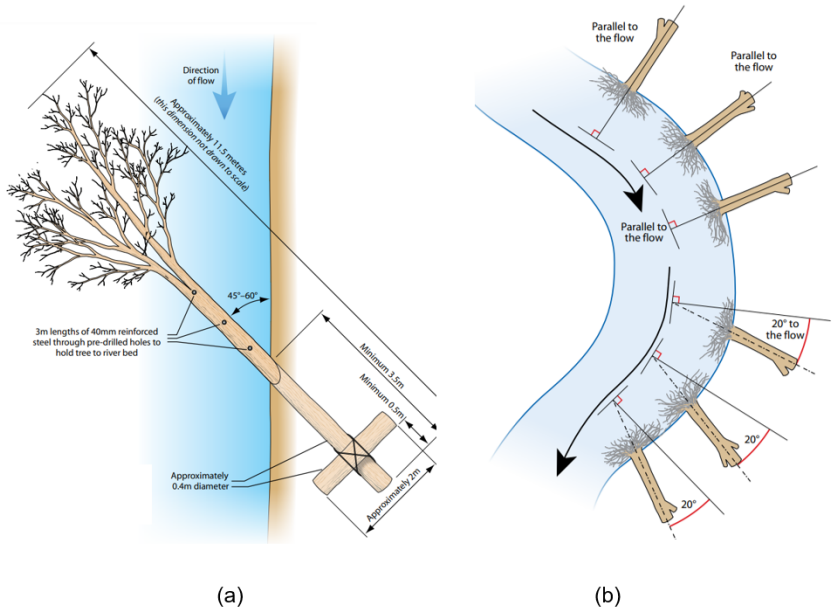


Figure 1.4: Examples of the use of instream wood from the RRC manual of river restoration techniques (a) Fixing whole trees into the river bank for flow diversity on the river Avon, England, from RRC (2009, p.2) (b) Bank protection using root wads on the River Dulais, Wales, from RRC (2013, p.1)

(a) **(b)**

Figure 1.5: Examples of engineered leaky dams (a) A leaky dam in Belford, England which replicates natural wood accumulations, from Quinn et al. (2013, p.12) and (b) A uniform, engineered leaky dam in Pickering, England, from North York Moors National Park (2015)

1.4.3 Design guidelines for leaky dams

Instream wood has primarily been used for habitat creation in river restoration, therefore, the majority of existing guidance focuses on designing instream wood structures to optimise geomorphological and ecological benefits. Guidance for the design, installation and monitoring of instream wood for flood risk management within the UK remains fragmented.

In Scotland the Scottish Environmental Protection Agency (SEPA) provides general guidance in the Natural Flood Management Handbook (Forbes et al., 2015) and specific guidance on leaky dams in the document 'Conceptual Design Guidelines - application of engineered logjams', which draws on experiences from the US (Herrera Environmental Consultants, 2006). Advice is also available in Scotland for the design of instream wood features for the passage of fish (Dodd et al., 2016).

In England and Wales general guidance on how to manage and implement NFM projects is given by the Environment Agency (CBEC Eco Engineering, 2017) but in the absence of specific national NFM leaky dam design guidance various local guidance documents have been produced by practitioners (e.g. Mott, 2006; Woodland Trust, 2016; Yorkshire Dales National Park Authority et al., 2017; Yorkshire Dales Rivers Trust, 2018). These guidelines are generally in agreement that leaky dams should be 1.5 times the channel width in length, be raised approximately 0.3 m above baseflow to allow for passage of fish, be placed perpendicular to the direction of flow, extend onto the floodplain and be anchored to the river banks. The RRC draws on experiences in the UK to present some project specific guidance in the form of case studies in their Manual of River Restoration Techniques (e.g. RRC, 2009).

There is detailed design guidance from locations where the discipline is more mature, such as the United States and Australia. For example, Washington State (USA), provides comprehensive Stream Habitat Restoration Guidelines (Cramer, 2012) which give in-depth information about the design, installation and monitoring of instream wood features. Abbe and Brooks's (2011) chapter 'Geomorphic, engineering, and ecological considerations when using wood in river restoration' provides detailed information for taking an engineering approach to the design

of instream wood features. However, these guidelines aim to provide guidance for the restoration of habitat rather than flood risk management and are based on research carried out in the US, therefore, their applicability to UK upland watercourses may be limited.

The Environment Agency commissioned report of Linstead and Gumell (1999), 'Large Woody Debris in British Headwater Rivers - Physical Habitat Role and Management Guidelines', remains one of few documents which specifically addresses instream wood in upland watercourses and questions the applicability of research from the US to UK headwater streams. However, guidance on natural flood management is due to be published in the UK by the Construction Industry Research and Information Association (CIRIA). The guidance will include information about the design philosophy, objectives and criteria for leaky dams installed for the purpose of flood risk management (CIRIA, 2021).

1.4.4 Hazards of instream wood

It is crucial for the success of leaky dams and river restoration activities that not only the benefits, but also the hazards of introducing large wood to rivers is taken into account (Dixon and Sear, 2014; Wohl et al., 2016). Instream wood can present hazards to people, infrastructure and properties, especially if it is present in or near urban areas (Mazzorana et al., 2011; Ruiz-Villanueva et al., 2014; Schmocker and Weitbrecht, 2013). Wood in rivers present three main hazards: increased water levels during floods, altering erosion and deposition regimes and mobile wood (Wohl et al., 2016). Mobile wood can become trapped behind structures such as bridges and culverts (Fenn et al., 2005; Lagasse et al., 2009) (Figure 1.6), reducing the cross-sectional area of the channel (Ruiz-Villanueva et al., 2014). Blockage of bridges has led to floodplain inundation, bed aggradation, channel avulsion, scouring and collapse of embankments and bridges (Comiti et al., 2006, 2008; Diehl, 1997; Lyn et al., 2007; Mao and Comiti, 2010), and increased upstream flooding (Lagasse et al., 2009; Schmocker and Weitbrecht, 2013).

Even without blocking structures, the impacts of large wood on localised sediment dynamics can induce aggradation or scour, accelerate bank erosion and

Figure 1.6: Woody debris blocking a bridge in Boscastle, Cornwall, during floods in 2004 (Environment Agency, 2007b, p. 1)

encourage lateral movement of the channel (Brummer et al., 2006; Collins et al., 2012; Montgomery and Buffington, 1997; Wohl and Cadol, 2011) causing flooding and presenting a hazard to infrastructure (Wohl et al., 2016). Large wood can also present entrapment hazards to recreational river users such as swimmers or kayakers (Wohl et al., 2016). Furthermore, if wood becomes mobile during high flows it can collide with watercraft, bridges and pipelines (Wohl et al., 2016). The collapse of beaver dams, which consist of accumulations of large wood, have led to outburst floods on numerous occasions which have been linked to the deaths of 13 people and damage to infrastructure in the US and Canada (Butler and Malanson, 2005).

As discussed in Section 1.4.3, there is little guidance about the design and consideration of stability of leaky dams applicable to UK rivers. There is a need to address the potential hazards of introducing large wood to rivers to avoid inadvertently exposing downstream communities at risk (Hankin et al., 2020), and to allow risk management authorities, such as the Environment Agency and lead local flood authorities in England & Wales, to assess the risks as well as the benefits of installing leaky dams (Wohl et al., 2016). Failure to do so can exacerbate negative perceptions of leaky dams, and ultimately rejection of NFM measures by flood prone communities (Gapinski et al., 2021; Waylen et al., 2018).

1.5 Impact of leaky dams on the flood hydrograph

Despite monitoring efforts, the impact of NFM style leaky dams on flood peak magnitude and timing has not been quantified in steep watercourses (Burgess-Gamble et al., 2017; Moors for the Future Partnership, 2015). Previous research has generally focused on using either an empirical approach or a numerical hydraulic or hydrological modelling approach to detect NFM impacts on downstream flood risk, but neither has been able to quantify impacts of leaky dams on flood hydrograph in upland watercourses for a range of flows. Empirical studies have not been able to draw conclusive results because of a lack of comparable events in the before and after monitoring period (Connelly et al., 2020) or high

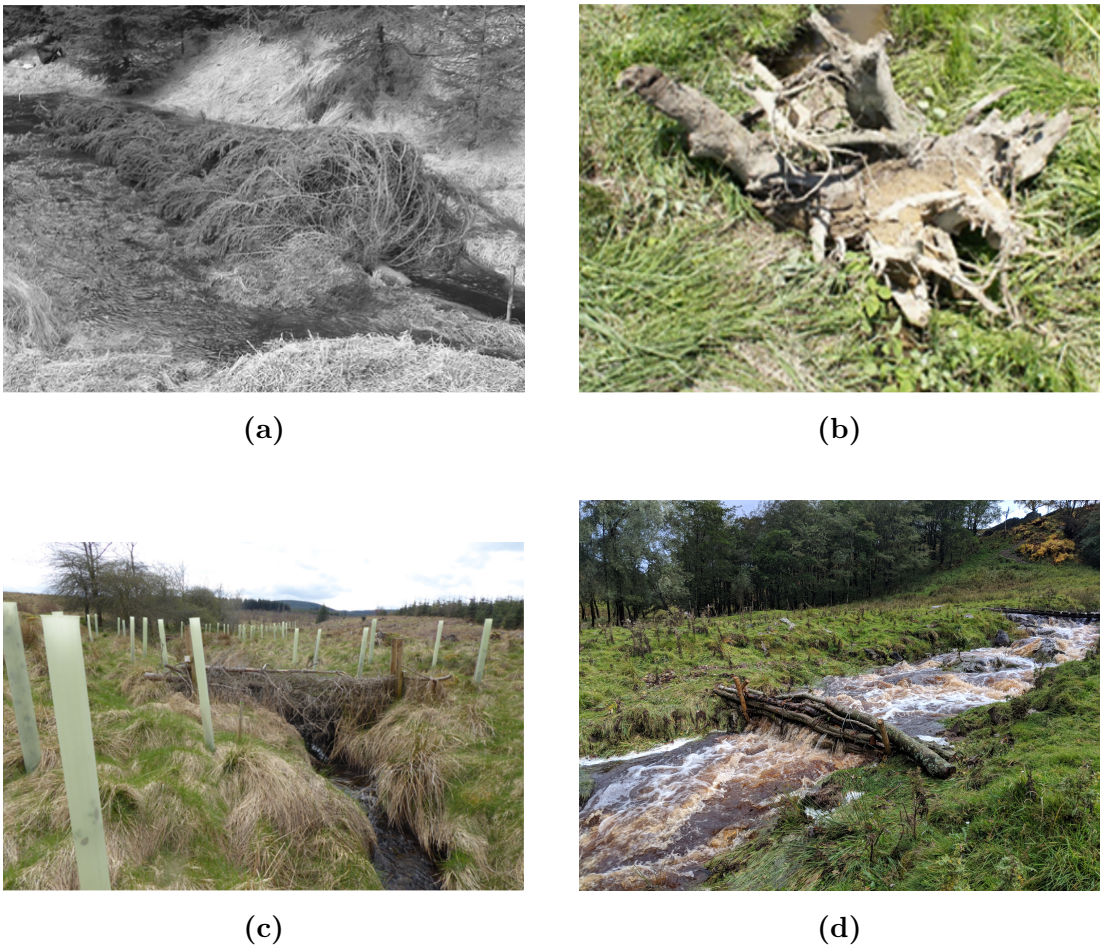
levels of uncertainty which mask the signal of the intervention (Black et al., 2021; Gebrehiwot et al., 2019; Lane, 2017). As a result, there is a lack of data with which to validate the way leaky dams are represented in hydraulic and hydrological models which means that confidence in the outputs of modelling studies remains low (Addy and Wilkinson, 2019).

1.5.1 Evidence from empirical studies

There is little empirical evidence to assess whether leaky dams can affect flood peaks, especially at larger scales (Burgess-Gamble et al., 2017). The stochastic nature of flood events, long timescales associated with monitoring of flood events has led to both limited pre and post intervention data, often producing only a small number of comparable flood hydrographs before and after placement of interventions (National Trust, 2015; Nisbet et al., 2015b; Thomas and Nisbet, 2007). Empirical results are site, scale and event specific compounding the comparison and generalisation of results. Furthermore, monitoring is often carried out on combined NFM measures (e.g. National Trust, 2015; Nisbet et al., 2015b) preventing isolation of the impacts of leaky dams (Connelly et al., 2020).

Two studies in which artificial flood waves were released through controlled reservoir releases have offered some insight into the impacts of instream wood on the flood peak magnitude and timing (Keys et al., 2018; Wenzel et al., 2014). The studies were limited to small streams (bankfull channel width <1 m) and the types of large wood associated with river restoration. The instream wood in both studies was placed on the river bed, rather than above baseflow, were in line rather than perpendicular to the flow, and did not extend onto the floodplain (Figure 1.7).

The placement of nine spruce tops (average length 8 m, average maximum trunk diameter 0.2 m) longitudinally in a 282 m first order, headwater stream reach (gradient 3.7%, width 0.8 m and average flow depth 0.3 m) in the Ore Mountains in south eastern Germany delayed the peak of an artificial 3.5 year return period flood event by 166 seconds, which equates to 10 minutes over a 1 km reach (Wenzel et al., 2014). The flood peak was reduced by only 2.2%, but the 30 m^3 of additional storage altered the shape of the flood hydrograph so that



(a) **(b)**

(c) **(d)**

Figure 1.7: Instream wood researched in (a) the Ore Mountains, Germany, from Wenzel et al. (2014); (b) mid-Atlantic region of the United States, from Keys et al. (2018); (c) Eddleston Water, Sottish Borders (Black et al., 2021), from Interreg North Sea Region Programme (2017) (d) Coverdale, North Yorkshire (from this study)

22% of the flow volume no longer coincided with the most erosive part of the flood peak (Wenzel et al., 2014). Three pieces of large wood with their rootwad facing upstream in a 50 m reach of a headwater stream in the mid-Atlantic region of the United States reduced peak magnitude by 8% for an artificial <1-in-1 year flood event (Keys et al., 2018).

It is known from observations of instream wood in a low gradient, headwater stream in the New Forest, Hampshire that the delaying effect of the large wood on the flood peak declined during more extreme events (Gregory et al., 1985; Kitts, 2010), likely due to overtopping of the wood (Dixon et al., 2016). This effect may also apply to the impact on event peak magnitude (Dadson et al., 2017; Lane, 2017) but as neither reservoir release study assessed instream wood impacts on more than one event magnitude their results are of limited use. Moreover, by testing only one, single peaked event hydrograph the impact of antecedent conditions or different event types (e.g. multi peaked) cannot be assessed. Finally, because the instream wood was not designed as NFM leaky dams the timing and impact on the flood hydrograph is likely to differ and may not be applicable to leaky dams installed for the purpose of NFM.

Recently, progress has been made in quantifying the impact of NFM features, including leaky dams, on event peak timing at a range of spatial scales (Black et al., 2021). In the 69 km^2 Eddleston water catchment in the Scottish Borders of the United Kingdom, a combination of leaky dams, on-line ponds and riparian tree planting increased lag time by up to 7.3 hours for sub-catchments up to 26 km^2 (Black et al., 2021). The study provides much needed empirical evidence to support the results of modelling studies which suggest that NFM measures can have significant impacts on the flood peak at the catchment scale by de-synchronising tributary flows (e.g. Dixon et al., 2016; Pattison et al., 2014). Impacts on flood peak magnitude were not quantified, however, due to significant uncertainty in the flow data (Black et al., 2021).

1.5.2 Evidence from hydraulic and hydrological modelling studies

Models have been used to address key evidence gaps of NFM: impacts at larger spatial scales; impacts on more extreme flood events; and the effects of tributary (de-)synchronisation (Dadson et al., 2017; Lane, 2017). Such studies are also providing insights about strategic placement of leaky dams in the catchment (Hankin et al., 2020).

Although early empirical research showed that the impact of instream wood features was drowned out in large flood events (e.g. Gregory et al., 1985), modelling studies suggest they may be effective in high exceedance probability events; for example, Thomas and Nisbet (2012) found a delay in flood peak of 15 minutes over a 0.5 km reach of a small tributary in Wales in a 1-in-100 year event. Modelling of instream wood in an upland, headwater catchment in the mid-Atlantic region of the United States which was empirically shown to reduce peak magnitude during a <1-in-1 year event, showed that effects could be seen for events with a return period of up to 25 years (Keys et al., 2018). In some cases model results indicate that the effectiveness of leaky dams combined with river restoration measures can increase with event peak magnitude (Odoni and Lane, 2010; Thomas and Nisbet, 2012). This is attributed to 'expandable field storage', or increased use of storage on the floodplain during larger events (Black et al., 2021; Hankin et al., 2020; Kay et al., 2019; Odoni and Lane, 2010; Thomas and Nisbet, 2012).

Coupled hydrological-hydraulic models are being used to gain a better understanding of the flood peak response to instream wood features at the catchment scale. Metcalfe et al. (2017) used a coupled semi-distributed hydrological model with a hydraulic channel model to assess the impact of upstream in-channel measures on the flood hydrograph at a catchment scale (up to 100km^2) to show that 70,000m^3 of storage could reduce the flood peak by 11% in moderate events (Metcalfe et al., 2017). The study found that fewer larger structures lower in the catchment were more effective at reducing and attenuating peak flows than a larger number of small, upland features, because more floodplain storage was

available. However, the study considered only one event which is acknowledged as a limitation (Metcalfe et al., 2017).

At the catchment scale, relative tributary timing is a crucial factor in determining the magnitude of the downstream flood peak (Pattison et al., 2014). Using a reduced complexity hydrological model Dixon et al. (2016) found that instream wood features had a highly variable impact on the downstream flood peak at the catchment scale, depending on whether the subcatchment flows were synchronised or desynchronised as a result of the intervention. Whilst Odoni and Lane (2010) found that interventions were more effective in the upper reaches of the catchment, Dixon et al. (2016) found that interventions in the lower and mid-catchment had the greatest impact on the flood peak.

By developing a system performance model of a catchment with different configurations of between 10 and 20 leaky dams Hankin et al. (2020) showed that the strategic placement of leaky dams in shallower sections of streams enhanced storage capacity and avoided storage provided by dams being underutilised. The model included consideration of failure of leaky dams and concluded that cascade failure of dams could be avoided by placing them on tributaries rather than on the main channel (Hankin et al., 2020).

Although modelling studies provide insights about the functioning of leaky dams at the catchment scale, confidence in modelling results is generally low (Addy and Wilkinson, 2019). Validation data has been available in relatively few studies, which means representation of leaky dams in models is generally heuristic (Addy and Wilkinson, 2019). Misrepresenting leaky dams in models can cause erroneous and misleading results at the local, reach and catchment scale (Addy and Wilkinson, 2019). Yet, to date, there is a lack of guidance on how to represent leaky dams in hydraulic and hydrological models (Addy and Wilkinson, 2019; Burgess-Gamble et al., 2017; Metcalfe et al., 2018; Pinto et al., 2019).

1.6 Monitoring of river restoration projects and natural flood management

The importance of monitoring river restoration projects is widely accepted (Boon, 1998). Monitoring allows the success of a scheme in meeting its objectives to be evaluated; provides evidence for environmental benefits of river restoration; ensures compliance with the WFD and other legislation; and allows changes in the scheme to be monitored over time (Environment Agency, 2007a). Actively monitoring a scheme also reduces overall project risk as unexpected impacts of the scheme can be perceived as they arise (England et al., 2008). This is important in informing funders, other stakeholders and the public of the impacts of the project (Skinner and Bruce-Burgess, 2005).

However, despite the importance of monitoring being emphasised in restoration guidance (e.g. Forbes et al., 2015; Mott, 2011; Woodland Trust, 2016) it is not generally carried out (Hammond et al., 2011). In a study of 50 restoration projects involving wood in Central Europe only 44% of projects were subject to monitoring (excluding photographs and visual inspections) whilst changes in channel morphology and biological quality measures were monitored in only 26% of projects (Kail et al., 2007). Monitoring is often limited by the resources that are available (England et al., 2008). Where resources are limited it is recommended that fewer schemes are monitored well, rather than compromising scientific standards by attempting to monitor all schemes (England et al., 2008). Poorly designed monitoring schemes can waste resources and add little to the scientific knowledge base (England et al., 2008).

In impact assessment an environmental impact is defined as a change in means which can be correlated to the start of some human activity (Smith, 2002). To demonstrate that the restoration has had an environmental impact the monitoring design aims to provide data which allows a change in means to be detected, and attributed to the restoration activity. It is therefore advised to use a Before After Control Impact (BACI) procedure (Underwood, 1994) to monitor the impact of NFM measures on the flood hydrograph (Ellis et al., 2021) which requires the collection of baseline and control site data. Baseline data describes the ecological and physical condition of the site before project implementation; this is

important as it allows the impact of the project to be quantified once it has been implemented (Skinner and Bruce-Burgess, 2005). A control site is a site which is similar to the restoration site but has not been impacted by the restoration activity; this allows the impacts of the restoration to be separated from background changes in the monitored data, such as seasonal trends (England et al., 2008).

In a BACI experiment a relationship is established between the response of a control and impact catchment during baseline monitoring (before impact); after the impact a change in this relationship can be attributed to the impact measure (Clausen and Spooner, 1993). Although the impact and control catchments should be *"comparable in size, topography, vegetation and climate"* (Clausen and Spooner, 1993) this approach does not require the two catchments to be the same, it only requires the response of the catchments to be correlated (Clausen and Spooner, 1993) thereby overcoming many of the inherent difficulties of working in the natural environment. This statistical control means that not all variables which could cause a change need to be monitored and background hydrological and climatic variability can be separated from the impact measure (Clausen and Spooner, 1993; England et al., 2008; Ssegane et al., 2013).

However, designing BACI experiments with flood peak data presents challenges. Alila et al. (2010) argue that the BACI method is flawed because it does not take into account the relationship between flood magnitude and frequency; suggesting a paradigm shift to frequency, rather than chronologically based comparison of events (Alila et al., 2009). Further concerns include the lack of true replication, applicability of results to other watersheds and the impact of extreme climatic events during the impact monitoring period (Ssegane et al., 2013). Temporal autocorrelation of hydrological data can occur where hydrological BACI data has a fine temporal resolution, because measurements are not independent (Som et al., 2012). However, this can be overcome by using appropriate statistical methods (Som et al., 2012). Spatial and temporal replication increase the reliability of the BACI experiment results but is limited by available resources; time and money to design, install and monitor instream wood structures.

1.7 Evidence Gaps

Leaky dams are increasingly being installed with the aim of delaying and reducing the magnitude of flood peaks for the purpose of flood risk management (Grabowski et al., 2019). However, confidence in their efficacy as a flood risk management measure is low because of a lack of empirical evidence (Burgess-Gamble et al., 2017; Ellis et al., 2021). Recently, leaky dams have been shown to delay flood peaks at a range of spatial scales, and for a range of event magnitudes (Black et al., 2021). However, due to high levels of uncertainty in hydrological data, and lack of baseline data combined with the stochastic nature of flood events (Connelly et al., 2020; Ellis et al., 2021; Lane, 2017), their impacts on flood peak magnitude have not yet been empirically quantified for a range of event types and magnitudes, even at the stream scale. To mainstream leaky dams as flood risk management measures empirical evidence for their effectiveness is required for a range of environments, in the same way there is for traditional flood defences (Ellis et al., 2021). Increased confidence in their impacts would allow their benefits to be quantified for implementation in flood risk management schemes, and environmental land management under the current CAP and the proposed ELMS.

Crucially, as well as the benefits, the risks associated with introducing wood to rivers need to be addressed (Dixon and Sear, 2014). Hazards presented by instream wood are well documented, particularly in steep environments (Ruiz-Villanueva et al., 2014; Schmocker and Weitbrecht, 2013; Swanson et al., 2021). However, there is not yet any quantification of the resilience of leaky dams during flood events. Progress has been made in including the potential for leaky dams to fail in models, which has led to insights about strategic placement of leaky dams (Hankin et al., 2020). However, such models are limited by basing their predictions of leaky dam failures on assumptions, rather than quantification of leaky dam fragility (Hankin et al., 2020). Being able to predict the rate of failure of leaky dams, and therefore supply of mobile wood is crucial to the success of leaky dams as flood risk management measures (Dixon and Sear, 2014) and to avoid inadvertently placing downstream communities at increased risk of flooding (Hankin et al., 2020).

1.8 Research questions

The UK's proposed Environmental Land Management Scheme provides an opportunity for the UK to become a world leader in sustainable land management (Klaar et al., 2020). However, to be able to do this, the benefits and trade-offs within and between ecosystem services need to be taken into account (Ellis et al., 2021; Vira and Adams, 2009). For leaky dams, trade-offs exist within flood risk management because of their potential to change the shape of the flood hydrograph (Black et al., 2021; Gregory et al., 1985; Kitts, 2010), and their potential to present a hazard in case of structural failure (Ruiz-Villanueva et al., 2014; Wohl et al., 2016). Because uncertainty in the benefits and risks of installing leaky dams for flood risk management is high (Burgess-Gamble et al., 2017; Ellis et al., 2021; Hankin et al., 2020; Lane, 2017), it is difficult for risk management authorities to assess whether the benefits of installing leaky dams in upland catchments outweigh the risks. Assessing whether the benefits of an intervention outweigh the risks is essential to ensure that communities are not inadvertently being exposed to increased risk (Hankin et al., 2020; Wohl et al., 2016).

This research aims to quantify the impact of installing leaky dams in steep, upland streams on downstream flood risk, both in terms of its impacts on the flood hydrograph, and the probability of failure of engineered leaky dams.

Aim: To quantify the risks and benefits for flood risk management of installing engineered leaky dams in steep, upland streams.

The first part of the aim, quantifying the impact of leaky dams on the flood hydrograph in upland streams, was addressed by setting up an empirical BACI study in an upland, headwater catchment in North Yorkshire, England. An extensive hydrological monitoring network was installed to monitor stage at the upstream and downstream extent of 5 stream reaches for a three year period. Halfway through the monitoring period a series of six to eight leaky dams were built on four of the streams, leaving one stream as a control.

Although notable flood events occurred during the monitoring period, it was not possible to assess the impact the leaky dams had on the flood hydrographs because uncertainty in the data masked the signal of the leaky dams. Particularly,

changes in the stage datum, brought about by erosion or deposition of material in or near the gauging cross-sections, precluded comparison of flood hydrographs recorded before and after the leaky dams were installed.

High levels of uncertainty are common in the type of empirical hydrological data collected in this study (Black et al., 2021; Gebrehiwot et al., 2019; Lane, 2017; Wilkinson et al., 2014). To address this problem, the value of applying a data based time series modelling technique to detect the impact of NFM leaky dams from uncertain empirical data was assessed in this study. Data based time series models are chosen based on the statistical properties of the data (Hipel and McLeod, 1994) and therefore allow some data quality problems to be overcome. Specifically, by applying data transformations the model can be made independent of the stage datum, thereby avoiding problems associated with changes in the stage datum.

Given an upstream flood hydrograph, a validated data based time series model fitted to the baseline data of a stream can be used to make predictions of the downstream flood hydrograph to a known level of accuracy (von Asmuth et al., 2002). Every flood hydrograph recorded after an intervention is made in the stream can therefore be compared to the simulated baseline response of the stream, allowing the impact of the intervention to be detected. Although this approach has been applied to assessing impacts on hydrological processes (e.g. Dickson et al., 2012; Gomi et al., 2006; Watson et al., 2001) it has not yet been used to quantify leaky dam impacts on the flood hydrograph. To assess whether this methodological approach is of value for the detection of the impact of NFM measures such as leaky dams on the flood hydrograph the following research questions, **Q1** and **Q2**, are addressed in Chapters 2 and 3.

To assess whether the benefits associated with installing leaky dams in streams outweigh the risks, a robust, evidence based approach to quantifying their capacity to resist failure is needed. When failure modes are not well understood or consist of complex interactions, empirical fragility analysis can be used to estimate failure probability, conditional on a loading condition, based on observations of leaky dam resilience and failure (Schultz et al., 2010). Fragility analysis is more widely used in earthquake engineering (Porter et al., 2007) but has recently been shown to be valuable in quantifying the failure probability of railway bridge assets

conditional on the flood return period (Lamb et al., 2019). Hence, their applicability to quantifying leaky dam resilience is addressed in research question **Q3**, in Chapter 4. The three research questions addressed in this study are:

Q1: What role could data-based time series modelling techniques play in quantifying NFM impacts from short and uncertain BACI data?

Q2: In upland streams, what is the impact of leaky dams on the flood peak magnitude of a range of flood events?

Q3: What potential does empirical fragility analysis have for quantifying the resilience of engineered leaky dams during extreme flood events?

Chapter 2

Using a data-based modelling approach to assess leaky dam impacts on downstream flood risk *Part I– model fitting and validation*

2.1 Introduction

It is generally accepted that flood risk is increasing in temperate climates due to the effects of climate change (Blöschl et al., 2019; Hirabayashi et al., 2013) and socio-economic drivers such as land-use change (Winsemius et al., 2016). Increased flooding has been recorded across northwest Europe in the last five decades due to higher levels of autumn and winter rainfall (Blöschl et al., 2019). In the UK, this trend is highly likely to continue, with unprecedented levels of rainfall predicted to be recorded every winter (Thompson et al., 2017). In response there has been a shift towards holistic, catchment wide approaches to flood risk management across Europe (Commission of the European Communities, 2009). Working with natural processes (WWNP) to manage flood risk, also referred to as Natural Flood Management (NFM), aims to restore or emulate the natural functioning of river catchments to increase infiltration, slow flows and

store water (Forbes et al., 2015). NFM techniques include woodland creation, improved land management, and river and floodplain restoration (Forbes et al., 2015). It is used alongside traditional flood defences and is particularly suited to sparsely populated rural and upland areas where traditional flood defence schemes are less feasible (Sayers et al., 2002). In the UK, the move towards integrating Natural Flood Management into flood risk management policy has been evident since 2005 (Defra, 2005); yet the efficacy of many NFM measures at reducing downstream flood risk remains unquantified (Burgess-Gamble et al., 2017; Lane, 2017; Wilkinson et al., 2019).

Understanding the level of protection offered by NFM interventions is increasingly important as the UK Government is set to replace the EU's Common Agricultural Policy (CAP) with the Environmental Land Management Scheme (ELMS) in England (Klaar et al., 2020). ELMS aims to fulfil the requirements of the Government's 25 year plan for the environment using the principle of public money for public goods (Defra, 2020). This 'payments for outcomes' approach includes making payments to farmers and land managers for protection from environmental hazards such as flooding (Defra, 2020). However, in the absence of quantitative evidence of the level of protection delivered by interventions, the 'public goods' delivered by an NFM scheme are difficult to quantify. Similarly, England's Environment Agency calls for WWNP to be considered in all publicly funded Flood and Coastal Erosion Risk Management (FCERM) strategies (Environment Agency, 2010a), but the spending of public funds requires a cost benefit approach to be taken (Defra, 2009), which requires the quantification of the performance of NFM measures as flood risk management assets. Furthermore, uncertainty about the efficacy of NFM features presents a barrier to uptake amongst stakeholders such as farmers and land managers whose support is key for the implementation of NFM (Bark et al., 2021).

A lack of sufficient baseline data has been identified as a key barrier to quantifying the hydrological impacts of NFM measures (Connelly et al., 2020; Ellis et al., 2021; Lane, 2017) and validation of the technique as a whole (Dadson et al., 2017). To address this evidence gap a Before-After-Control-Impact (BACI) monitoring methodology, which is common practice in the field of ecology, is recommended (Ellis et al., 2021). The BACI methodology was originally developed for

ecological monitoring (Stewart-Oaten et al., 1986) and extends the comparison of observation before and after an impact to include information from a control site which accounts for natural variation so that causal inference can be made (Smith, 2002). The BACI monitoring methodology has been widely applied in hydrology, for example, to assess the impacts of land use change and stormwater management measures on streamflow (e.g. Alila et al., 2009; Hughes et al., 2020; Rhea et al., 2015; Shuster and Rhea, 2013). However, it has not yet been successfully applied to assessing the impacts of leaky dams on flood peak magnitude in upland watercourses (Burgess-Gamble et al., 2017; Ellis et al., 2021). There are two main challenges with applying the BACI approach to the monitoring of NFM impacts on flood peak magnitude and timing. Firstly, the opportunistic nature of NFM projects means there is a lack of lead time to collect long enough periods of baseline data to account for the stochastic nature of floods (Connelly et al., 2020; Ellis et al., 2021). For example, despite best efforts, a paucity of high flow events observed during the monitoring period hampered the empirical quantification of NFM impacts of two of the three government funded pilot projects initiated in 2009 (National Trust, 2015; Nisbet et al., 2015a) and continues to affect the collection of evidence from the £15 million worth of NFM projects funded by the UK government in 2017 (Environment Agency, 2019). Secondly, even when baseline data is available, high levels of uncertainty typical of hydrological data can mask reach scale impacts of NFM measures (Ellis et al., 2021; Lane, 2017).

As a result of the difficulties associated with collecting hydrological data there are few empirical studies which have successfully quantified the impacts of leaky dams in upland watercourses on downstream flood risk. Recent reviews of both the academic and 'grey' literature identified only two studies in which the impact of instream wood on the flood hydrograph was successfully quantified (Addy and Wilkinson, 2019; Burgess-Gamble et al., 2017). In both of these studies the problems presented by the stochastic nature of flood events were avoided by generating artificial reservoir releases to emulate flood events (Keys et al., 2018; Wenzel et al., 2014).

In the Ore Mountains in Southeastern Germany nine spruce treetops with an average length of 8.5 m and diameter of 0.2 m were placed lengthways in a first order, high gradient stream with average bankfull width of 0.3 m. Flood surges

equivalent to a 3.5 year return period event were released before and after the treetops were placed in the stream and discharge was measured at one second intervals at the upstream and downstream extent of the 282 m test reach. The treetops had a small but significant impact on the flood peak magnitude and timing; flood peak magnitude was reduced by 2.3% and the peak was delayed by 2.8 minutes, although the accuracy of the gauging equipment was not stated. However, 22% of the total flow volume was relocated from the most erosive peak of the event to lower discharges (Wenzel et al., 2014).

A similar experiment was conducted in a 50 m reach of a headwater stream with an average bank full width 0f 0.93 m in the Mid-Atlantic region of the United States. Three pieces of large wood, which spanned the channel width, were placed horizontally on the channel bed with their rootwads facing upstream. Artificial flood waves equivalent to a <1 year return period event were released before and after the large wood was installed and discharge was measured at the upstream and downstream extent of the test reach. The large wood reduced event peak discharge by 8% and increased floodplain inundation extent by 34% (Keys et al., 2018).

Although both of these studies indicate that large wood in rivers can have significant impacts on flood peak magnitude and timing, neither study assessed the types of engineered large wood typically installed for NFM, often referred to as leaky dams. These engineered leaky dams differ from the type of large wood used in the studies of Wenzel et al. (2014) and Keys et al. (2018) in that (i) they were placed perpendicular to the direction of flow, rather than lengthways, to maximise friction effects (Gippel et al., 1996; Shields and Alonso, 2012); (ii) they were raised above baseflow level, rather than placed on the stream bed, so that their storage capacity was not used up during normal flows; and (iii) they extended onto the river banks to increase their storage capacity and interaction with flood flows (Forbes et al., 2015; Herrera Environmental Consultants, 2006; Yorkshire Dales Rivers Trust, 2018). Furthermore, each study empirically assessed impacts of large wood on only one flood event each, limiting understanding of its performance during different magnitudes and types of events.

In the UK, Black et al. (2021) empirically quantified the impact of leaky dams installed for the purpose of NFM on flood peak timing at multiple spatial scales.

The study monitored the 69 km^2 Eddleston catchment (Scottish Borders, UK) in 13 locations for a period of nine years and found significant delays in headwater catchments upto 26 km^2 (Black et al., 2021). However, interventions in the headwater catchments also included on-line ponds and riparian planting and therefore the impacts cannot be attributed to leaky dams alone. Furthermore, due to high levels of uncertainty in the hydrological data the impact on flood peak magnitude was not quantified (Black et al., 2021).

The majority of attempts to overcome the difficulties associated with empirically quantifying leaky dam impacts have been made using numerical fluvial hydraulic and hydrological models (Burgess-Gamble et al., 2017) which rely on *a priori* assumptions about the physical processes governing the impacts of leaky dams. These processes are poorly understood (Dixon et al., 2016; Lane, 2017) and the lack of quantitative validation data means the representation of leaky dams in numerical hydraulic and hydrological models remains heuristic (Addy and Wilkinson, 2019). Critics of this reductionist, or 'bottom up' approach to hydrological modelling have long argued that *a priori* conceptions of how the system works based on academic judgement and intuition can lead to overconfidence in the resulting models (Young, 1978).

An alternative is to follow an inductive, 'top-down' approach, which minimises the need for *a priori* assumptions. This approach, described by Beven (2001) as 'doing hydrology backwards' uses statistical time series modelling methods to infer model structure and parameter values from empirical data (Young, 2003). Once the model structure and parameters have been determined from the data the model represents the dynamical properties of the system and can therefore be used to make predictions of the values of the output series and its uncertainty for unobserved periods (von Asmuth et al., 2002).

By definition, this type of modelling requires the underlying processes which generate the time series to be stationary, or in a state of 'statistical equilibrium' over time. If the statistical properties of the time series changed over time the inferences, forecasts or simulations generated using the fitted model would not be valid unless the underlying non-stationarity of the data was taken into account. There is an array of time series modelling approaches and techniques to account for underlying non-stationarity; artificial neural networks (Dorofki et al., 2012;

Piotrowski and Napiorkowski, 2013; Thirumalaiah and Deo, 1998), support vector machines (Han et al., 2007; Lin et al., 2006), classification and regression trees (Noymanee and Theeramunkong, 2019; Yin et al., 2018) and transfer function models (Beven et al., 2008; Leedal et al., 2010; Romanowicz et al., 2008; Young, 2003) have all been applied to the modelling of hydrological data. The appropriate model is chosen based on the statistical properties of the data (Hipel and McLeod, 1994) to handle features such as seasonality, non-stationarity, and autocorrelation typical of hydrological data (Beven and Westerberg, 2011).

The transfer function noise (TFN) family of models are predominantly used when a time series can be modelled by linearly transforming one or more predictor time series and the resulting residuals of that transformation are autocorrelated. TFN models are therefore particularly well suited to modelling hydrological data (von Asmuth et al., 2002). They are more widely applied in the fields of systems engineering, econometrics and the social sciences (Okiy et al., 2015) but have been used to model hydrological data for several decades (e.g. Dooge, 1959; Jakeman et al., 1990; Young, 1986). In hydrology, TFN models are most commonly used to model the rainfall-runoff relationship (e.g. Katimon et al., 2013; Ratto et al., 2007; Young, 2003), but they have also been used to fill gaps in hydrological records (Tencaliec et al., 2015), real-time level to level forecasting (Leedal et al., 2010; Young, 2002), modelling groundwater fluctuations (von Asmuth et al., 2002) and to detect impacts on hydrological processes (Dickson et al., 2012; Katimon et al., 2013; O'Driscoll et al., 2016). Transfer functions have been shown to produce simulations of peak event magnitude to a high level of accuracy; based on upstream stage series a transfer function model was able to simulate downstream stage on the River Severn to within 0.006 m to 0.139 m (0.1%–3.7%) for varying lead times (between 2 and 14 hours) at the peak of an event (Romanowicz et al., 2008). They have been shown to produce accurate simulations of stage (R^2 = 0.94) even when fitted and validated using only a short period (20 days) of data (Young, 2003).

Transfer function noise modelling and other top-down, data-based time series modelling techniques present an opportunity to extract information from typically short periods of baseline data collected before NFM interventions are installed. This would allow for comparison between pre and post-intervention response of

a stream even if directly comparable flood hydrographs were not captured in both monitoring periods. Given the difficulties in monitoring and processing hydrological data typically collected in NFM projects (Arnott et al., 2018), this research addresses the following research question:

Q: What role could data-based time series modelling techniques play in quantifying NFM impacts from short and uncertain BACI data?

by addressing the objectives:

O1: Assess the potential of linear TFN modelling to present a solution to the problem of quantifying NFM impacts from short and uncertain baseline and post-intervention data.

O2: Assess whether linear TFN models are able to simulate the pre-intervention level to level response of upland streams during flood events to a high enough degree of accuracy to inform the baseline conditions prior to the installation of in-stream NFM interventions.

To answer these questions, TFN modelling techniques were applied to fit and validate models to baseline stage time series collected during a BACI monitoring campaign in three steep, upland streams. The BACI monitoring campaign was designed to assess the impacts of leaky dams on flood peak magnitude and timing in upland streams.

This chapter forms the first part of two linked chapters in which the leaky dam impacts on downstream flood risk were assessed using a data-based time series modelling approach. This chapter (part I) details how the BACI data were collected, fitting of the linear TFN models and the blocked out of sample cross-validation approach used to assess the predictive skill of the models. In the next chapter (part II, Chapter 3) the fitted and validated TFN models were used to simulate the baseline stage response of the streams for all high flow events observed in the post-intervention monitoring period. By taking this approach, differences in the simulated peak stage and observed peak stage could be attributed to the leaky dams.

2.2 Methods

Three years of hydrological data were collected from three streams located in a small, upland catchment in North Yorkshire, UK. The data comprised of 1-minute stage data for a period of 18 months collected on a control stream and prior to the installation of 7–8 leaky dams on two impact streams. Due to high levels of uncertainty and a lack of stationarity in the data, a TFN model was fitted to the data from each stream to simulate the pre-intervention response of the streams. Simulations of pre-intervention response of the streams are used in the next chapter (Chapter 3) to assess the change in peak magnitude of flood events observed after leaky dams were installed in the impact streams. A summary of the model fitting and validation process is given in Figure 2.1. First an overview of the study site and field methods is given, then the steps taken to prepare the data for modelling are outlined, followed by a description of the tests for stationarity. Thereafter a brief summary of the form of transfer function noise models is provided followed by details of the data driven procedures followed to fit a parsimonious model. Finally, the blocked out of sample cross-validation approach used to assess the accuracy of the model and estimate empirical prediction intervals is described.

The first step was to collect stage series upstream and downstream of three streams before and after the installation of leaky dams on the impact streams. The second step was to prepare the data for modelling by performing a thorough quality assurance process which is detailed in Appendix B. The time series modelling approach requires the assumption to be made that the properties of the forecast and predictor series do not vary with time, in other words, the series are required to be stationary. This means that the data do not have predictable patterns in the long term such as trends or seasonality, and that it has constant variance (Hyndman and Athanasopoulos, 2018). Hence, following data collection and quality assurance, in the third step, the non-stationary stage series data was transformed by taking the first order difference to make it stationary. First order differencing had the benefit of making the model independent of errors in the stage datum which contributed high levels of uncertainty in the stage data (see Appendix B).

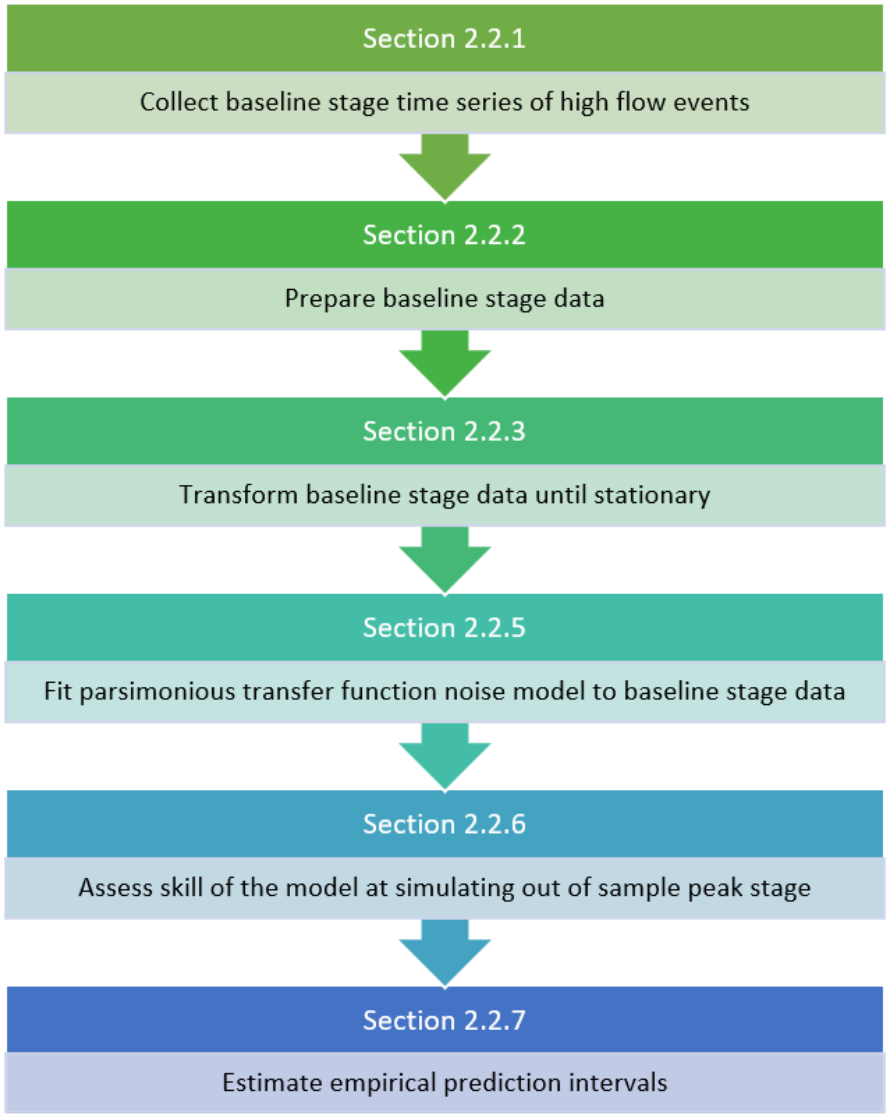


Figure 2.1: Summary of the data-based model fitting and validation process

The fourth step was to fit a time series model to the observed data with downstream stage as the forecast (dependent) variable and upstream stage as the predictor (explanatory) variable. The model was fit to the baseline, pre-intervention data so that, given the upstream stage series of any high flow event, it could simulate the downstream peak stage response of the stream with no leaky dams. To assess the accuracy of the model at simulating downstream peak stage, in the fifth step, the model was validated using events recorded during the baseline monitoring period, i.e., when there were no leaky dams in the stream so that the simulations should match the observations. To do this, each of the events in the baseline data was removed in turn, and the model, which was fitted to the remaining data, was used to simulate its peak stage. The simulated peak stage was compared to the observed peak stage to assess the accuracy of the simulation. By repeating this for all 45 events the model's ability to simulate out of sample peak stage was assessed for every event in the baseline monitoring period. Finally, the distribution of the model error at predicting peak stage was used to estimate empirical prediction intervals for each stream model.

2.2.1 Study site

The study site is located in the headwaters of the River Cover (54.20045 N, -1.98617 E) on the Eastern flank of the Yorkshire Dales National Park, North Yorkshire, England (Figure 2.2). The climate is cool and wet, with an average annual rainfall of 1270 mm (Environment Agency rain gauging station 57426 data 1988-2018, Figure 2.2). The River Cover flows in a north-easterly direction through a number of settlements including Carlton, West Scrafton and Melmerby towards a confluence with the River Ure approximately 20 km downstream of the study site. The Ure passes many larger settlements including Masham, Ripon and Boroughbridge before it flows through the city of York as the River Ouse. The glacial valley of Coverdale overlies Great Scar limestone with rocks of the Yoredale series forming the valley sides (Yorkshire Dales National Park Authority, 2002). The headwaters are formed of many small, parallel streams which flow into the River Cover at an altitude of 400 m AOD. The study focuses on three of these streams, with a combined catchment area of 4.7 km^2. The watercourses are of

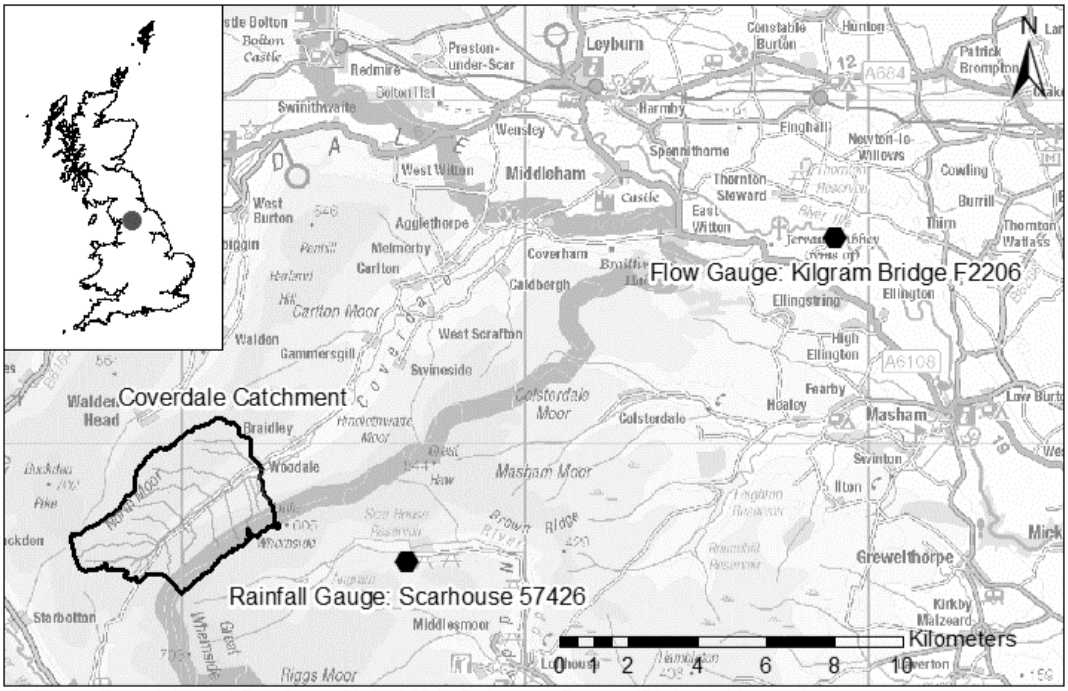


Figure 2.2: Location of study site in Coverdale, North Yorkshire, UK, and nearest Environment Agency operated flow and rainfall gauging stations

type A in the Rosgen classification; steep, partially entrenched and cascading with step-pool streams (Rosgen, 1994). Land use in the catchment is pastoral agriculture on open, unimproved grassland whilst the moorland is managed for grouse shooting. Since 2001 changes in management practice such as grip blocking and reductions in stock grazing have had impacts (anecdotally) on the hydrology and geomorphology of the river; better storage of water in the moorland has decreased large scale bed movements in the river and reduced grazing pressure has increased marginal vegetation which has stabilised the river banks (Wild Trout Trust, 2013). Desk-based studies and walk-over surveys were conducted prior to project initiation to choose streams with similar catchment area, land-use, steepness and typology. The hydrological characteristics of the streams are summarised in Table 2.1. Catchment area was calculated using a 30 m resolution digital elevation map in the global information system ArcMap, version 10.6. and manually adjusted based on elevation data from OS maps. The monitored stream lengths were chosen to avoid including lateral inflows within the monitored reach.

Table 2.1: Stream characteristics

Stream	**Gradient (m/m)**	**Catchment Area (km^2)**	**Monitored length (m)**	**Mean width (m)**
1	0.13	1.1	280	2.6
2 (control)	0.11	1.9	260	3.0
3	0.09	1.7	250	2.7

Water stage data, defined as water level above the gauge datum, were collected following a Before After Control Impact (BACI) methodology as described by Smith (2002). Figure 2.3 shows the control stream (Stream 2) and two impact streams (stream 1 and 3) in which leaky dams were installed at the end of the baseline monitoring period. Baseline data were collected between March 2017 and September 2018 for a period of 18 months before leaky dams were placed in the impact streams. Post-intervention stage data were collected for a further 17 months between September 2018 and February 2020. Stage was monitored at one-minute intervals using In-Situ Inc. (Redditch, UK) Rugged TROLL 100 non-vented pressure transducers ($\pm 0.05\%$ full scale accuracy) in stilling wells at

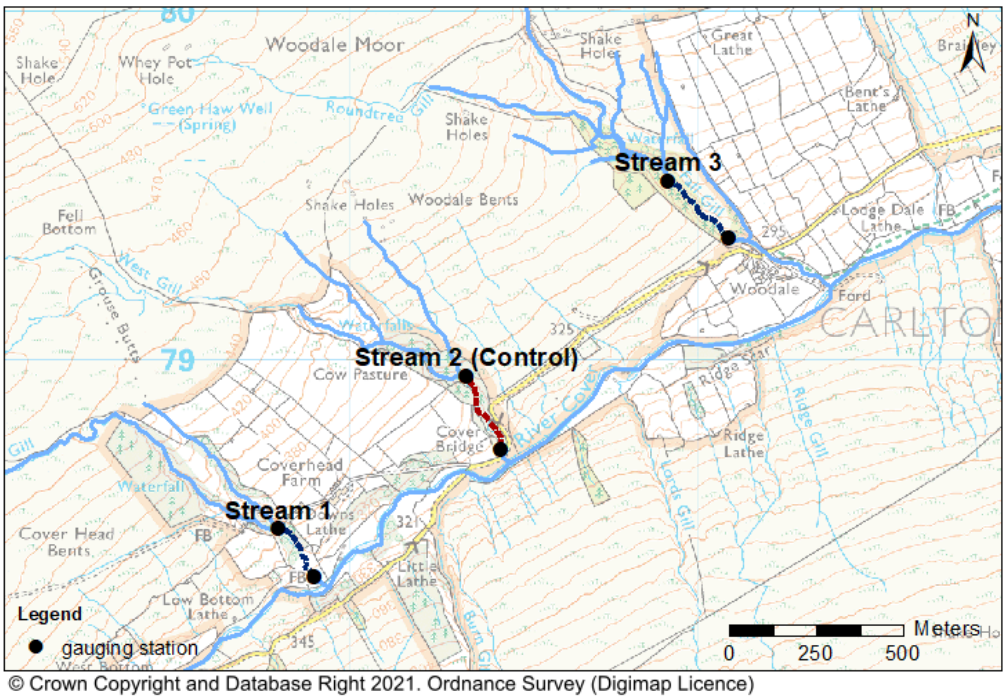


Figure 2.3: Water level monitoring upstream and downstream of two impact streams (navy) and one control stream (red) in the headwaters of the river Cover, Coverdale, North Yorkshire, UK

the upstream and downstream extent of the study reaches on each stream, Figure 2.3. The pressure readings from the transducers were corrected for atmospheric pressure using an In-Situ Inc. (Redditch, UK) Rugged BaroTROLL atmospheric pressure gauge (±0.05% full scale accuracy) which was installed near the bottom of stream 1, at a similar elevation to the non-vented pressure transducers. The stream labelled as the control stream in Figure 2.3 was monitored throughout the baseline and post-intervention monitoring periods without the placement of any interventions.

2.2.2 Data preparation

The data were visualised, prepared and analysed using R version 4.0.2 (R Core Team, 2020) throughout the study. Data preparation consisted of three steps: data smoothing, quality assurance and aggregating the data. The data were smoothed to discern the shape of the hydrograph from the noise in the data caused by fluctuations of up to ± 0.10 m in water level. Due to the large volume of data the computationally efficient locally weighted scatterplot smoothing (LOESS) method was used (Sharma et al., 2015). A second order polynomial LOESS was used to avoid problems of underestimating in regions of curvature (i.e. flood peaks) associated with linear weighted models (Hastie et al., 2017). The package paleoMAS v.2.0-1 (Correa-Metrio et al., 2012) and base R package stats v.4.0.2 were used to smooth the data.

To identify and, where possible, correct for data errors, quality assurance of the data was carried out by visual inspection of the smoothed data, as recommended by Crochemore et al. (2020). After Pastorello et al. (2014), general trends and patterns were identified in the data to detect anomalous values through single variable inspection, multi-variable inspection of correlated variables, and detailed relationship examination. Where errors in the data were identified, evidence was sought from field observations, field notes, photography, spot discharge gaugings and aerial photography to form hypotheses about the source of the errors. Where sufficient evidence was available to support these hypotheses and calculate the magnitude of error a correction was made to the stage series (details given in Appendix B). For example, in one case the collapse of a cattle bridge downstream

of a stage gauge had a backwater effect which was corrected by comparing spot discharge gaugings made before and after the collapse of the bridge. The periods of data error lasted between 1 hour and 11 months (where a permanent change to a cross section occurred) but the majority lasted less than one week. Further details are provided in Appendix B. The data were interactively visualised at different scales and timeframes using the R package Plotly v.4.9.2.1 (Sievert, 2020). Finally, to reduce the computational burden and complexity of the model the data were aggregated from 1-minute to 15-minute time-steps by taking the maximum value of the cleaned and smoothed data in each 15-minute period. The maximum rather than mean values were taken to preserve the maximum event peak magnitudes in the stage series.

2.2.3 Data transformation

To fit a TFN model to the stage time series data the data had to meet the requirements of stationarity (Hipel and McLeod, 1994). The condition of stationarity in the upstream and downstream stage time series data for each stream was tested using a unit root test, the Augmented Dickey Fuller (ADF) test (Fuller, 1996) and a stationarity test, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992). The presence of a unit root in the data indicates the presence of a stochastic trend. The KPSS test was used to test the null-hypothesis that the data series was stationary against the alternative of a unit-root; rejecting the null-hypothesis indicated that the data series was non-stationary. The null-hypothesis was rejected if the test statistic was greater than the critical value for the 95% confidence level (Kwiatkowski et al., 1992).

As the KPSS test has a high rate of type I errors (rejecting the null-hypothesis erroneously) (Caner and Kilian, 2001) it was combined with the ADF test for stationarity. Conversely to the KPSS test the ADF test was used to test for the presence of a unit root. Rejecting the hypothesis that a unit root was present in the data indicated that the data would be unlikely if there was a unit root (non-stationarity). The ADF test required the selection of the appropriate lag length to avoid biasing the results or reducing the power of the test. The lag length, l_{max}, (in number of timesteps) was chosen following the rule in Equation

2.1 suggested by (Schwert, 2002), where T is the sample size and square brackets indicate the integer value. Following Ng and Perron (1995) the ADF statistic was calculated for every lag up to l_{max} and the p value was reported for the biggest lag length at which the absolute value of the statistic was <1.6, or the smallest value of the statistic where all values >1.6 (after Ng and Perron (1995).

$$l_{max} = [12.(\frac{T}{100})^{\frac{1}{4}}] \tag{2.1}$$

Both the upstream stage series, U, and downstream stage series, D, in metres, from all streams were transformed to satisfy the requirement of stationarity by taking the first order difference after Box and Jenkins (1976) as shown in equations 2.2 and 2.3, where the subscript *'t'* denotes the time-step, *'t-1'* denotes the previous 15 minute timestep and the integrated upstream and downstream stage series are denoted $U*$ and $D*$ respectively (in metres). The KPSS and ADF unit root tests were performed on the differenced upstream and downstream stage series, $U*$ and $D*$, which verified that they were stationary as a result of the data transformation (Table 2.2).

$$U_t^* = U_t - U_{t-1} \tag{2.2}$$

$$D_t^* = D_t - D_{t-1} \tag{2.3}$$

Table 2.2: Results of the unit root tests for stationarity. ** indicates significance to 1% level, * indicates significance to 5% level

		Upstream (predictor variable)		Downstream (forecast variable)	
		KPSS test	ADF test	KPSS test	ADF test
Stream 1	Stage	39.6663**	-2.03*	35.6453**	-1.23
Stream 2	Stage	9.4579**	-1.75	9.0565**	-1.43
Stream 3	Stage	39.6663**	-3.51*	35.6453**	-1.61
Stream 1	Differenced Stage	0.0089	-33.1**	0.0116	-32.9.1**
Stream 2	Differenced Stage	0.0032	-44.9**	0.0026	-45.0**
Stream 3	Differenced Stage	0.0089	-41.3**	0.0116	-40.9**

2.2.4 Linear Transfer Function Noise models

Following confirmation of stationarity, transfer function noise models were identified as an appropriate class of models to represent the relationship between U^* and D^*. Transfer function noise models with application to hydrological data are described in detail by Hipel and McLeod (1994) and are therefore only summarised here. Linear transfer functions mathematically describe the dynamic linear relationship between a given input and output. The transfer function, or dynamic component of the model, is similar to a multiple linear regression, but one in which the predictive variables can include one or more lagged versions of a variable. Autocorrelation in the residual series, which would not meet the assumptions of independence required in linear regression, is taken into account by fitting a time series model to the residual series in the noise component of the model. Hence, the transfer function model consisted of a dynamic component and a noise component:

$$output = dynamic\ component + noise \tag{2.4}$$

The dynamic component of the TFN model was a linear function in which the forecast variable, D^*, was regressed against the predictive variable, U^* and k number of lagged values of U^*; thereby representing how the input, (U^*_t), dynamically affects the output, (D^*_t):

$$D^*_t = U^*_t + v_1 U^*_{t-1} + v_2 U^*_{t-2} + ... + v_k U^*_{t-k} \tag{2.5}$$

The parsimonious number of dynamic regression terms (lagged values of U^*), k, required to simulate D^* was determined by inference from the data as described in Section 2.2.5. Using the backshift operator *'B'*, a notational device which denotes the number of periods by which the data is shifted back, or lagged, this can be written as:

$$D^*_t = v(B)U^*_t \tag{2.6}$$

where $v(B) = (v_0 + v_1 B + v_2 B^2 + ... + v_k B^k)$ is the transfer function, the coefficients v_0, v_1, v_2, v_k, are the impulse response weights, and k is the maximum number of lags used. As for the number of lagged terms, k, the impulse response weights

were inferred from the data as described in Section 2.2.5. Influences other than the input series U_t which affect the output, D_t, were captured in the noise term, N_t, hence, in its simplest form, for series with zero mean, the transfer function noise model is:

$$D_t^* = v(B)U_t^* + N_t \tag{2.7}$$

the noise term, N_t, was represented by a time series model. As the model error of hydrological data is often auto-correlated (Hipel and McLeod, 1994) it is usually represented using an autoregressive moving average (ARMA) model (Bell et al., 2001; Katimon et al., 2013; Yuan et al., 2009). For the stationary noise series the ARMA model was of the form:

$$N_t = \sum_{j=1}^{p} \phi_j N_{t-j} + \sum_{j=1}^{q} \theta_j a_{t-j} + a_t \tag{2.8}$$
$$\text{where } a_t \sim N(0, \sigma^2)$$

where $\phi_j = (\phi_1, \phi_2, ..., \phi_p)$ and $\theta_j = (\theta_1, \theta_2, ..., \theta_q)$ are the vectors of model coefficients of order p and q, which denote the number of autoregressive and moving average parameters respectively, 'j' is the autoregressive or moving average term between one and p or q respectively, and a_t is the residual series. To obtain estimates for the model parameters, a_t was required to be independent and normally distributed with zero mean and fixed variance, which is denoted $N(0, \sigma^2)$ in time series modelling convention. Using the backshift operator, B, Equation 2.8 can be denoted:

$$N_t = \frac{\theta(B)}{\phi(B)} a_t \tag{2.9}$$

where $\theta(B)$ is the moving average (MA) term of order q, $\phi(B)$ is the autoregressive (AR) term of order p. The number of terms, p and q, in the ARMA model were determined by inference from the data as described in Section 2.2.5. The transfer function and noise terms were combined to simultaneously model the dynamic and noise components of the transfer function noise model with zero mean:

$$D_t^* = v(B)U_t^* + \frac{\theta(B)}{\phi(B)} a_t \tag{2.10}$$

2.2.5 Model Fitting

To fit the TFN model in Equation 2.10 the Minimum Akaike Information Criterion Estimation (MAICE) procedure introduced by Akaike (1973) was followed. The MAICE procedure is an adaptation of Box and Jenkin's model fitting procedure (Box and Jenkins, 1976) which uses the Akaike Information Criterion (AIC) as an indicator of predictive ability to discriminate between models. The AIC is defined in Equation 2.11, where $lnML$ is the value of the maximised log likelihood function of the model and k is the number of model parameters. The lowest AIC score was used to implement the principles of model parsimony by penalising the statistical fit measure ($lnML$) for the number of model parameters, k, used to obtain the fit. Hurvich and Tsai (1989) have shown the application of AIC score to the fitting of ARMA models.

$$AIC = -2lnML + 2k \tag{2.11}$$

Fitting a parsimonious model, obtaining a good statistical fit to the data with the fewest number of parameters possible (Checkland, 1981), was relevant as the forecasting skill of the data driven model would be reduced if too much or too little of the complexity of the data was captured by the model (Astrup et al., 2008; Bailey et al., 2015; Wagener et al., 2001). The AIC score was used to determine which model parameters were useful in increasing the model's predictive ability in favour of approaches which use the statistical significance of parameter estimates because p-values are not a good indicator of predictive ability for time series models (Hyndman and Athanasopoulos, 2018). Furthermore, as p-values can be misleading when predictors are correlated, they could not be used to determine statistical significance of the dynamic regression parameters (Hyndman and Athanasopoulos, 2018).

The two parts of the transfer function noise model were fitted simultaneously. The dynamic linear regression component of the model, given by Equation 2.5 and the ARMA noise model given by Equation 2.8 were fitted by maximum likelihood estimation (MLE) using the package forecast in R (Hyndman and Khandakar, 2008).

Specifying range of parameters

To fit Equation 2.10 to the data, the number of lagged regressors, k, in the transfer function and the order, p and q of the ARMA noise were determined by inference from the data. The cross correlation function was used to assess the correlation between upstream and downstream stage. The cross-correlation function plot between U^* and D^* had its highest value at lag zero, the value decreased gradually to near zero by a lag of 20 timesteps (five hours) but remained significant and positive, for higher lags it remained significant but oscillated around zero taking both positive and negative values. Based on the form of the cross-correlation function between U^* and D^* the dynamic regression terms considered were lag 0 to 12, lag 20, 30, 40 and 50. Lag 0 was the instantaneous value of stage, lag 1 was the stage in the previous 15-minute time step, lag 2 was the stage two timesteps previously (30 minutes) and so on to lag 50, the stage 12.5 hours previously. Higher order lags were considered in case consideration of antecedent conditions increased the predictive ability of the model. In the interest of model parsimony, and interpretability the orders p and q of the ARMA noise model were restricted to five. Models with both autoregressive (AR) and moving average (MA) terms were considered because mixed models often require less parameters in total than pure AR or MA models (Pollock, 1992).

Estimating parameters and AIC for each model

In the first stage of model fitting the number of dynamic regressors, k, was determined using a forward stepwise approach after Hyndman and Athanasopoulos (2018). It is acknowledged that a stepwise regression does not always result in the best model fit, but it has been shown to almost always result in a good model fit (James et al., 2014). At each step the full TFN model was estimated so that the assumptions underlying the parameter estimation would not be broken. Whilst performing the forward stepwise regression it was assumed that the noise series was appropriately represented by an ARMA noise model obtained using an automated AIC optimisation algorithm (Hyndman and Khandakar, 2008). The AIC score was calculated for the best model at each step of the forward stepwise

regression and plotted against the number of regressors, k, to identify the points at which additional regressors had little impact on the AIC score.

Additionally, simple graphical methods of model identification of Box and Jenkins (1976) were used to check whether the data exhibited a pure autoregressive or moving average signature. Namely, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the noise series were plotted along with the approximate 95% confidence intervals for a white noise series, $\pm 1.96\sqrt{T}$, where T is the length of the series, to check whether the values of the ACF or PACF were significantly different from zero. An autoregressive (AR) signature was recognised by a cut-off in the PACF followed by no significant values of the PACF with a gradually dampening ACF plot. A moving average (MA) signature was signified by a gradually dampening PACF and a cut-off in the ACF with no further significant values of the PACF at higher lags. A model with mixed AR and MA terms was recognised by the absence of a clear AR or MA signature in the data (Hipel and McLeod, 1994; Pollock, 1992). To fit a mixed model the automated algorithm developed by Hyndman and Khandakar (2008) was used to test all possible combinations of AR and MA orders to optimise the AIC score. After Hipel and McLeod (1994) the model was checked for overfitting by comparing the optimised AIC score to that of the model with one less AR term and one less MA term in turn to check whether the model complexity could be reduced without significantly impacting the model fit. Underfitting was checked by comparing the optimised AIC score to that of the model with one more AR, one more MA term and one more of each of the terms in turn to check whether the AIC score was improved.

The algorithm simultaneously fitted the dynamic regression component of the model, with number of lagged predictors, k (determined in the stepwise forward regression) and carried out a search over the ARMA model order space to identify the optimum combination of terms by minimising the Akaike Information Criterion (AIC) score. The ARMA noise model was fit by maximum likelihood estimation (MLE). The Forecast v.8.12 package (Hyndman et al., 2020) was used to fit the models.

Residual diagnostic checks

Before attaching confidence to the results, the statistical assumptions of the model were validated. Namely, the residuals were checked for autocorrelation, heteroscedasticity, and non-normal distribution. Residuals were checked for autocorrelation by studying the ACF plots and carrying out the Breusch-Godfrey test for serial correlation (Godfrey, 1978). Unlike other tests for serial correlation such as the Ljung-Box test and Durbin Watson test, the Breusch-Godfrey test allows for lagged dependent variables and could therefore be used on the residuals of the autoregressive model (Maddala and Lahiri, 2009). Normality and heteroscedasticity of the residuals were checked using standard residual diagnostic plots. Non-normality of the residuals would not invalidate the model simulations, but would affect the validity of the theoretical prediction intervals (Hyndman and Athanasopoulos, 2018).

2.2.6 Model validation

A blocked cross validation procedure was used to assess whether the models were suitable for the intended purpose of simulating the pre-intervention downstream stage response to high flow events. Similar to models used for forecasting into the future, the model required good predictive skill in simulating downstream stage response to flood events which were not used to train the model. This is known as 'out of sample' testing and is preferred to goodness of fit measures based on 'in sample' data used to train the model (Bartolomei and Sweet, 1989; Pant and Starbuck, 1990; Tashman, 2000) which have been shown to underestimate the forecasting error (Makridakis et al., 1982). Out of sample data for validation is commonly obtained by retaining a portion of the collected time series data from the end of the monitoring period (Tashman, 2000). However, this approach has been criticised for not making full use of the available data for model training, particularly for relatively short datasets (Bergmeir and Benítez, 2012). For this reason cross-validation is commonly used to evaluate the performance of machine learning and other regression models (Bergmeir and Benítez, 2012). Similar to other statistical methods, cross-validation methods can be adapted for application

to time series data by carrying out a blocked form of cross-validation (Bergmeir and Benítez, 2012; Roberts et al., 2017)

Based on the recommendations of Bergmeir and Benítez (2012) & Roberts et al. (2017) multiple out of sample time series were simulated by removing one ten-hour block of data from the series at a time and testing the model trained on the remaining data on the excluded block. The block length of ten hours was chosen because the simulation error for every timestep is additive and is therefore reduced by keeping the length of the simulation to a minimum (Hyndman and Athanasopoulos, 2018; Tashman, 2000). The blocks were centred on the peak of discrete high flow events so that the skill of the model at predicting the out of sample event peak magnitude could be assessed using the goodness of fit measures described in this section. Every high flow event in the baseline data was included in the removed data block in turn so that an out of sample prediction of the event peak stage was available for each event whilst using the most complete dataset for training the model as possible.

Identifying discrete high flow events

A consistent, objective rule-based method was used after Deasy et al. (2009); Glendell et al. (2014) to identify discrete high flow events in the stage series by requiring the following two criteria to be met: (1) A stage peak was considered a discrete high flow event if it was part of a defined flow event with duration >60 minutes and the upstream peak stage exceeded the mean stage recorded on the stream; (2) Events were classed as independent if they were separated by at least 15 minutes of stage below or within 10% of baseflow stage. Following the approach of Bezak et al. (2015) a consistent estimate of the baseflow stage series was obtained using the methods described in the World Meteorological Organisation's manual on low-flow estimation and prediction (WMO, 2009). The method identifies turning points based on minima found in five day time windows of daily time-series and was successfully used by Bezak et al. (2015) to separate over 2,500 events from daily streamflow records. The accompanying R package 'lfstat' v. 0.9.4 (Koffler et al., 2016) was used to implement the method. To account for the flashy nature of the streams a three-day time window and turning

point factor of 0.95 was used. Finally, a visual inspection of the time series data was used to check that all events were extracted from the data and that the identified events were independent. The Hydrological Model Assessment and Development (HydroMAD) v.0.9-26 R package (Andrews and Guillaume, 2018) was used to identify discrete storm events using the above criteria.

Goodness of fit measures

Goodness of fit measures were calculated for each of the out of sample event simulations. Although there has been much discussion of appropriate choice of forecast error statistic for the comparison of models (e.g. Armstrong and Collopy, 1992; Tashman, 2000) this study was concerned primarily with the goodness of fit of the simulated event peak magnitude. Therefore, following similar studies (e.g. Katimon et al., 2013; Padiyedath Gopalan et al., 2019; Romanowicz et al., 2008) the measures used to assess model performance were absolute peak error (PE), peak error percentage (PEP), peak timing error (PTE), root mean square error (RMSE) and Nash-Sutcliffe Efficiency (NSE) given by Equations 2.12 to 2.16, where $\hat{D}_{peak}$ is the model simulated peak magnitude, D_{peak} is the observed peak magnitude, $\hat{D}$ is the model predicted stage, D is the observed stage and N is the number of observations. $\hat{T}_{peak}$ and T_{peak} are the simulated and observed event peak timing respectively. PE, PEP and PTE describe the ability of the model to predict the event peak magnitude, whilst the RMSE and NSE describe the goodness of fit of the simulation to the observed data throughout the event. Although the NSE requires the assumption that the errors are independent to be met it is used here because of its wide application to assess the goodness of fit of hydrological models (McCuen et al., 2006).

$$PE = \hat{D}_{peak} - D_{peak} \tag{2.12}$$

$$PEP = \frac{(\hat{D}_{peak} - D_{peak})}{D_{peak}} \tag{2.13}$$

$$PTE = \hat{T}_{peak} - T_{peak} \tag{2.14}$$

$$RMSE = \sqrt{\sum_{i=l}^{N} \frac{(\hat{D} - D)^2}{N}} \tag{2.15}$$

$$NSE = 1 - \frac{\sum^{N}(\hat{D} - D)^2}{\sum(D - \bar{D})^2} \tag{2.16}$$

Skill of the model at predicting event peak magnitude

To assess the model's overall performance at predicting event peak magnitude for the range of events tested the observed peak magnitudes were linearly regressed against the simulated peak magnitudes and the confidence interval of the relationship was calculated based on the standard errors of the linear relationship. The confidence interval describes the uncertainty associated with a population parameter such as the mean or a regression coefficient (Hahn and Meeker, 1991). It was therefore expected to be narrower than the prediction intervals, which quantify the uncertainty associated with the prediction of an individual data point. The NSE_p (Equation 2.17), which has been widely used to assess goodness of fit of hydrological models (McCuen et al., 2006), was used to test the closeness of the relationship to the one-to-one line. In Equation 2.17 n is the total number of events and $\bar{D}_{peak}$ is the mean of the observed values of peak magnitude.

$$NSE_p = 1 - \frac{\sum^{N}(\hat{D}_{peak} - D_{peak})^2}{\sum(D_{peak} - \bar{D}_{peak})^2} \tag{2.17}$$

2.2.7 Empirical prediction intervals

Theoretical prediction intervals are calculated based on the standard error of the innovation series and the residuals of the fitted model (Chatfield, 2001; Lee and Scholtes, 2014) under the assumption that the model is correctly specified, residual errors follow a normal distribution and that they are independent and identically distributed (Hyndman and Athanasopoulos, 2018). However, although they are commonly used, it is widely accepted that theoretical prediction intervals are almost always too narrow in practice because they account only for the uncertainty due to random error (Hyndman et al., 2002; Makridakis and Winkler, 1989) and may not provide adequate cover if the assumptions of normal,

independent and identically distributed residuals are not strictly met (Hyndman and Athanasopoulos, 2018). Therefore, a common alternative approach was used to calculate prediction intervals based on the empirical out of sample forecasts which account for random, parameter and model specification errors whilst only assuming that the error distribution of future simulations is similar to the error distribution of the out of sample simulations (Lee and Scholtes, 2014; Williams and Goodman, 1971). Empirical prediction intervals have been successfully applied in a wide range of fields (Isengildina-Massa et al., 2011; Lee and Scholtes, 2014; Rayer et al., 2009). After Williams and Goodman (1971), who introduced the approach, the prediction interval at each forecasting timestep was estimated using specified quantiles of the empirical error distribution at that timestep. As the error for multiple step ahead simulations was additive (Hyndman and Athanasopoulos, 2018) the simulation window was centred on the peak of the event so that the event peak estimation error was always calculated for the same timestep ($\frac{N}{2}$), where N was the number of simulation timesteps. The 95% prediction interval, and the 80% prediction interval were calculated in this way. Although the 95% prediction interval is more stringent, the 80% prediction interval has been recommended for use with error distributions with outliers, or 'tail problems' (Chatfield, 2001).

2.3 Results

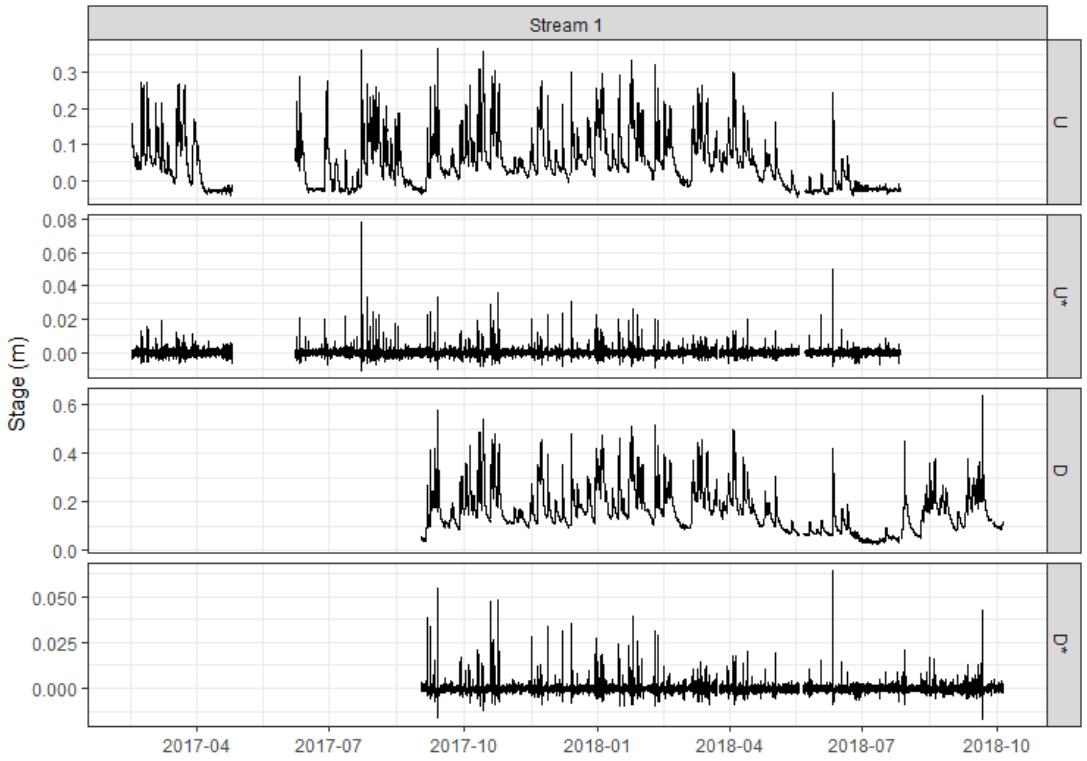
The stage data collected in three small, upland streams were prone to datum errors which made the data too uncertain to assess leaky dam impacts directly (Appendix B). However, by transforming the data to meet the requirements of stationarity and fitting a time series model the data were used to simulate the baseline response of the streams. For one of the impact streams the model overestimated peak stage, particularly for the largest events, but on the other impact stream the model was able to accurately replicate event peak stage across the range of events observed during the baseline monitoring period. On the control stream the simulations were similarly accurate for all but the six largest observed events. Empirical prediction intervals obtained from the model simulations indicate the accuracy with which the models will be able to predict what downstream

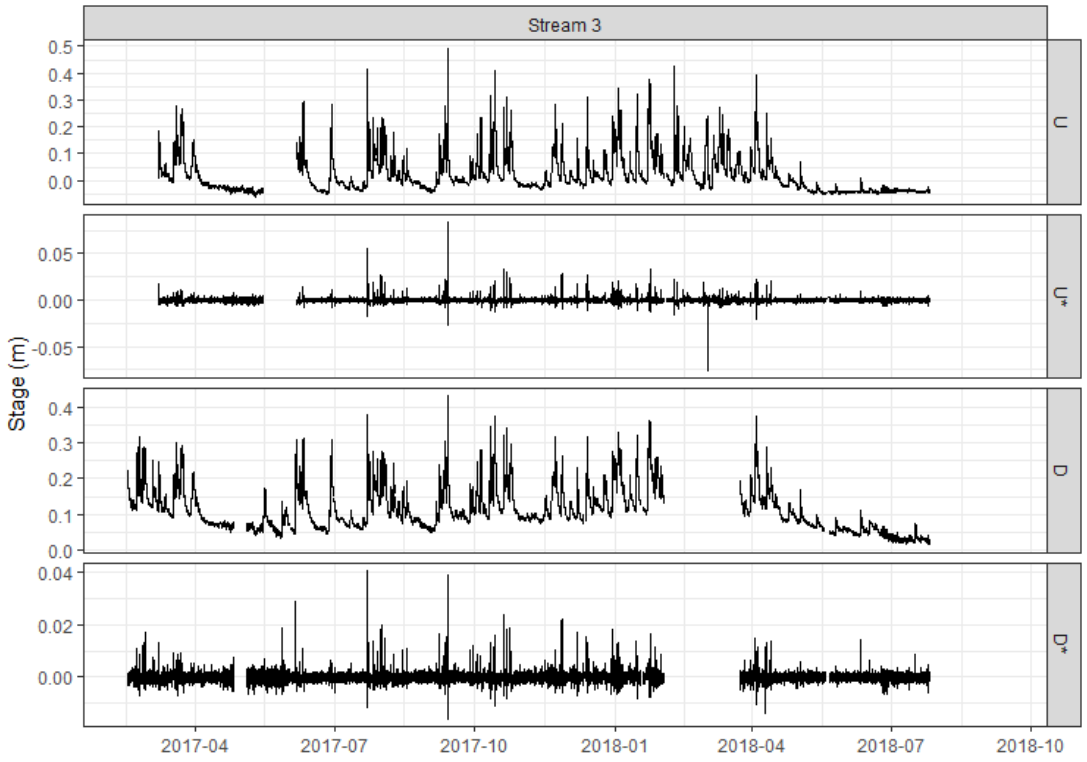
event peak magnitude would have been had no leaky dams been installed in the post-intervention monitoring period.

2.3.1 Baseline model training data

The upstream and downstream baseline stage series collected on the two impact streams and the control stream are shown in Figure 2.4, along with their first order differenced series which form the predictor and forecast series of the model. The range of stage collected during the baseline monitoring period was between 0.45 m and 0.79 m and varied between gauging stations depending on the shape of the gauging cross section (Figure 2.4). Negative values of stage resulted from the correction of the non-vented pressure transducer data with atmospheric pressure recorded closer to the altitude of the downstream gauges than the upstream gauges. The downstream stage was highly cross-correlated to upstream stage at lag zero (i.e. instantaneously) (CCF 0.97-0.99) on all three streams reflecting the steep nature of the catchment. Seasonality was present in all three streams with periods of lower baseflows in the summer months, particularly in summer of 2018, which was exceptionally dry across the UK (Met Office, 2018). Apart from the summer of 2018, high flow events were recorded regularly throughout the baseline monitoring period (Figure 2.4). The highest stage peaks were recorded during Storm Aileen (13 September 2017), ex-hurricane Ophelia (14 October 2017) and Storm Bronagh (20 September 2018). Rainfall totals and peak discharge recorded at the downstream Environment Agency operated gauging stations for the ten largest events recorded on the impact stream can be found in Table 3.5 in Chapter 3. Storm Aileen and Storm Bronagh caused significant flooding of transport links and properties in North Yorkshire (Flood List, 2020). Not all events were recorded on all three streams because equipment failure led to periods of missing data on each of the streams.

During the quality assurance process high levels of uncertainty were identified in the stage data. Whilst some error sources could be corrected (e.g. the backwater effect from the collapse of a cattle bridge downstream of a gauging station), others could not. Small, gradual changes in the relationship between upstream and downstream stage of magnitude up to ± 0.05 m were identified frequently


Stream 1
Stage (m)
U
U*
D
D*
0.3
0.2
0.1
0.0
0.08
0.06
0.04
0.02
0.00
0.6
0.4
0.2
0.0
0.050
0.025
0.000
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10


Stream 3
Stage (m)
U
U*
D
D*
0.5
0.4
0.3
0.2
0.1
0.0
0.05
0.00
-0.05
0.04
0.02
0.00
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10

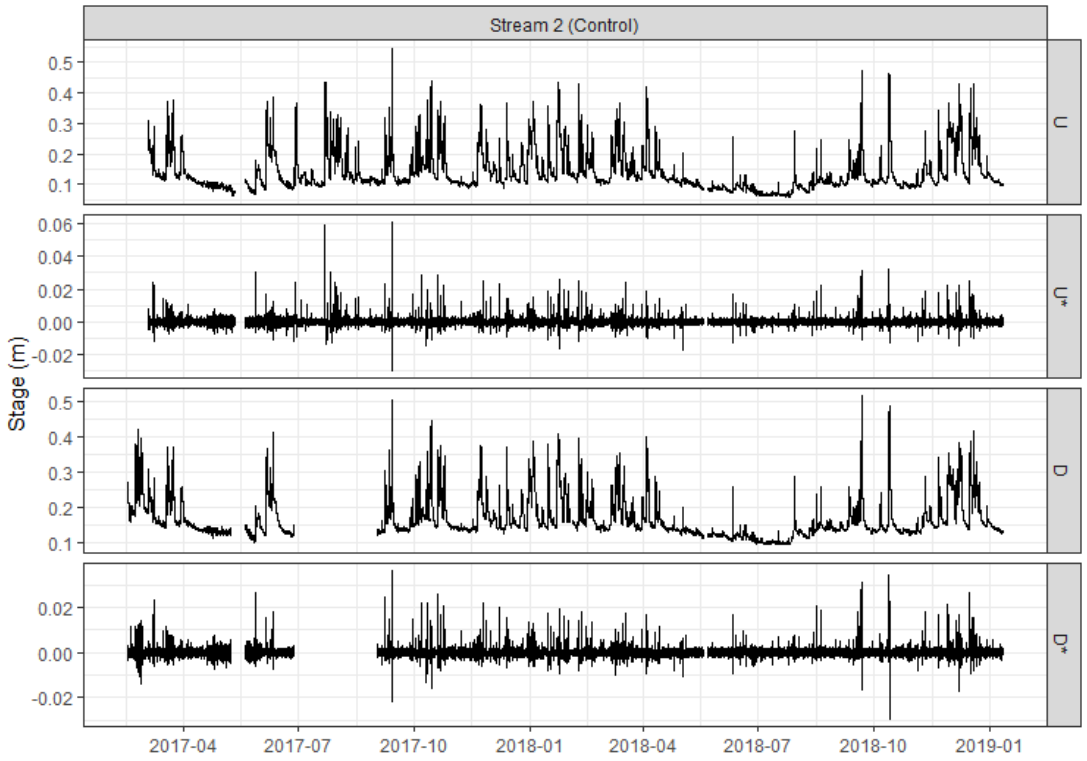


Figure 2.4: Upstream (U) and Downstream (D) stage series collected in the three study streams with first order differenced series U^* and D^*

but as these changes were gradual rather than instantaneous, they could not be readily corrected. Because the errors were constant over the range of stage (i.e., when there was a shift of 0.05 m at baseflow there was also a shift of 0.05 m at the peak) they were assumed to be 'datum errors' brought about by a change in the reference datum or flow conditions at the gauging station. Based on field observations the most likely source of datum error was frequent blockage of the gauging stations with material on the outside of the gauge stilling well, and sediment on the inside of the stilling well.

The upstream (U) and downstream (D) stage series (Figure 2.4) were non-stationary and were therefore transformed by first order differencing. The differenced stage time series fluctuated about a mean of zero between -0.06 m and 0.09 m meaning that the biggest recorded rise in maximum stage in a 15-minute

period was 0.09 m. The differenced stage series, U^* and D^* passed the ADF and KPSS tests for stationarity (Table 2.2) and form the predictor and forecast series used to train the time series model. Spikes in the differenced stage series were found to relate to steep rises in the event hydrographs and were generally positive because the stage fell more gradually after the peak of an event than it rose. Where a positive or negative spike occurred in U^* it was accompanied by an equivalent spike in D^* and is therefore likely to reflect actual changes in water level, rather than errors in the data. Because the differenced data was the rate of change of stage in each 15-minute time step, rather than the absolute stage, it had the benefit of being independent of the datum errors identified during the quality assurance process.

2.3.2 Parsimonious TFN model form

For stream 1, the first five regression terms notably improved the model AIC score whilst beyond five terms adding more regressors had little impact on the AIC score (Figure 2.5). The assumption of independence of the regression residuals was met only when the 6^{th} regressor was added to the model. For stream 2 (control) the AIC value was decreased by including lags 1 to 4, but beyond $k = 5$ there was little improvement to the AIC score and little change in the form of the ACF plot. On stream 3 there was a more gradual decrease in the AIC score, with up to eight regressors notably improving the AIC score and little difference in the ACF plots. To check for overfitting, the results of model validation for the model with $k = 8$ were compared to the results of the model with $k = 7$, 6, 5, 4, and 3 in turn. There was little improvement in the model fit for the model with $k >4$ and so the $k = 4$ model was taken forward. Hence, the data for stream 1, stream 2 (control) and stream 3 were modelled with six, five and four dynamic regression terms, respectively. The model for stream 1 included lag 0 to 3, lag 11 and lag 20. The model for stream 2 (control) included lag 0 to 4 and lag 11. The model for stream 3 included lag 0, 1, 3, and 4.

The ACF and PACF plots for the noise series of the transfer function, N_t, for each stream (Figure 2.6) indicated that it was necessary to use mixed AR and MA models to represent the data as none of the plots exhibited a pure AR or MA

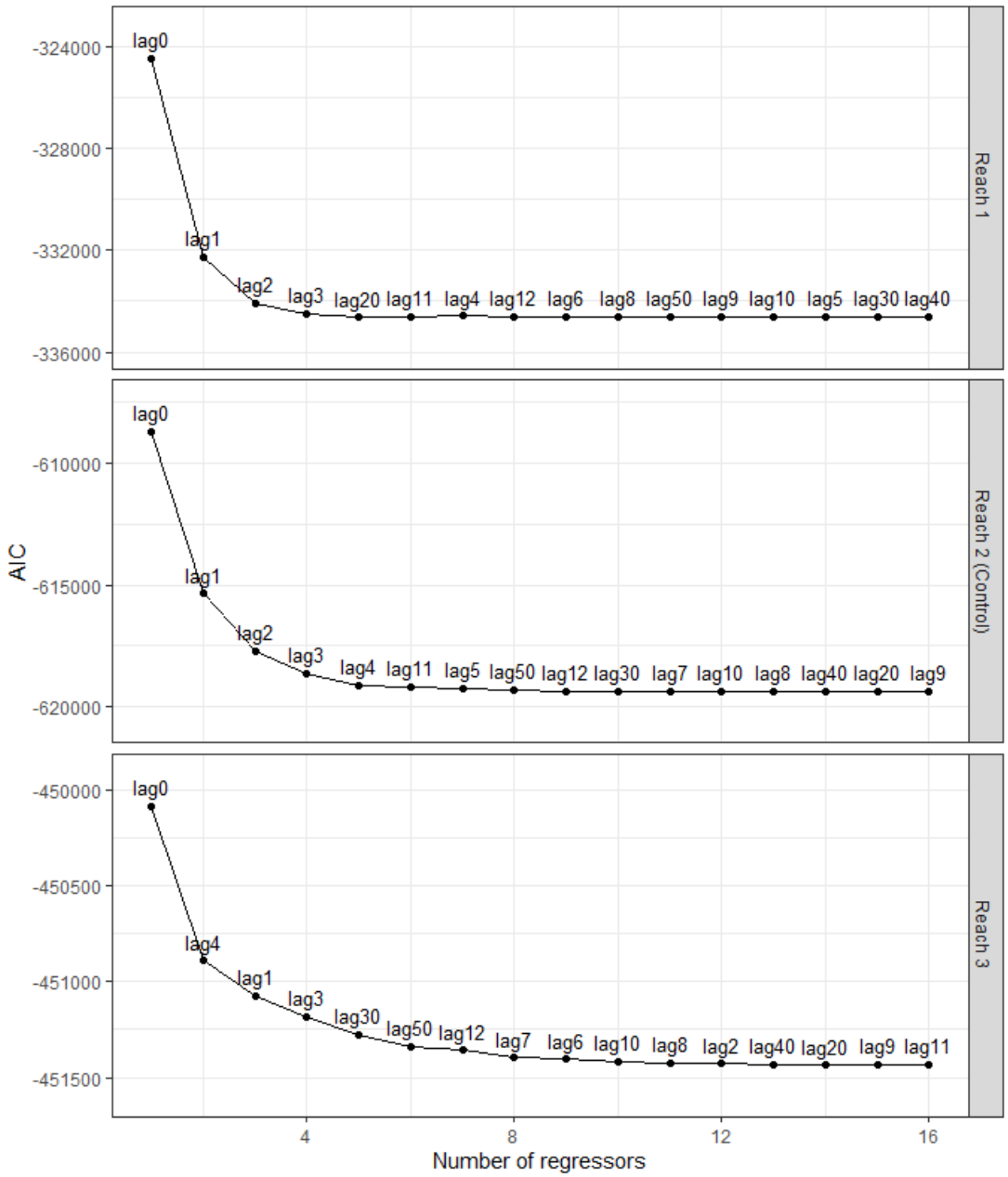


Figure 2.5: Results of forward stepwise regression to find the appropriate number of dynamic regression terms for a parsimonious form of the transfer function. The order of the noise model was allowed to vary up to p=5 and q=5 at each step. Each point shows the regressor which added the most predictive skill (model with lowest AIC score) to the model for each stream at each step of the forward stepwise regression

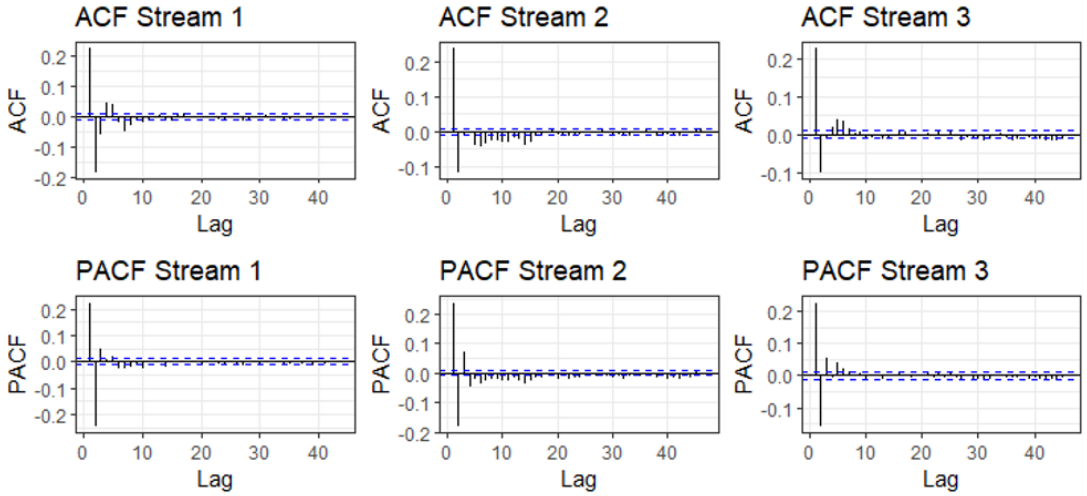


Figure 2.6: ACF (first row) and PACF (second row) of regression noise series for all three streams, blue dotted lines indicate significance level. A noise series with an autoregressive signature has a cut-off in the PACF followed by no significant values, and a gradually dampening ACF plot. A noise series with a moving average signature has a gradually dampening PACF and a cut-off in the ACF followed by no significant values.

signature. By testing all possible combinations of AR and MA terms for each noise series the model with minimum AIC score was found to be the ARMA (5, 0, 3) model for stream 1, the ARMA (4, 0, 2) model for stream 2 (control) and the ARMA (2, 0, 3) model for stream 3.

2.3.3 Modelling assumption checks

Diagnostic plots (Figure 2.7) show that the assumptions of independence and homoscedasticity underlying the innovation series were approximated by the residuals of the models. The model for stream 1 met all of the assumptions; there were two significant values of the ACF (Figure 2.8), but this is permissible at the 95% confidence limit. The residuals of stream 2 (control) and stream 3 contained significant values of the ACF at higher order lags. Results of the Breusch-Godfrey test indicate that residuals were uncorrelated up to order 7, 14 and 18 for stream 1, 2 and 3 respectively. There was some heteroscedasticity in the fit of the data on stream 2 (control) and stream 3 but further transforming the data, using different lagged regressors or changing the order of the ARMA model did not improve the autocorrelation or heteroscedasticity present in the data from stream 2 (control) and stream 3.

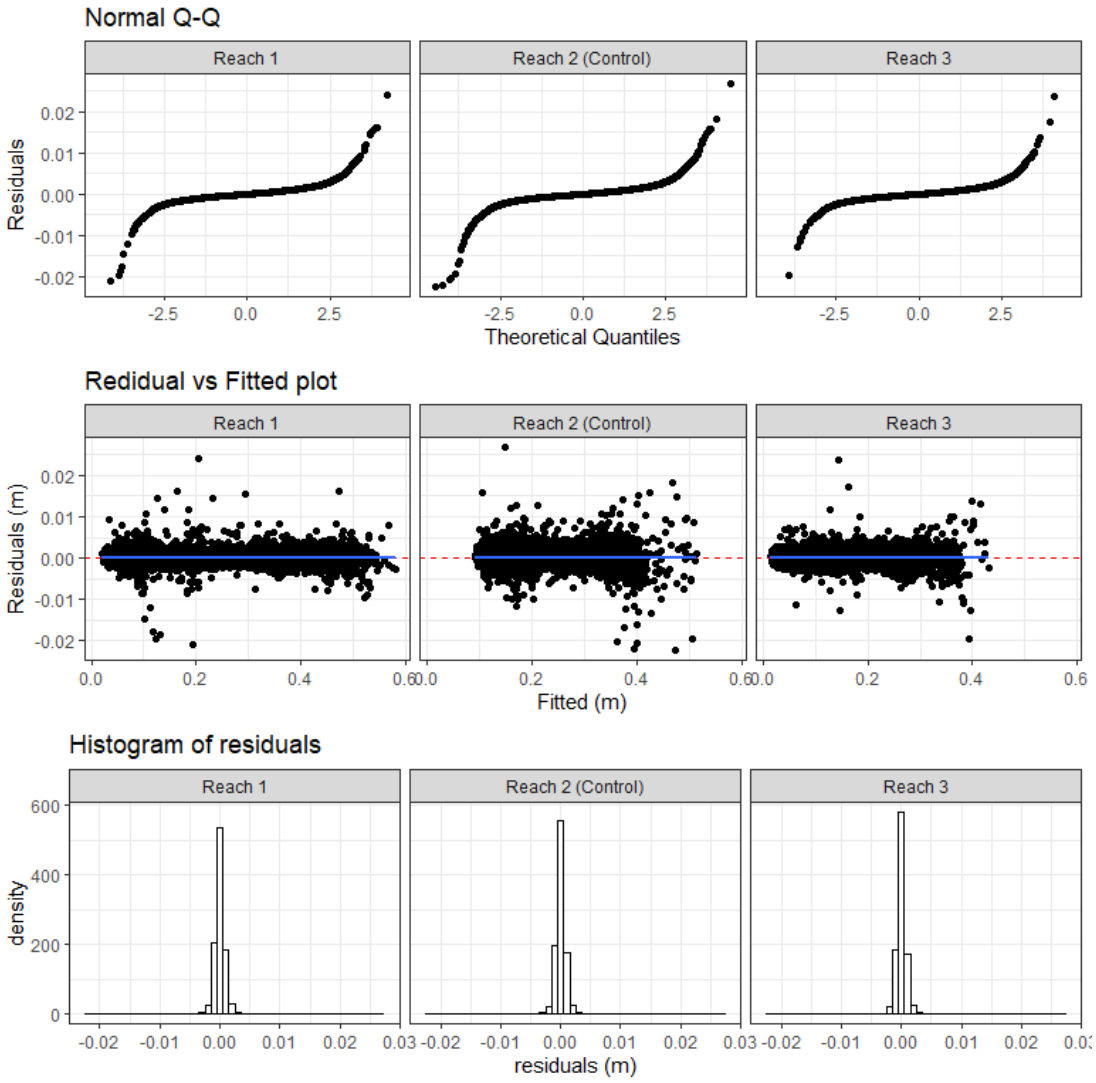


Figure 2.7: Residual diagnostic plots for fully fitted TFN models

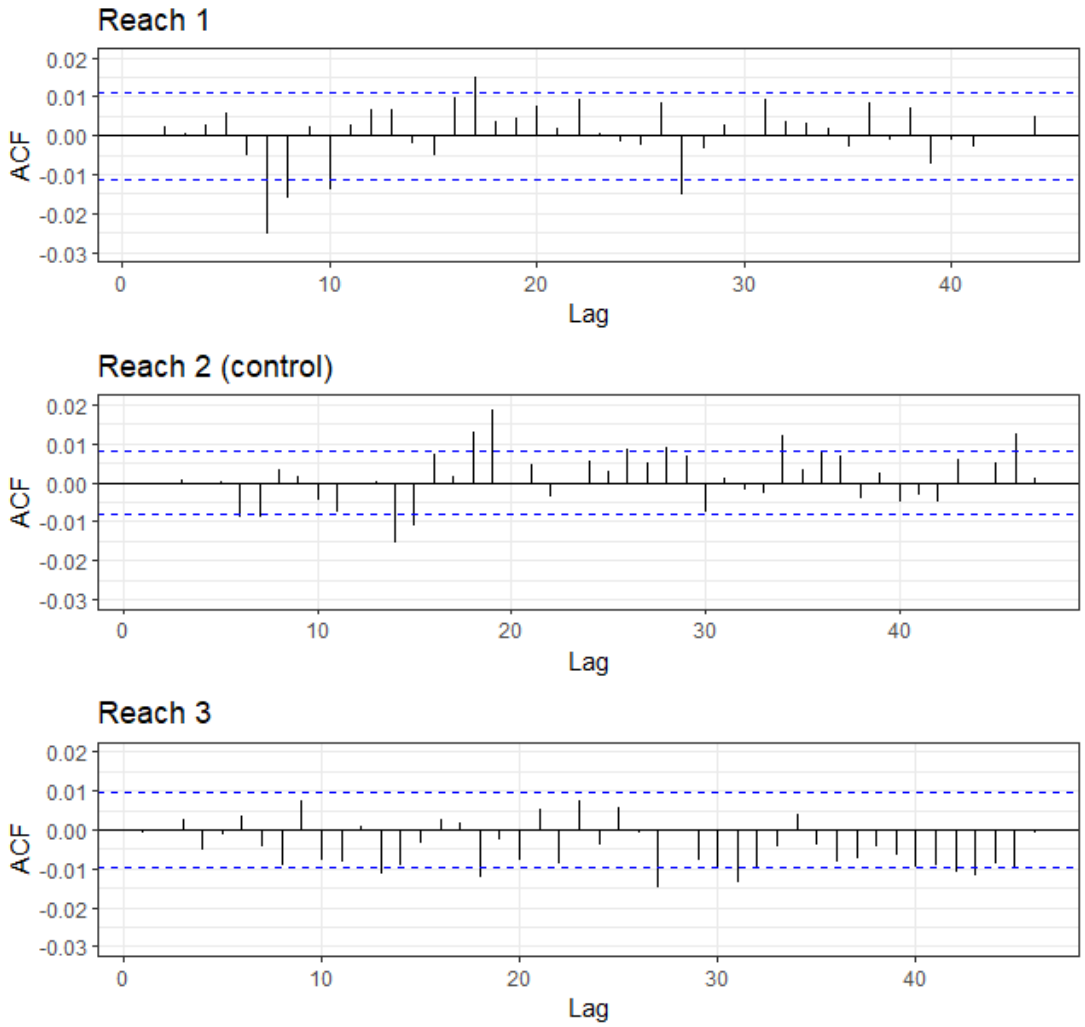


Figure 2.8: Autocorrelations in residuals of the fully fitted TFN models

2.3.4 TFN model equations

The TFN models describing baseflow conditions in stream 1, stream 2 (control) and stream 3 are given in Equations 2.18–2.20. The coefficients of the identified TFN model parameters were estimated by MLE and are given in Table 2.3 along with their standard error. The close fit of the data to the one-to-one line on the plots of fitted against observed values in Figure 2.9 show that the good in-sample fit of the model to the data (RMSE 0.00097 m).

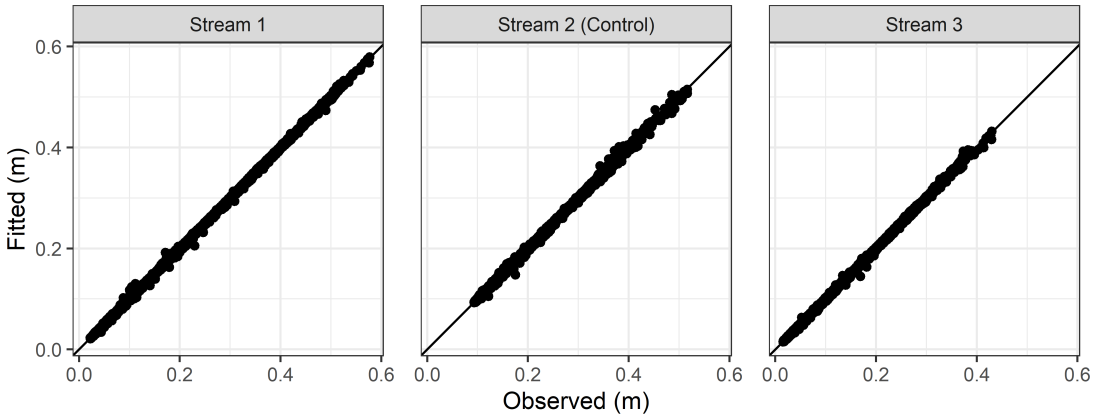


Figure 2.9: Points of fitted downstream stage plotted against observed downstream stage with the one-to-one line for reference.

Stream 1:

$$\begin{aligned} D_t^* &= 0.687U_t^* + 0.332U_{t-1}^* + 0.165U_{t-2}^* + 0.102U_{t-3}^* - 0.014U_{t-11}^* \\ &\quad -0.01U_{t-20}^* + N_t \\ N_t &= 0.893N_{t-1} + 0.179N_{t-2} - 0.437N_{t-3} + 0.288N_{t-4} - 0.162N_{t-5} \\ &\quad -0.554a_{t-1} - 0.654a_{t-2} + 0.453a_{t-3} + a_t \end{aligned} \tag{2.18}$$

Stream 2 (Control):

$$\begin{aligned} D_t^* &= 0.430U_t^* + 0.198U_{t-1}^* + 0.114U_{t-2}^* + 0.083U_{t-3}^* \\ &\quad +0.042U_{t-4}^* + N_t \\ N_t &= 1.035N_{t-1} - 0.196N_{t-2} + 0.173N_{t-3} - 0.062N_{t-4} - 0.739a_{t-1} \\ &\quad -0.253a_{t-2} + a_t \end{aligned} \tag{2.19}$$

Stream 3:

$$\begin{aligned} D_t^* &= 0.582U_t^* + 0.0386U_{t-1}^* + 0.0396U_{t-3}^* + 0.0404U_{t-4}^* + N_t \\ N_t &= 1.147N_{t-1} - 0.441N_{t-2} - 0.865a_{t-1} - 0.021a_{t-2} + 0.030a_{t-3} + a_t \end{aligned} \tag{2.20}$$

2.3.5 Simulations of downstream baseline stage during high flow events

Between 32 and 54 high flow events were identified in the data on each of the streams. Different numbers of high flow events were identified on the streams

Table 2.3: Parsimonious TFN model parameter coefficients; coeff. = coefficient, s.e. = standard error in metres

		Stream 1		Stream 2 (control)		Stream 3	
	Parameter	coeff.	s.e	coeff.	s.e.	coeff.	s.e.
Transfer Function Parameters	U^*_t	0.687	0.005	0.430	0.003	0.583	0.003
	U^*_{t-1}	0.332	0.005	0.198	0.003	0.039	0.004
	U^*_{t-2}	0.165	0.005	0.114	0.003	-	-
	U^*_{t-3}	0.102	0.005	0.083	0.003	0.040	0.004
	U^*_{t-4}	-	-	0.042	0.003	0.040	0.004
	U^*_{t-11}	-0.014	0.004	-	-	-	-
	U^*_{t-20}	-0.010	0.003	-	-	-	-
AR terms	N_{t-1}	0.893	0.083	1.035	0.039	1.147	0.040
	N_{t-2}	0.179	0.097	-0.196	0.050	-0.441	0.037
	N_{t-3}	-0.437	0.061	0.173	0.020	-	-
	N_{t-4}	0.288	0.032	-0.062	0.008	-	-
	N_{t-5}	-0.162	0.014	-	-	-	-
MA terms	a_{t-1}	-0.554	0.083	-0.739	0.039	-0.865	0.039
	a_{t-2}	-0.654	0.074	-0.253	0.039	0.021	0.030
	a_{t-3}	0.453	0.051	-	-	0.229	0.011

because of periods of missing data, and due to the interventions being installed a few months earlier in stream 3 than in stream 1. To validate the model simulations using data which were not used to train the model, each event was removed in turn and the model coefficients were re-estimated on the remaining data. The downstream stage of the 'hold out' event was then simulated by providing the model with the upstream stage series of the 'hold out' event. The simulated downstream stage was compared to the downstream stage observed during the event to assess the accuracy of the simulation. Overall, the model simulations fit the observed stage well (NSE 0.996–0.976), but it can be seen (Figure 2.10) that the peak of the event was not always captured by the model. Figure 2.10 gives examples of a large, medium and small event simulation on each stream together with the empirical prediction intervals and the observed downstream stage.

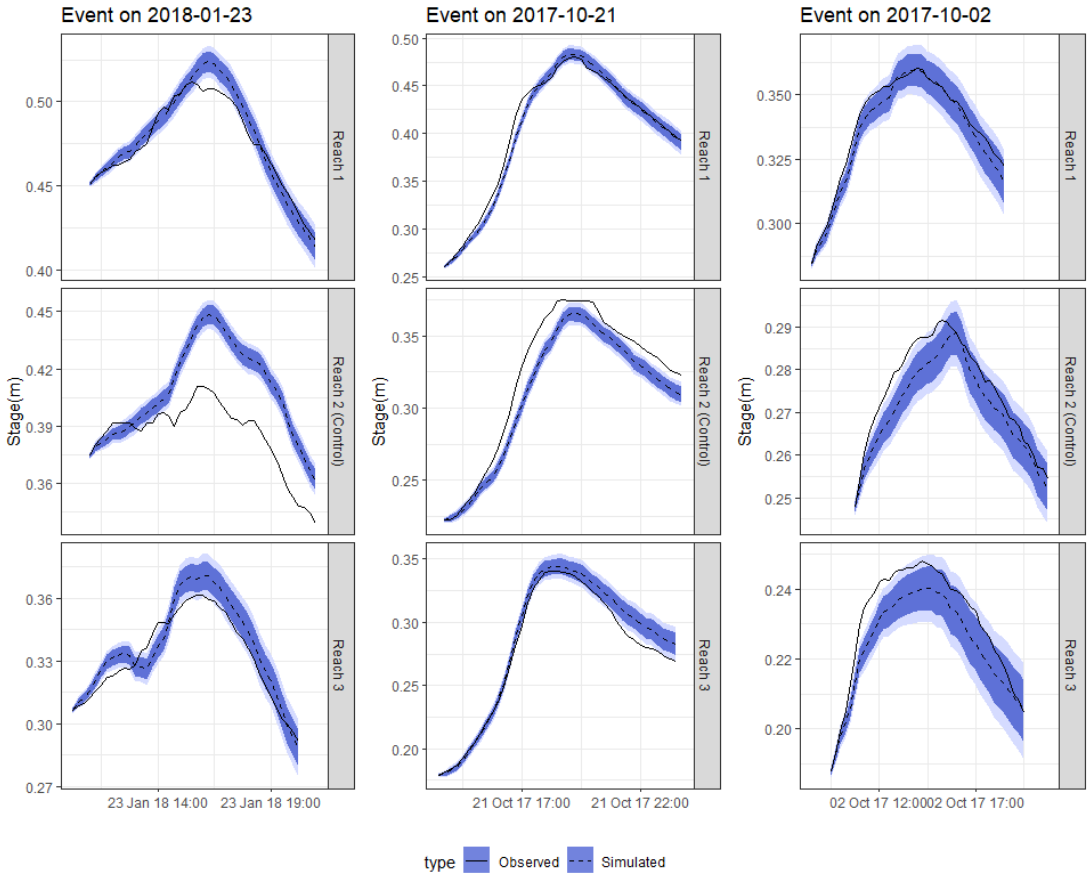


Figure 2.10: Observed (solid line) and simulated (dashed line) downstream stage during a large (first column), medium (second column), and small (third column) high flow event with 80% (light blue shading) and 95% (dark blue shading) empirical prediction intervals. Note the changing y-axis on the plots.

2.3.6 Skill of the model at simulating out of sample downstream peak magnitude

To assess how well each of the models performed at simulating the downstream peak magnitude of events the peak error (PE), and peak error percentage (PEP) were calculated (Figure 2.11). The error in simulating event peak magnitude was smaller than ± 0.03 m for all but four of the simulated events. On stream 1 the PE was smallest and was distributed evenly above and below zero, indicating that the simulations were not biased. On stream 2 (control) and stream 3 the range in PE and PEP was larger, and although the median value of PE was close to zero, the majority of event magnitudes were under predicted on stream 2 (control), and over predicted on stream 3. Whilst the PEP was below ± 5% for all simulations on stream 1, it was up to ± 11% on stream 2 (control) and ± 9% on stream 3. The interquartile range (IQR) of PEP was within ± 3%, for all three streams, with the IQR on stream 1 being as low as ± 1%.

By plotting the observed event peak magnitude against the simulated event peak magnitude (Figure 2.12), it can be seen that the points lay close to the one-to-one line for stream 1 (NSE 0.994) with a residual standard error of 0.008 m. The simulated peak magnitudes were both over and under predicted and were not affected by the event peak magnitude. On stream 2 (control), however, it can be seen that the PE was relatively small (RMSE 0.006 m) for events below 0.35 m but increased to a RMSE of 0.027 m and 0.036 m for events with peak magnitude greater than 0.35 m and 0.40 m respectively. For the largest events the peak magnitude was under predicted more frequently than it was over predicted, resulting in a linear relationship which lay below the one-to-one line (NSE 0.97). There was a relationship between PE and event magnitude on stream 3; events with peak magnitude below 0.25 m were under predicted whilst events with peak magnitude greater than 0.25 m were over predicted. Although the NSE was relatively high (0.98) the coefficients of the linear regression reflect the bias in the model simulations.

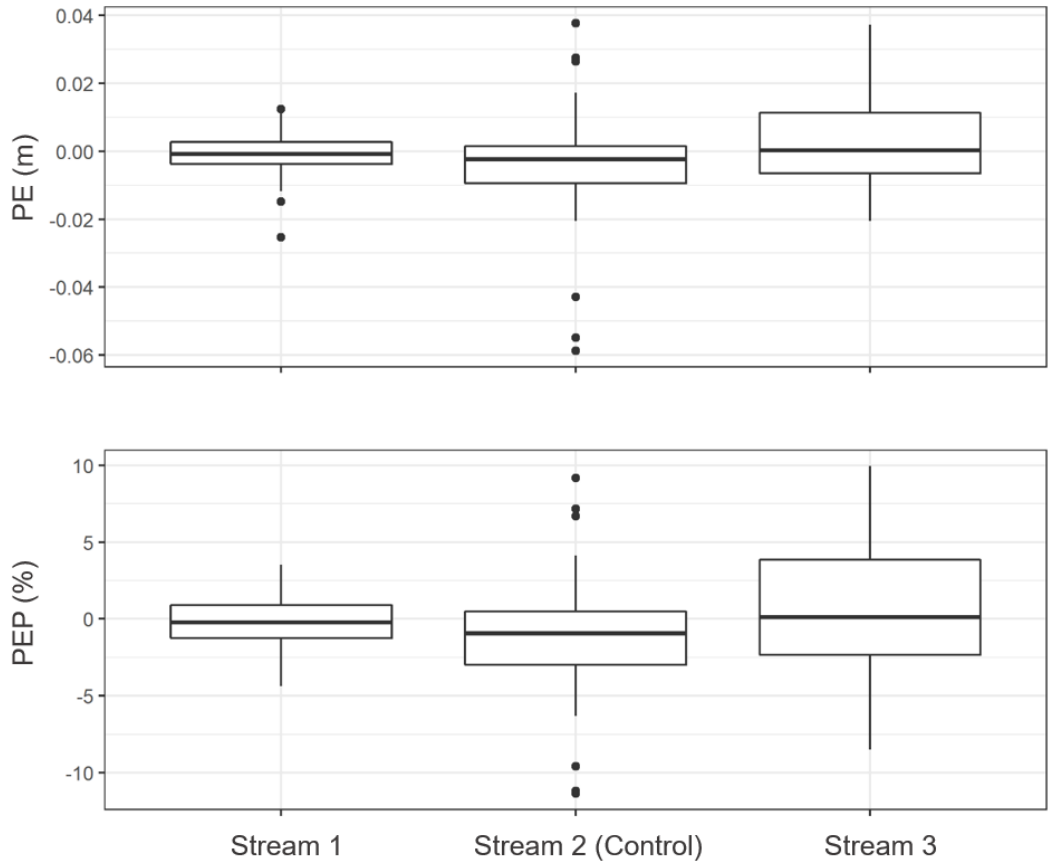


Figure 2.11: Skill of TFN models at predicting out of sample event peak magnitude. Peak error (PE) and Percentage error in peak (PEP) is the difference between the observed and simulated downstream peak stage.

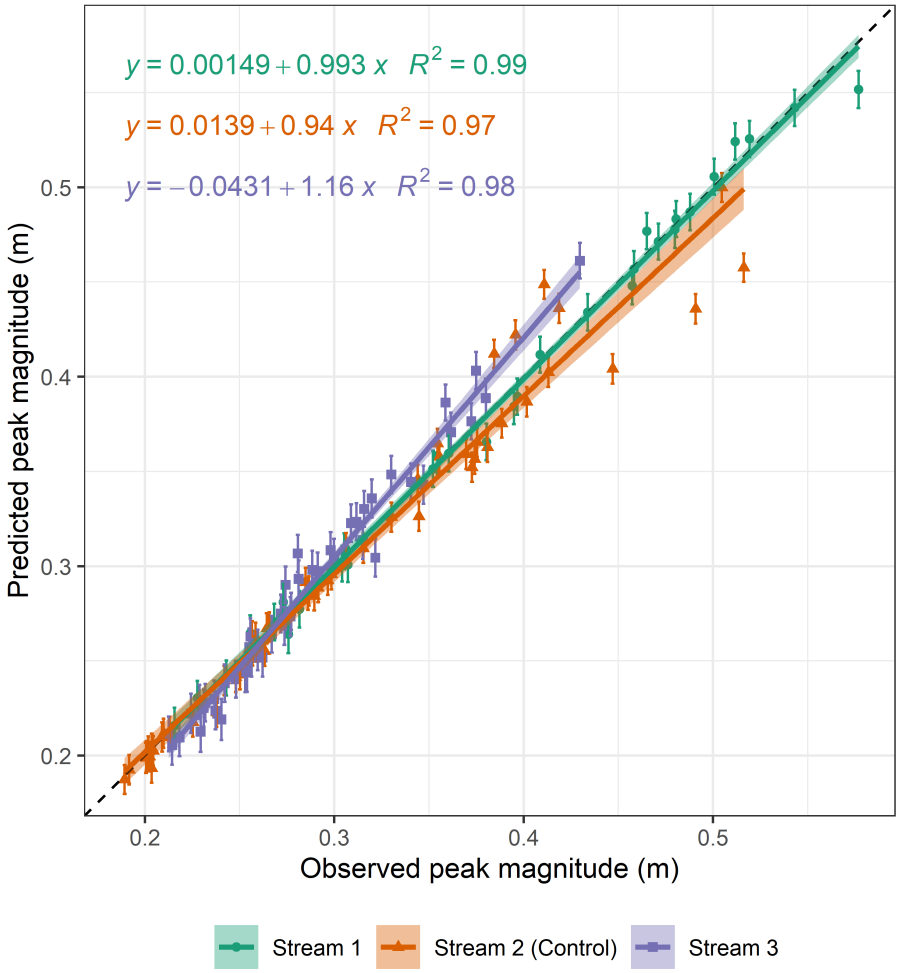


Figure 2.12: Goodness of fit of out of sample simulations of event peak stage with theoretical prediction intervals. The shaded areas shows the confidence interval of the linear relationship between simulated and observed peak magnitude and the dashed line shows the 1-to-1 relationship

2.3.7 Skill of the model at simulating out of sample downstream peak timing

For each high flow event the peak of the event recorded upstream occurred at the 20^{th} timestep, five hours into the simulation. The majority of observed downstream event peaks occurred after the upstream event peak, as expected at timesteps >20. However, it can be seen in Figure 2.13 that the model had little skill at simulating the timing of the downstream event peak; peak timing of the majority of events was predicted to be earlier than they occurred in the observed data. However, this may be related to issues with event peak timing identified in the data described in Appendix B.

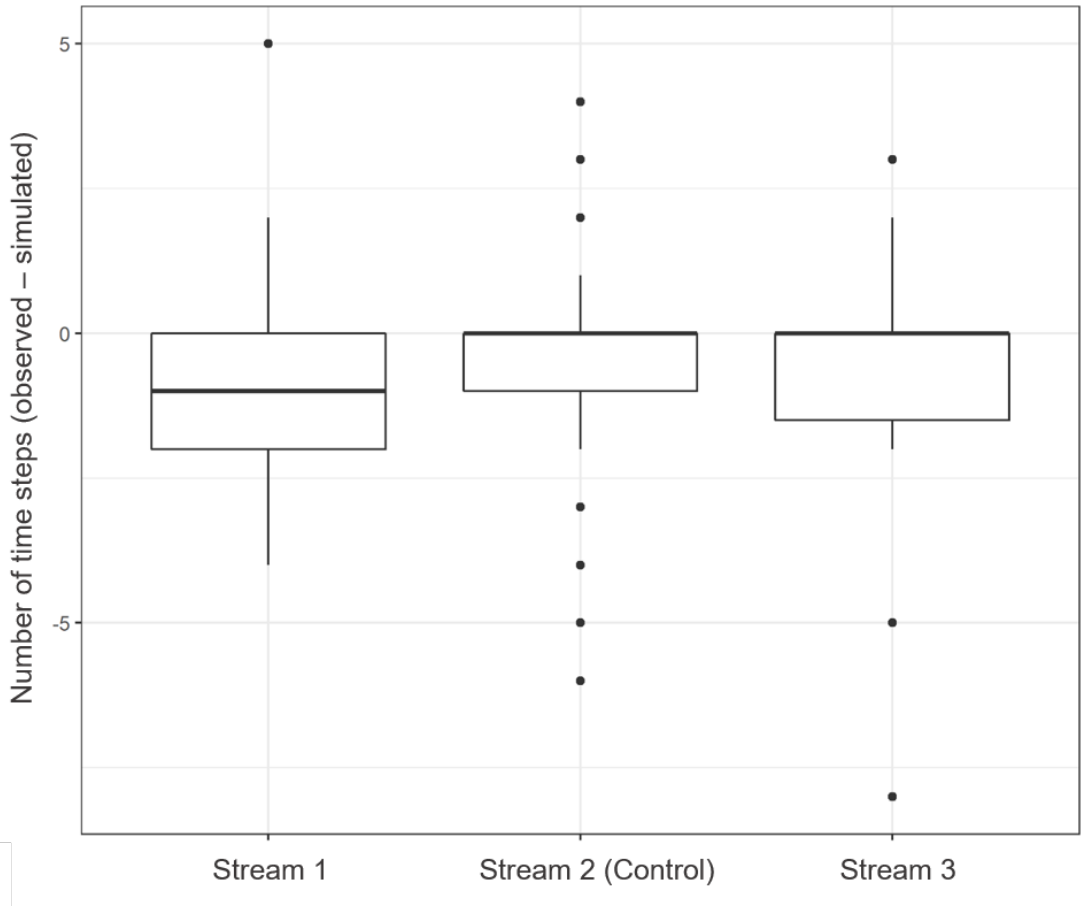


Figure 2.13: Skill of the model at simulating event peak timing given by peak timing error, the difference between the observed and simulated peak timing given in number of 15-minute timesteps

2.3.8 Empirical prediction intervals for out of sample simulations of baseline stage

The peak error (PE) distributions shown in Figure 2.14 were used to determine the magnitude of the empirical 80% and 95% prediction intervals at the peak of the simulated events given in Table 2.4. By definition the simulated peak magnitude prediction intervals captured the observed peak magnitude for 80% and 95% of the events. In comparison, the theoretical prediction intervals at peak were on average 0.019 m, 0.015 m, 0.02 m for stream 1, 2 and 3 respectively and captured the peak two thirds of the time.

Table 2.4: Empirical prediction interval width at event peak; PI = prediction interval

	Lower 95% PI	Upper 95% PI	Lower 80% PI	Upper 80% PI
Stream 1	-0.017	0.012	-0.011	0.008
Stream 2 (control)	-0.051	0.027	-0.018	0.007
Stream 3	-0.016	0.032	-0.009	0.017

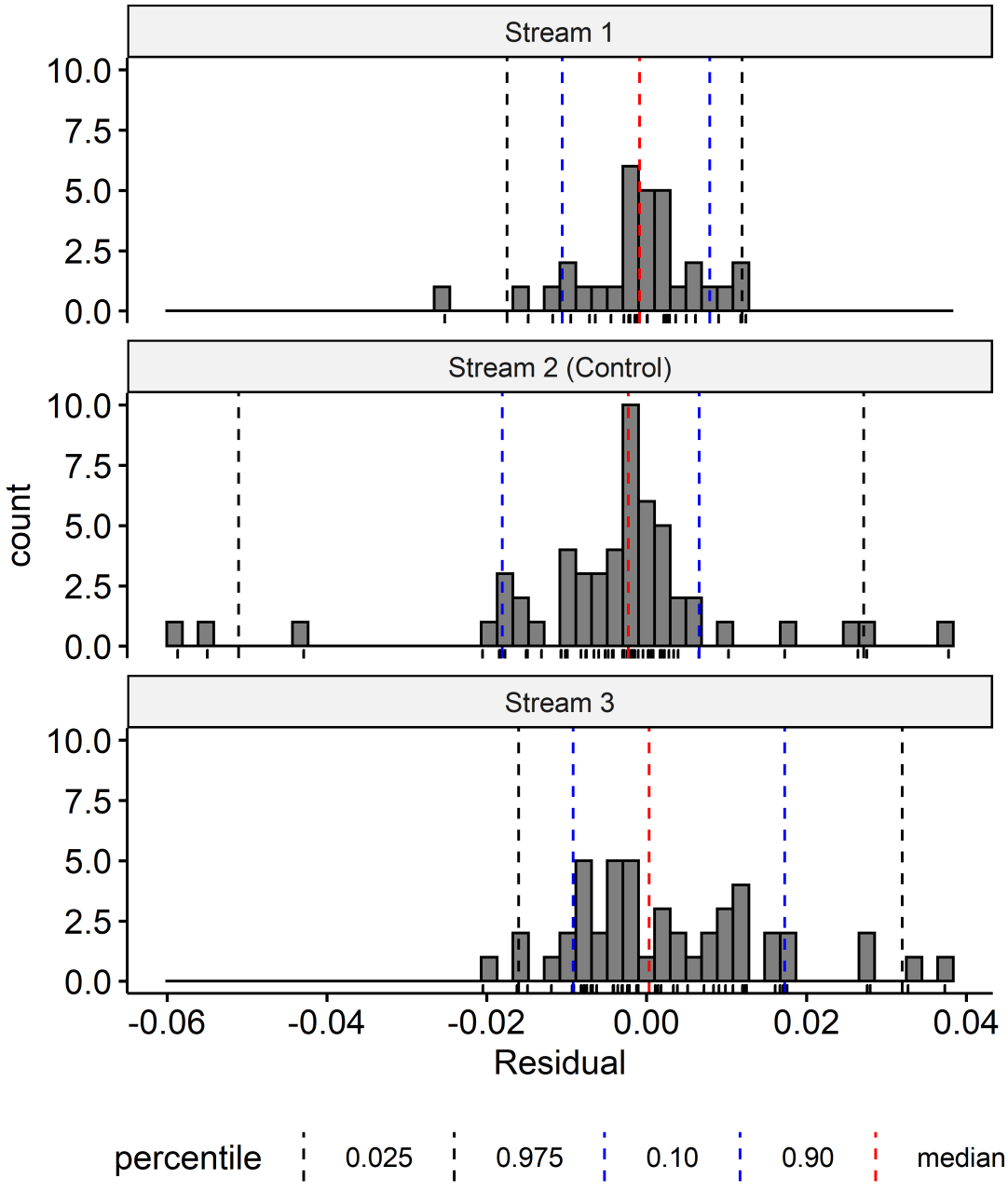


Figure 2.14: Empirical prediction intervals at 80% and 95% confidence level determined from distribution of the peak error (PE)

2.4 Discussion

Efforts to address the lack of quantitative evidence of leaky dam impacts on flood peak magnitude (Burgess-Gamble et al., 2017) have been hampered by lack of baseline data (Burgess-Gamble et al., 2017; Ellis et al., 2021), lack of comparable events monitored before and after the installation of leaky dams (National Trust, 2015), and high level of uncertainty in stage and/or discharge series (Connelly et al., 2020; Lane, 2017). This study has demonstrated that data-based time series modelling can be used to overcome the barriers presented by poor quality, relatively short baseline data series.

The potential of using a TFN model to extract useful information about flood peak magnitude and timing from highly uncertain baseline data was demonstrated using data from three hydraulically similar upland streams. The stage time series, collected using a BACI experimental design, were prone to shifts in the stage datum which meant the peak stage magnitude of the events recorded during the monitoring period were too uncertain to directly detect changes in peak stage magnitude before and after the installation of 7–8 leaky dams in the impact streams. Instead, the stage time series data from the control and impact streams was used to fit and validate a class of time series models called transfer function noise (TFN) models to make accurate simulations of baseline response of the streams to high flow events.

The TFN models were able to simulate downstream stage for events which had not been used to fit the model to within ± 2 cm at the peak of the event at the 80% confidence limit. The model for one of the impact streams performed particularly well, simulating downstream event peak magnitude to within ± 2 cm at the 95% confidence level across the range of stage. There were some issues with bias and heteroscedastic errors in the other two models, indicating a different type of model or data transformation may be more suitable on these streams. Nevertheless, the model for the control stream was found to perform well (NSE 0.97) for all but the six largest of 54 observed events despite its relatively poorer fit at higher flows. All three models were less effective at simulating the timing of the event peak, with errors of up to two hours.

Stage data uncertainty

In this study small datum shifts (<0.05 m) were common in the stage time series data and could not be readily corrected because they were small and gradual (Wilby et al., 2017). Field observations imply that these small datum shifts were commonly caused by deposition of coarse material on the outside of the stilling well and deposition of fine material within the stilling well. Frequent inspection, maintenance and calibration of gauging stations could mitigate these problems where sufficient resource is available. This has implications for resource allocation to the monitoring of NFM pilot schemes, the results of which are key in securing future funding (Hammond et al., 2011).

The finding that hydrological data was highly uncertain is not unique to this study; it has long been acknowledged that hydrological data is 'messy' (Beven and Westerberg, 2011). As a result, a wide range of methods have been developed to quantify the uncertainty in streamflow series (see Coz (2012) & Kiang et al. (2018) for recent reviews) but these methods are largely focused on how the uncertainty in the rating relationship propagates to the discharge series (Kiang et al., 2018) without taking into account the uncertainty in the stage series (Guerrero et al., 2012) which can be one of the most significant sources of error (Di Baldassarre and Montanari, 2009).

For example, Westerberg et al. (2011) found changes of ± 20% at intermediate to high flows in the gauging station of an alluvial river with a catchment area of 1766 km^2 and (Guerrero et al., 2012) found that discharge for a given stage could vary by a factor of two or more in a temporal study of six gauging stations for river basins which varied in size by two orders of magnitude in the same area. For six hydrometric stations in France with catchment areas ranging from 54 km^2 to 11,000 km^2 systematic stage errors in the stage series measurements of between ± 0.5 cm and ± 6.8 cm accounted for uncertainty equivalent to 4% to 12% of the daily average discharge (Horner et al., 2018).

Stage datum errors are often assumed to be negligible (Horner et al., 2018) but as for gauging stations in larger catchments in the above examples, the stage datum errors at all gauging stations in the study site outweighed the uncertainty

introduced by the error in the rating relationships (see Appendix B). Identifying such errors necessitates a thorough quality assurance process which requires sufficient resource allocation to post-processing of the data.

This study has demonstrated that valuable information can be extracted from such highly uncertain data by using time series modelling techniques, which requires further resource. Crucially, it has shown that the use of an empirical BACI experimental approach, which is recommended for the assessment of natural flood impacts on flood peak magnitude (Ellis et al., 2021) alone is not sufficient unless very low levels of monitoring uncertainty are achieved. The findings of this study are in agreement with the monitoring guidance developed for the UK Government funded £15 million worth of NFM pilot projects (HM Treasury, 2016) that proper specialist resource, such as an academic study or report, is required to monitor the extent to which the NFM measures reduce downstream flood risk (Arnott et al., 2018) but would extend this recommendation to include post-processing and analysis of the data.

Model accuracy

It is difficult to make comparisons between levels of accuracy achieved by TFN models in different studies because it depends on many factors including the quality of the input data; model timestep and model structure (Sene and Tilford, 2004). The accuracy of the models were therefore assessed in terms of the intended purpose of the model: simulating downstream stage under baseline conditions to a high enough degree of accuracy to be able to compare it to observations of downstream stage response after the leaky dams were installed. This presents difficulties as the impacts of leaky dams in upland watercourses are not yet known (Burgess-Gamble et al., 2017). However, it is known that wood placed in upland channels for the purpose of river restoration reduced event peak magnitude by 8% and 2.2% in steep watercourses (Keys et al., 2018; Wenzel et al., 2014) and in lowland rivers peak magnitude reductions of 21% have been observed for combined planform and large wood restoration (Kitts, 2010).

The previous research of Kitts (2010); Wenzel et al. (2014) & Keys et al. (2018) suggests that simulating event peaks to within ± 5% on stream 1 may be

sufficiently accurate to detect leaky dam impacts on individual events. Furthermore, by combining the simulations of all events and calculating the confidence interval of the relationship between observed and simulated stage it becomes clear that there is a very high level of confidence in the ability of the model to replicate downstream event peak magnitude on average, particularly on stream 1.

The results on stream 1 demonstrate the potential that linear TFN models have for extracting information about peak magnitude from highly uncertain baseline data. Further work is required to identify models which provide a better fit to the data on stream 2 (control) and stream 3. It is likely that the lack of fit is due to non-linearity in the response (e.g. due to differences in the geometry of the gauging cross-sections (Romanowicz et al., 2008)), which can be incorporated in the TFN approach by including a non-linear transform to the data. The DBM approach, for example, has been developed to model typically non-linear relationship between rainfall and runoff (Young, 2003) and has been successfully applied to model non-linear level to level responses (Beven et al., 2008; Leedal et al., 2010; Romanowicz et al., 2008; Young, 2002).

2.4.1 Implications

Whilst there are many examples of the high level of predictive ability achieved using transfer functions in hydrology the majority of applications are in rainfall runoff modelling (see for example, Katimon et al. (2013); Ratto et al. (2007) & Young (2003)). This study shows that it is possible to achieve sufficiently high levels of accuracy in predictions of the level-to-level response of small upland watercourses to detect leaky dam impacts. The implication of being able to make highly accurate simulations is that observations of post-intervention events can be compared to simulations of the baseline scenario with a high degree of confidence. Assuming stationarity in the first difference into the future and given adequate control data, any differences in the simulated baseline response and observed intervention response can be attributed to the NFM interventions. Notably, baseline and post-intervention monitoring datasets are therefore not required to contain events which are so similar that their responses can be compared directly.

This is significant as the stochastic nature of flood events has precluded the assessment of NFM impacts in several studies (Kitts, 2010; National Trust, 2015), or provided evidence of only one or two comparable events (Nisbet et al., 2015a; Thomas and Nisbet, 2007; Wilkinson et al., 2010). By simulating downstream baseline stage, the impact of interventions can be assessed for every event in the post-intervention monitoring period providing replicates and allowing impacts to be assessed at a range of event magnitudes.

The problem of capturing a range of events may not always be overcome by using this approach; in this study several notable high flow events were captured, but baseline data which does not contain many high flow events cannot be used to train a model to simulate high flow events. However, Young (2003) showed for the case of rainfall-runoff modelling that high levels of accuracy ($R^2 = 0.94$) can be achieved using a model fitted and validated using only 20 days of hourly observations. Hence, by using a data-based modelling approach assessing the impacts of NFM measures, even where only very short periods of baseline data are available, may become viable.

Increasing the quantitative evidence base of NFM impacts is crucial for measures such as leaky dams to become more mainstream in flood risk management (Ellis et al., 2021) and environmental land management, particularly following the UK's exit from the EU (Klaar et al., 2020). Quantitative evidence of leaky dam effectiveness is needed to assess whether they are a viable flood risk management technique in upland catchments, during what types of flood events they are effective, and to inform the design of natural flood management schemes. Furthermore, empirical evidence is needed to validate the representation of leaky dams in hydraulic and hydrological models (Addy and Wilkinson, 2019) which can be used to design and assess the effectiveness of flood risk management schemes at a range of spatial scales. Such models can be used to take into account the potential for synchronisation or de-synchronisation of relative tributary flows which can have significant impacts on downstream flood risk (Dixon et al., 2016; Pattison et al., 2014).

2.4.2 Limitations of the approach

The linear transfer function model was able to provide a good fit to the data on stream 1 but violated the assumptions of independent and identically distributed residuals on the other two streams. This was reflected in the out of sample performance of these models; when used to simulate events which had not been used to train the model the error in peak magnitude increased with peak magnitude on both stream 2 (control) and stream 3, with an obvious bias in the peak estimates of stream 3. The relationship between peak magnitude and peak error on both streams undermined confidence in the model fit and its simulations on these streams.

The lack of fit of the models to the data recorded on stream 2 (control) and stream 3 serves to demonstrate some of the limitations of the relatively simplistic linear TFN family of time series models. Linear rather than non-linear TFN models were applied, in the first instance, in the interest of model parsimony. In future work the knowledge gained by applying linear TFN models to the data in this study will inform the choice of model needed to improve the model fit. The lack of fit on stream 2 and 3 is a limitation of the single input linear TFN approach taken here, rather than a limitation of the data-based modelling approach more generally; it is likely that the model fit could be improved by using a different type of time series model.

The dependence observed in the residuals of the stream 2 (control) and stream 3 models indicates that there was information present in the data which was not captured by the model. This suggests that the models may have been misspecified (Hipel and McLeod, 1994), either by not including important parameters, or by not fitting the right type of model. However, Hyndman and Athanasopoulos (2018) state that, whilst autocorrelated residuals affect the coverage of the theoretical prediction intervals they do not bias the model but merely indicate a more efficient model could be found.

The lack of fit of out of sample simulation on stream 2 and 3 is, therefore, more likely to stem from the violation of the assumption of identically distributed residuals. Larger error at larger values of stage indicate that there may be non-linearity in the downstream stage response at different magnitudes of stage (Beven

et al., 2008). Non-linearity of the stream level to level response could be due to the geometry of the cross-sections of the stream, particularly where there are differences in the form of the stream between the upstream and downstream gauging cross section (Romanowicz et al., 2008). The use of a non-linear transform, such as the state dependent parameter (SDP) of the data-based mechanistic (DBM) modelling approach described by Young (2002) may provide a better fit to the data and improve the model's predictive ability. Alternatively, a class of models which accounts for non-constant variance, such as autoregressive conditional heteroscedasticity (ARCH) models could be used. Applying a Box-Cox transformation, which accounts for non-constant variance by transforming variables to approximate a normal distribution (Hipel and McLeod, 1994) did not improve the model fit.

Another explanation for the lack of fit of out of sample simulations observed on stream 2 (control) and stream 3 is that this could be taken as a sign of model overfitting. Overfitting occurs when too many model parameters are included in the model which captures too much of the complexity of the training data and deteriorates the model's generalisability (Piotrowski and Napiorkowski, 2013). There is a consensus that simulation models with an intermediate level of complexity outperform the simplest and most complex models (Astrup et al., 2008; Beven, 1989; Jakeman and Hornberger, 1993). For rainfall-runoff models Beven (1989) argues that the information in hydrological records should be modelled using 3–5 model parameters, adding that over-parametrized models have many different sets of plausible parameters leading to problems of equifinality (Beven, 2001; Young, 2003), particularly in specifying the order of the dynamic component of the model (Young, 2003).

To avoid overfitting, model identification was carried out using the MAICE procedure, which combines an iterative approach to model fitting with an automated selection criterion, the AIC score. The AIC score is one of the most used selection criteria (von Asmuth et al., 2002) but has been criticised for selecting models with too many parameters (Shibata, 1976). To select a more parsimonious model a selection criteria with higher penalties for at higher model orders, such as the finite sample information criterion (FSIC) of Wensink and Broersen (1994) could be used in future. Alternatively, the DBM approach limits the number of

parameters by specifying that as well as using a statistically objective method, such as the AIC, to determine the model order, all parameters should have a valid physical interpretation (von Asmuth et al., 2002; Young, 2003).

Even where a good enough model fit is found to use the model for its intended purpose of assessing the impact of leaky dams, the assumption has to be made that there would have been no change in the relationship between upstream and downstream stage between the baseline and post-intervention monitoring period had the interventions not been installed. In other words, the output data is assumed to be stationary in the first difference into the future. Factors which affect all streams, such as land-use or climatic factors are controlled for by looking for differences between the baseline and post-intervention monitoring period in the control stream. However, changes which affect streams individually, such as changes in the rating relationship, cannot be controlled for. Homoscedasticity over time of both the in-sample residuals, and out of sample forecasting errors provides some basis for the assumption that the relationship will continue to be stationary into the future.

Datum change errors (small datum shifts at the stage gauge <0.05 m), which were the most commonly observed factor to affect individual streams were avoided by taking the first order difference. However, first order differencing does not avoid changes which affect the rate of change of stage, such as a change in the shape of the cross section. Therefore, it is important that the data be subjected to a thorough quality assurance process before applying a data-based time series modelling approach. This has implications for the resource required for the monitoring of NFM pilot schemes, the results of which are key in securing future funding (Hammond et al., 2011).

The prediction intervals of the models were calculated based on the empirical approach introduced by Williams and Goodman (1971) which takes into account random, parameter and model specification errors without making assumptions about the error distribution. Lee and Scholtes (2014) showed that empirical prediction intervals are not affected by model misspecification and are therefore robust despite the possible implications of the lack of fit indicated by residual diagnostics. They were therefore appropriate despite autocorrelation and heteroscedasticity in the residuals. However, although the method is robust it can

result in prediction intervals which are too wide to be informative. This is confounded by the limitation that this approach is not conditional on the state of the system, which resulted in overestimates of the uncertainty for smaller events on stream 2 (control) and stream 3. As suggested by Lee and Scholtes (2014) the empirical prediction intervals could be improved by taking an adaptive approach conditional on the recent state of the system to reduce the prediction interval width.

Another limitation of the empirical prediction interval approach is that it relies on the availability of sufficient data points (events) to estimate the error distribution (Isengildina-Massa et al., 2011). The number of data points needed depends on the required accuracy (Taylor and Bunn, 1999); Lee and Scholtes (2014) suggest a minimum of 120 data points are required, whilst Taylor and Bunn (1999) recommend 50 data points and Williams and Goodman (1971) found just 24 data points provided sufficient accuracy at the 80% confidence level, although including more observations improved accuracy at the 90% and 95% level. Hence, it is possible that some of the prediction intervals calculated in the study reflect model-fitting errors rather than true out of sample forecast errors (Bowerman and Koehler, 1989).

2.4.3 Further work

The development of a model with known accuracy allows for comparison of pre and post-intervention peak stage response of the stream. The next step is to use the models to simulate baseline peak stage for comparison to observations of peak stage after leaky dams are installed in the streams. Thereby impacts of leaky dams greater than the model error can be assessed for every event in the post-intervention monitoring period. Ideally, further work would be done to identify time series models which provide as good a fit to the data as was found for stream 1 on all three streams. This would allow high flow event peaks of the full range of magnitudes to be simulated with a high degree of accuracy on all three streams. However, the level of accuracy for the full range of stage on stream 1, and for all but the six largest events on the control stream is high enough to assess leaky dam impacts on stream 1 in Chapter 3.

2.5 Conclusion

Whilst previous research has focused on using either an empirical approach or a deterministic modelling approach to detect NFM impacts on downstream flood risk, this research demonstrates that where uncertainty masks the signal of the intervention, a top-down data-based time series modelling approach can provide the tools to make a meaningful comparison between empirical baseline and post-intervention data. Although different types of time series models may be necessary to demonstrate the full benefits of data-based time series modelling approaches, this study provides evidence that, where the underlying data generating processes are linear, TFN modelling can reproduce observed stage hydrographs to a high degree of accuracy. Given the upstream stage series, the baseline, pre-intervention response of a stream can therefore be accurately simulated for any chosen high flow event. This means that for every flood peak observed after an NFM intervention is made the baseline flood peak magnitude can be simulated for comparison to a high degree of confidence. The impact of the NFM intervention can thereby be quantified for the full range of events observed after the interventions are installed. Hereby it is demonstrated that, by using a data-based time series modelling approach, BACI data can be used to assess the impact of NFM features such as leaky dams on downstream flood peak magnitude, even when lead times to collect baseline data are short, the data are highly uncertain and comparable high flow events are not observed before and after the interventions are installed. Data-based time series modelling techniques, therefore, provide a promising solution to the problems associated with quantifying the flood risk management benefits of NFM interventions.

Chapter 3

Using a data-based modelling approach to assess leaky dam impacts on downstream flood risk *Part II Leaky dam impacts on flood peak magnitude*

3.1 Introduction

In recognition of the increasing flood risks posed by climate change and land-use change (Blöschl et al., 2019; Winsemius et al., 2016), policy in the UK and across Europe is moving towards a more holistic, landscape-scale approach to flood risk management (Commission of the European Communities, 2009). Working with natural processes to reduce flood risk, or natural flood management (NFM), is used alongside traditional engineered defences to restore processes which retain and slow water in the landscape (Forbes et al., 2015). NFM delivers multiple benefits including ecological, geomorphological and water quality improvements alongside aiming to reduce flood risk (Abbe and Brooks, 2011; Burgess-Gamble et al., 2017; Dadson et al., 2017; Lane, 2017). NFM may provide significant opportunities for flood risk management (Black et al., 2021; Dixon et al., 2016), however, the use of natural flood management as a mainstream management

approach is limited by a lack of confidence in the effects on downstream flood risk (Burgess-Gamble et al., 2017; Ellis et al., 2021). Quantifying the benefits of NFM for flood risk management is paramount to stakeholder buy-in (Bark et al., 2021) and for authorities to fulfil their obligations of cost-benefit analysis for the spending of public funds (Defra, 2009).

Natural flood management creates and restores features of catchments which slow, store or attenuate water in the landscape (Dadson et al., 2017). NFM includes measures such as improving the water retention capacity of soils, intercepting overland flows by planting buffer strips, and restoring the natural function of wood in rivers to slow fluvial flows (Yorkshire Dales National Park Authority et al., 2017). In the UK, flood risk management legislation requires that the inclusion of NFM measures is considered in all publicly funded flood and coastal erosion risk management strategies (Environment Agency, 2010a). Furthermore, following the UK's exit from the EU, the UK Government have proposed to replace the EU's Common Agricultural Policy (CAP) with the Environment Bill House of Commons (57, 2019-21 & 2021-22) and Agriculture Act (2020) which support integrated catchment management approaches such as NFM (Klaar et al., 2020). The Agriculture Act (2020) sets out the Environmental Land Management Scheme (ELMS) which aims to pay landowners directly for the provision of public goods, including flood risk management (Defra, 2020). Hence, whilst NFM has been evident in UK policy since 2005 (Defra, 2005), it is likely to become more mainstream if the UK is able to successfully implement its "public money for public goods" approach set out in the Agriculture Act (2020).

To be able to implement a payments for outcomes approach it is necessary to quantify the benefits delivered by NFM measures. Likewise, its use in publicly funded flood and coastal erosion management strategies requires the quantification of the performance of NFM interventions as flood risk management assets (Defra, 2009). Uptake of the approach depends on buy in of stakeholders (Wingfield et al., 2019) who see uncertainty in the efficacy of NFM measures as a key barrier to their implementation (Bark et al., 2021). Hence, the quantification of natural flood management impacts on flood peak magnitude and timing is paramount to support a paradigm shift in the way that flood risk and the environment are managed in the UK.

To generate quantitative evidence to demonstrate the effectiveness of working with natural processes to reduce flood risk, the UK government's Department for Environment and Rural Affairs (Defra) invested £1.7 million in three multi-objective flood management projects in 2009 (Moors for the Future Partnership, 2015). Two of these projects included the use of leaky dams in upland catchments (National Trust, 2015; Nisbet et al., 2015b). Leaky dams consist of one or more pieces of large wood installed across small streams perpendicular to the direction of flow. They typically extend onto the floodplain and are raised above baseflow to allow for passage of fish (Burgess-Gamble et al., 2017; Forbes et al., 2015). Although they are traditionally used for river restoration (Grabowski et al., 2019; Kail et al., 2007), they are increasingly being installed in rivers for the purpose of flood risk management (Grabowski et al., 2019). Leaky dams have been installed in upland streams across the UK (National Trust, 2015; Nisbet et al., 2011; Slow the Flow Calderdale, 2017) and can range from a small number of interventions placed in one stream to hundreds of interventions spread across a catchment (The Rivers Trust, 2021). In England and Wales, their installation for the purpose of flood risk management in uplands has been incentivised through agricultural subsidies (Defra et al., 2016).

Leaky dams are thought to delay the flood peak and reduce its magnitude because they increase in-channel roughness, reduce flow velocity and increase floodplain connectivity (Abbe and Brooks, 2011; Gregory et al., 1985; Keys et al., 2018). Leaky dams differ from the type of in-stream wood typically installed for river restoration in that they are designed to maximise interaction with the flood peak by spanning the channel and extending onto the floodplain to increase storage potential, blockage ratio and overbank routing/roughness gains (Woodland Trust, 2016). By being raised above baseflow, their storage capacity is not depleted under baseflow conditions so that it is available during flood events, and by being perpendicular to the direction of flow they maximise friction effects (Gippel et al., 1996). Their impacts on flood peak magnitude are therefore expected to be greater than those of in-stream wood installed for the purpose of river restoration.

Leaky dams have been shown to delay flood peak timing at a range of spatial scales (Black et al., 2021; Gregory et al., 1985; Kitts, 2010) but their impact on downstream flood peak magnitude remains unquantified. The impact on flood

peak magnitude of in-stream wood placed longitudinally on the channel bed, however, has been measured in an upland test catchment in the mid-Atlantic region of the United States (Keys et al., 2018), and in the Ore Mountains of Germany (Wenzel et al., 2014). In both cases the impacts of the in-stream wood on the magnitude of an artificial flood wave, generated using reservoir releases, was measured. In the US, the magnitude of a <1-in-1 year flood event was reduced by 8% (Keys et al., 2018), whilst a 1-in-3.5 year flood event was reduced by 2.2% in the Ore Mountains (Wenzel et al., 2014). Because the studies tested different event magnitudes it is unclear whether the results varied due to the event magnitude, or because of site or wood specific factors, such as the channel geomorphology or the size of the wood relative to the channel.

Modelling of leaky dam schemes at the catchment scale suggests that leaky dams in tributaries with a high gradient (>0.005 m/m) have little impact on event peak magnitude (Dixon et al., 2016; Thomas and Nisbet, 2012) because hydraulic resistance has less effect on flood waves in steep rivers (Sholtes and Doyle, 2011), and the available storage volume sharply declines (Thomas and Nisbet, 2012). However, confidence in the results of modelling studies is low because the representation of leaky dams is usually not validated (Addy and Wilkinson, 2019). To increase confidence in the impacts of leaky dams on event peak magnitude, empirical data from a range of environments is needed, in a similar manner to traditional engineered defences (Ellis et al., 2021).

Applying a Before After Control Impact (BACI) monitoring methodology to the collection of hydrological data has been identified as a way to gather the evidence required to mainstream NFM approaches (Ellis et al., 2021). However, the BACI approach alone is unlikely to overcome the difficulties associated with quantifying the impacts of leaky dams and other NFM measures. Difficulties in isolating impacts of one type of measure, variation in a measure's effectiveness with event magnitude, insufficiently long monitoring timescales, and complexities of context and scale of interventions all contribute to the uncertainty in quantifying their impacts (Connelly et al., 2020). For example, of the three Defra-funded multi-objective flood management demonstration projects initiated in 2009 two reported that a lack of comparable high flow events during the monitoring period precluded the assessment of NFM impacts, despite the use of a BACI style

monitoring approach (National Trust, 2015; Nisbet et al., 2015b). This issue continues to hamper efforts to collect evidence from the £15 million of NFM projects funded by the UK government in 2017 (Environment Agency, 2019) and presents a common problem as many NFM studies are based on short periods of observation (Connelly et al., 2020). Additionally, even when sufficient flood events are recorded, the signal of an intervention can be masked by the high levels of uncertainty typical of hydrological data (Black et al., 2021; Gebrehiwot et al., 2019; Lane, 2017; Wilkinson et al., 2014).

The stochastic nature of flood events further compounds the comparison of impacts between sites and catchments as they have varying characteristics (Wingfield et al., 2019). Impacts on different types of events, such as typically intense summer and long winter storms could have important economic implications for flooding of agricultural land which is most vulnerable during the summer months (Morris and Brewin, 2014; Posthumus et al., 2009). Furthermore, modelling suggests that considering the impacts of multi-peaked events on leaky dams can lead to dramatically different conclusions due to the time required for the system to recover its effectiveness between peaks (Metcalfe et al., 2017). To manage expectations of NFM efficacy, which has been identified repeatedly as key to sustaining efforts to mainstream NFM (Collentine and Futter, 2018; Nisbet et al., 2015b; Wells et al., 2020), a clearer picture is needed of the impacts on single and multi-peaked events with a range of peak magnitudes and event characteristics.

In light of the difficulties associated with collecting empirical evidence, researchers have employed fluvial hydraulic based models to gain a better understanding of leaky dam impacts (Burgess-Gamble et al., 2017). These modelling studies have been carried out using a 'bottom up' or 'reductionist' approach in which subjective *a priori* assumptions about model structure and parameters are made based on academic judgement and intuition (Young, 2003). Young (1978) argued that such *a priori* conceptions of how the system works can lead to overconfidence in the resulting models. This is a particularly pertinent point when considering the impacts of NFM features because the physical basis of the processes governing their impacts are poorly understood (Lane, 2017) and there is little evidence based guidance as to the representation of leaky dams in such models (Addy and Wilkinson, 2019). Instead, a different approach to quantifying

leaky dam impacts on flood peak magnitude is suggested here in which empirical data, collected using a BACI design, is combined with a data-based time series modelling approach.

Data-based time series modelling is a top-down approach which does not rely on *a priori* assumptions about the model, but instead allows the data to guide the form of the model. Because the type of model is chosen based on the statistical properties of the data they can handle data with different spectral properties (distribution of variance with frequency) and undesirable features such as non-stationarity, non-constant variance and seasonality which confound change detection (Hipel and McLeod, 1994). They can be trained using short periods of data (Young, 2003) which means they do not rely on long periods of baseline data to be collected before leaky dams are installed. Once the model structure and parameters have been determined from the data, the model represents the dynamical properties of the system and can be used to make predictions of the values of the output series and its uncertainty for unobserved periods (von Asmuth et al., 2002). Hence, a data-based time series model can be trained to simulate the downstream response of a stream without leaky dams to an upstream flood hydrograph. These simulations can be compared to observations of the stream response after leaky dams have been installed to assess the difference the leaky dams have made to the stream response.

The approach of comparing simulated time series to observed time series to quantify the effects of a treatment has been successfully applied to assess the temperature response of a headwater stream in British Columbia, Canada to changes in riparian vegetation (Gomi et al., 2006), changes in thermal response of alpine rivers brought about by anthropogenic flow regulation (Dickson et al., 2012), and to quantify the effect of different forest treatments on streamflow in forested catchments in south-eastern Australia (Watson et al., 2001). Although accurate models of level to level response have been developed for rivers in the UK (Beven et al., 2008; Romanowicz et al., 2008; Young, 2003), the approach has not yet been applied to the assessment of leaky dam impacts on downstream flood peak magnitude. The results of Chapter 2 showed that the peak stage of high flow events recorded during the baseline monitoring period of a BACI style study could be simulated to within ±2 centimetres at the 80% confidence interval for an

impact and a control stream. In this chapter, the data-based time series models developed in Chapter 2 were used to simulate what peak stage would have been, with two centimetre accuracy, had leaky dams not been installed in the impact stream. Differences >2 cm were attributed to the effect of the leaky dams on peak stage. Leaky dam effects were quantified using this approach, for the first time, for events with a variety of peak magnitudes, durations, total volumes, and numbers of event peaks. This chapter addresses the second research question posed in the introduction:

Q: In upland streams, what is the impact of leaky dams on the flood peak magnitude of a range of flood events?

by addressing the objectives:

O1: Compare and calculate differences between simulations of baseline stage and observations of post-intervention stage at the peak of all flood events observed during the post-intervention period on the impact and control streams.

O2: Determine the statistical significance of the differences between simulated and observed peak stage.

O3: Assess the variability of the effectiveness of leaky dams during different types of high flow events, including single and multi-peaked events.

3.2 Methods

A semi empirical data-based time series modelling approach was taken, after Dickson et al. (2012); Gomi et al. (2006); Watson et al. (2001) & O'Driscoll et al. (2016), to assess the impact of installing eight leaky dams in an upland, headwater stream in Coverdale, England, on the peak magnitude and timing of high flow events. The study stream was one of five streams which were part of a larger monitoring effort to empirically quantify the impact of leaky dams using four impact streams and one control stream with a BACI experimental design. The stage series were found to be too uncertain to directly assess leaky dams impacts

on flood peak magnitude and timing (Appendix B) and therefore the data-based time series approach summarised in Figure 3.1 was taken to assess the difference between baseline and post-intervention reach response on an impact stream and the control stream.

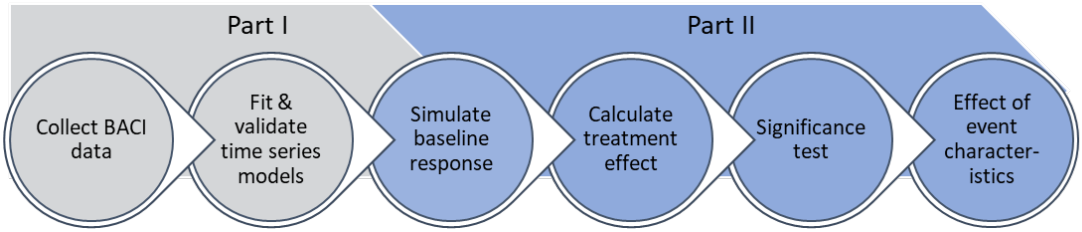
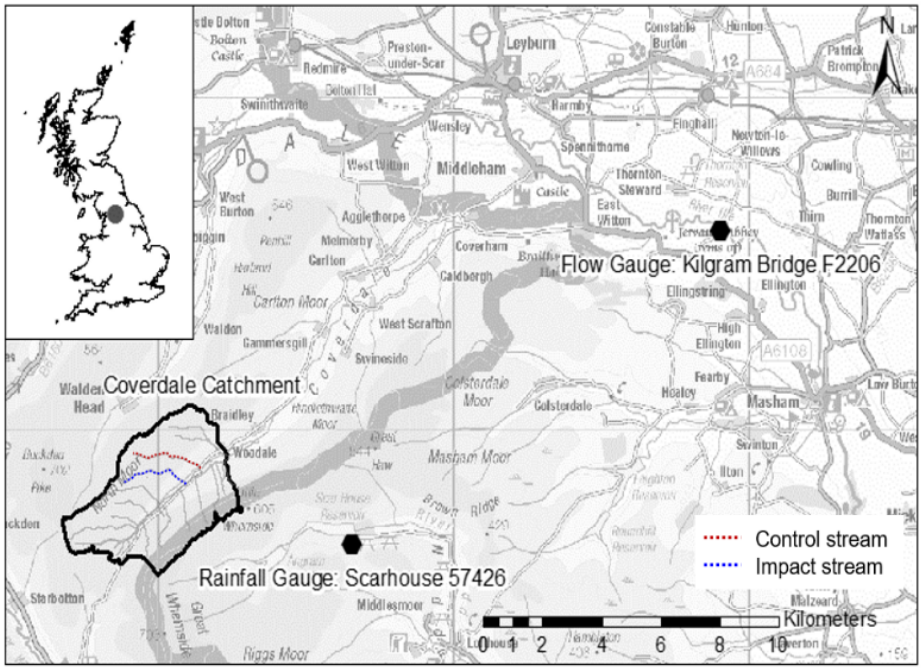


Figure 3.1: Data-based time series analysis approach used within this study

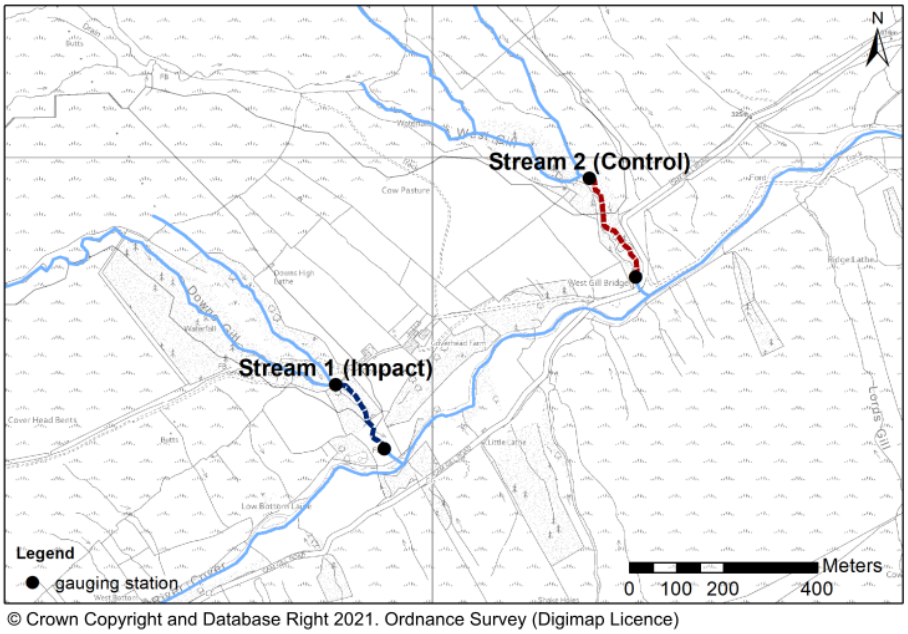
As summarised in Figure 3.1, the models were fitted and validated in the previous chapter (Part I, Chapter 2), in this chapter (Part II) the models were used to assess the impact of leaky dams on the peak magnitude of the flood hydrograph. Following fitting and validation of the data-based time series models outlined in Chapter 2 (Part I), the following steps were taken to assess the impacts of leaky dams on event peak magnitude: (1) Identification of high flow events in upstream, post-intervention stage series; (2) Simulation of downstream baseline stage response to upstream event stage time series; (3) Calculation of differences between simulated and observed downstream stage for both baseline and post-intervention monitoring periods on the impact and control stream ;(4) Testing for statistical differences between the simulated and observed response in the baseline and post-intervention periods; (5) Analysis of the impact of high flow event characteristics such as peak magnitude, duration and the number of peaks in the event on leaky dam effectiveness.

3.2.1 Field Methods

The study site was located in the headwaters of the River Cover (54.20045 N, -1.98617 E) on the Eastern flank of the Yorkshire Dales National Park, North Yorkshire, England (Figure 3.2a). The climate is cool and wet, with an average annual rainfall of 1270 mm (EA rain gauging station 57426 data 1988-2018). The



(a)



(b)

Figure 3.2: (a) Location of study site in Coverdale, North Yorkshire, UK, and nearest Environment Agency operated flow and rainfall gauging stations (b) Water level/flow gauging network in Coverdale catchment

glacial valley of Coverdale overlies Great Scar limestone with rocks of the Yoredale series forming the valley sides (Yorkshire Dales National Park Authority, 2002). Land use in the catchment is pastoral agriculture on open, unimproved grassland whilst the moorland is managed for grouse shooting. The headwaters are formed of many small, parallel watercourses which flow into the River Cover at an altitude of 400 m AOD. This study focuses on two of these watercourses; an impact stream (Stream 1) and a control stream (Stream 2), shown in Figure 3.2b, with similar hydrological characteristics (Table 3.1) and no established lateral inflows within the monitored reach. The watercourses are of type A in the Rosgen classification; steep, partially entrenched and cascading with step/pool streams (Rosgen, 1994).

Table 3.1: Characteristics of the study streams

Stream	Gradient (m/m)	Catchment Area (km^2)	Monitored length (m)	Mean width (m)	leaky dams (count)
1 (impact)	0.13	1.1	280	2.6	8
2 (control)	0.11	1.9	260	3.0	0

River stage was recorded at the upstream and downstream extent of both streams from September 2017 to February 2020 at one-minute intervals using In-Situ Rugged TROLL 100 (Redditch, UK) non-vented pressure transducers with 0.05% full scale accuracy (± 0.0045 m) in stilling wells corrected using an In-Situ Rugged BaroTROLL (Redditch, UK) atmospheric pressure gauge (± 0.05% full scale accuracy). Equipment failure resulted in some data gaps on the impact stream, particularly a 72 day gap in summer 2018 (baseline) and the data from both streams contained considerable uncertainty due to instabilities in the stage datum at the gauges. Rating relationships were developed for each of the gauging sites by calibrating 1D hydraulic models of the sites to measurements of stage-discharge pairs at each site (Appendix A). Discharge was measured during a range of high flow events using slug-injection dilution methods (Moore, 2005), with salt pulses recorded using conductivity as a proxy for concentration at one second intervals using a Campbell Scientific CR200 Data logger and conductivity probe.

There were no long-term flow or level gauging records available for the site; however, an Environment Agency maintained flow gauging station is in operation

at Kilgram Bridge (station number F2206), on the River Ure, 6 km downstream of its confluence with the River Cover (Figure 3.2a). The figure also shows the location of EA rainfall gauging station 7426 at Scar House (54.184828 N, -1.900625 E), 6 km east of the study site, from which 15-minute rainfall time series are available from 1988 onwards. Peak Over Threshold (POT) and Annual Maximum (AMAX) series from Kilgram Bridge station F2206 were obtained from the National River Flow Archive (UK Centre for Ecology and Hydrology, 2021). The AMAX and POT series were used together with Met Office named storms (Met Office, 2021) to contextualise the events observed during the study period.

(a) **(b)**

Figure 3.3: Photographs of (a) elevation and (b) plan view of typical leaky dams on the impact stream

The leaky dams, shown in Figure 3.3, were built following the guidance of the Yorkshire Dales Rivers Trust (2018) which was developed by local NFM practitioners based on experience from catchments with similar hydraulic conditions and site constraints. The design of the dams is shown in Figure 3.4 and the average as-built dimensions of the dams are given in Table 3.2. The dams were built from 2-5 locally felled tree stems with a minimum length of 1.5 times channel width. The stems were installed to span the channel perpendicular to the direction of flow. The dams had an average height of 0.8 m above the riverbed and were installed to provide approximately 0.3 m clearance from baseflow for fish passage. The dams were anchored using strainer posts and wire and brash was

used to stuff the dams to enhance biodiversity impacts. To fulfil the consenting requirements of the lead local flood authority a flood risk impact assessment was carried out which considered the risk to downstream assets of structure blockage and flood wave surge in case of failure.

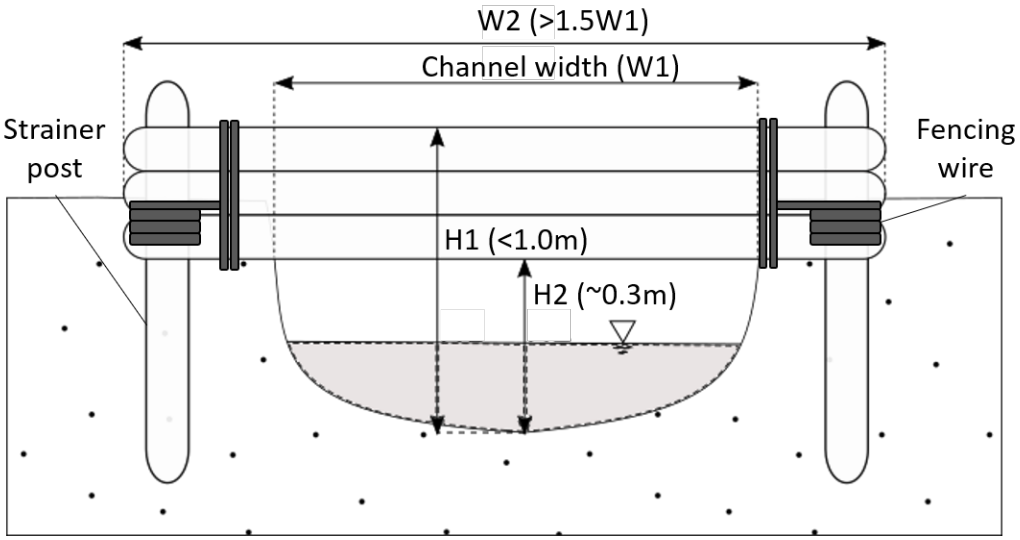


Figure 3.4: Diagram of leaky dam dimensions (referred to in Table 3.2)

Table 3.2: Average as-built dimensions of leaky dams installed in the impact stream, dimensions refer to Figure 3.4

Stream	1
Count of dams	8
Average of W1 (m)	4.58
Average of W2 (m)	2.61
Average of H1 (m)	0.86
Average of H2 (m)	0.3
Average dam spacing (m)	25

Decisions about the placement of leaky dams were made by balancing site constraints such as material availability and site accessibility with water storage potential arising from the site topography. Three types of flood water storage

mechanisms were identified and opportunities were sought in the following priority order, designed to maximise flood storage volume: (1) Increased flood-plain connectivity with opportunities for re-routing of flood water to offline storage areas; (2) Increased flood-plain connectivity with in-line (floodplain) storage areas; (3) Increased in-channel storage. Figure 3.5 shows the location of the eight leaky dams on the impact stream.

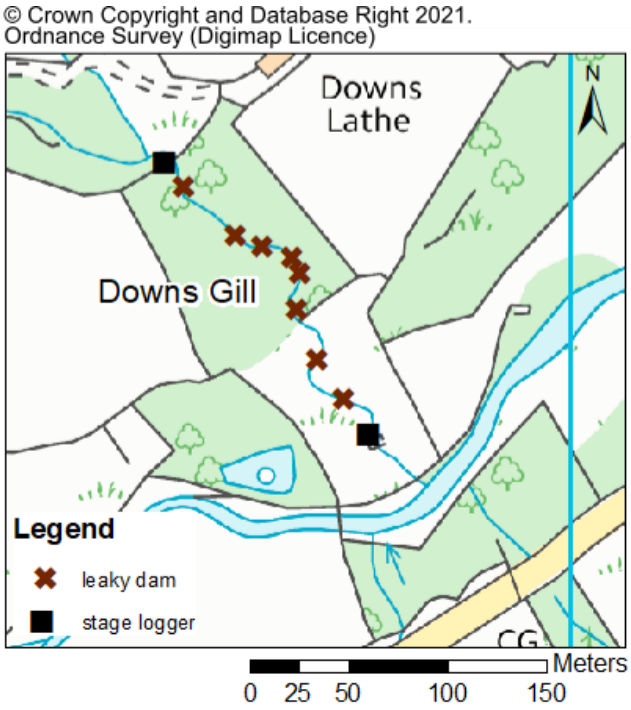


Figure 3.5: Location of leaky dams and stage gauges in the impact stream

3.2.2 Data Analysis

High flow events

High flow events were identified from the post-intervention stage time series using a similar rules based methodology to Deasy et al. (2009) & Glendell et al. (2014) as described in Chapter 2 Section 2.2.6. The following two criteria were used to identify discrete high flow events: (1) A stage peak was considered a discrete high flow event if it was of duration >60 minutes and the upstream peak stage exceeded

the mean stage recorded on the stream during the monitoring period (mean stage was 0.04 m on the impact stream and 0.15 m on the control stream); (2) Events were classed as independent if they were separated by at least 15 minutes of stage below baseflow stage +10%. Baseflow stage determined that start and end points of the events and was calculated, after Bezak et al. (2015), using the methods described in the World Meteorological Organisation's manual on low-flow estimation and prediction (WMO, 2009) using the accompanying R package 'lfstat' v. 0.9.4 (Koffler et al., 2016). The Hydrological Model Assessment and Development (HydroMAD) v. 0.9-26 package (Andrews and Guillaume, 2018) for the statistical computing environment R was used to identify discrete storm events using the above criteria. High flow events were permitted to be single or multi-peaked.

Characteristics describing the hydrograph were calculated for each of the discrete storm events. The metrics used to describe the magnitude of the event were peak stage, S_p, duration, D, and total stage, S_t. The peak stage was defined as the maximum stage between the start and end of the event and was therefore based on the largest peak of multi-peaked events. Event duration was calculated as the time from the start to the end of the event and total stage was the sum of stage from the start to the end of the event, which was used as a proxy for event volume. After Potter (1991) the hydrograph rise time, T_{rise}, defined as the time between the start and peak of the event, was calculated to describe the peakedness of the event. The time since the previous event, D_a, was also calculated for each event. Correlations between the characteristics were assessed using Spearman's rank order correlation (Spearman, 1904).

Event simulations

Transfer function noise models were fitted to the stage time series collected in the baseline, pre-intervention monitoring period on the impact and control stream. Fitting and validation of the models is described in detail in Chapter 2. In summary, the models, given by Equation 3.1 and 3.2, were fitted to the baseline stage time series on each stream using the upstream stage series as the predictor variable and the downstream stage series as the forecast variable. The stage

time series were aggregated from 1-minute to 15-minute temporal resolution by taking the maximum stage in each 15-minute timestep. Both the upstream and downstream stage series were transformed to meet the requirements of stationarity by first order differencing. The fitted and validated models consisted of a dynamic regression and ARMA noise component, the parsimonious form and coefficients of the models were inferred from the data as described in Chapter 2. Blocked out of sample cross-validation showed that the models were able to simulate high flow event peak stage to within ± 0.02m at the 80% confidence level. In this chapter, the models in Equation 3.1 and 3.2, were used to simulate downstream pre-intervention stage for high flow events in the post-intervention monitoring period. The simulated pre-intervention stage was used as a basis for comparison to observed post-intervention stage to assess differences in the pre- and post-intervention response of the stream to high flows.

Simulations of baseline stage were made for every discrete high flow event in the post-intervention monitoring period for the impact and control stream respectively. The forecast variable was downstream stage, D_t^*, at time t, transformed to be stationary by first order differencing. The upstream stage series at time t formed the predictor variable, U_t^*, which was also transformed by first order differencing. N_t was the autocorrelated noise term which was modelled using the ARMA model, and a_t was a random noise term which was independently and identically distributed. The package 'forecast' v.8.12 (Hyndman and Khandakar, 2008) was used in the statistical software programme, R v. 4.0.2 (R Core Team, 2020), to make simulations using the TFN models. The simulations were made for 20-hour windows of upstream stage data at 15-minute timesteps centred on the peak of the event as illustrated in Figure 3.6.

Impact Stream (Stream 1):

$$\begin{aligned} D_t^* &= 0.687U_t^* + 0.332U_{t-1}^* + 0.165U_{t-2}^* + 0.102U_{t-3}^* - 0.014U_{t-11}^* \\ &\quad -0.01U_{t-20}^* + N_t \\ N_t &= 0.893N_{t-1} + 0.179N_{t-2} - 0.437N_{t-3} + 0.288N_{t-4} - 0.162N_{t-5} \\ &\quad -0.554a_{t-1} - 0.654a_{t-2} + 0.453a_{t-3} + a_t \end{aligned} \tag{3.1}$$

Control Stream (Stream 2):

$$\begin{aligned} D_t^* &= 0.430U_t^* + 0.198U_{t-1}^* + 0.114U_{t-2}^* + 0.083U_{t-3}^* \\ &\quad +0.042U_{t-4}^* + N_t \\ N_t &= 1.035N_{t-1} - 0.196N_{t-2} + 0.173N_{t-3} - 0.062N_{t-4} - 0.739a_{t-1} \\ &\quad -0.253a_{t-2} + a_t \end{aligned} \tag{3.2}$$

Simulations of baseline peak stage and empirical prediction intervals were taken from the blocked out of sample cross validation procedure described in Chapter 2 Section 2.2.6. This approach was taken so that the model prediction intervals were based on data which was not used to train the model.

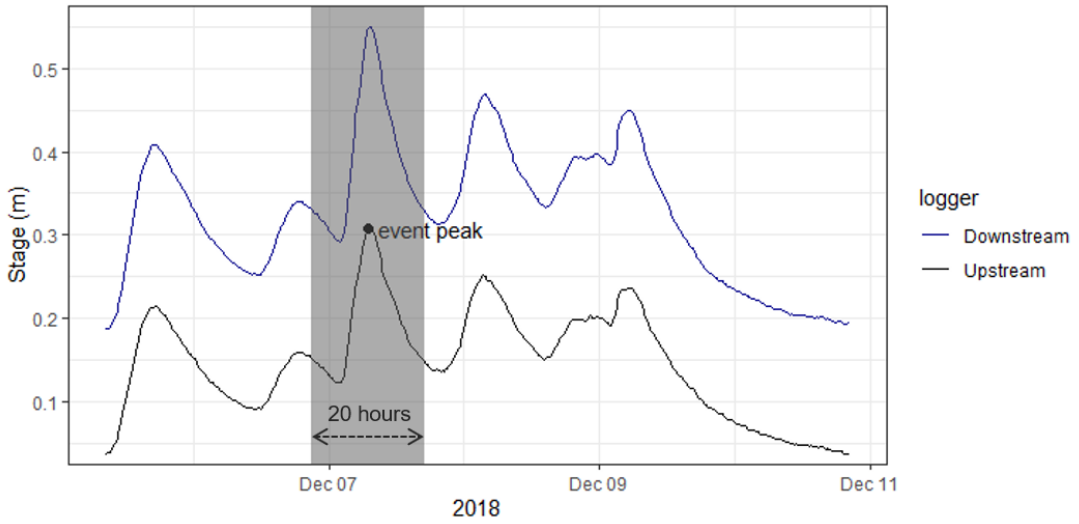


Figure 3.6: Example of a 20-hour simulation window centred on the peak of an event

Treatment Effect

The difference in peak stage with and without leaky dams was assessed by calculating the difference between the simulated peak stage and the observed peak stage (Dickson et al., 2012; Gomi et al., 2006) and was called the treatment effect, (T_e),

$$T_e = D_{peak} - \hat{D}_{peak} \tag{3.3}$$

where D_{peak} is the observed downstream stage in metres at the peak of the event (i.e., with leaky dams), and $\hat{D}_{peak}$ is the simulated baseline downstream stage in metres at the peak of the event (i.e., without leaky dams). To aid in comparability between sites and studies the treatment effect was also expressed in terms of the percentage change in peak discharge, T_{eQ}. The simulated and observed peak stage were converted to discharge, Q_{peak} and $\hat{Q}_{peak}$, respectively using the rating relationships developed in Appendix A. The treatment effect was calculated as a percentage difference in event peak discharge,

$$T_{eQ} = 100(\frac{(Q_{peak} - \hat{Q}_{peak})}{Q_{peak}}) \tag{3.4}$$

Although expressing treatment effect in terms of percentage reduction in peak discharge allowed the treatment effect to be compared to other sites, it introduced uncertainties associated with the rating relationship and so statistical significance testing and investigation of variation in treatment effect was conducted in terms of peak stage alone. The model's empirical prediction intervals were used to assess whether the treatment effect was greater, lesser or within the same magnitude as the expected model error at the peak of the event. The prediction intervals were estimated in Chapter 2 based on the distribution of peak stage error in the out of sample blocked cross-validation and were estimated at the 95% and 80% confidence levels. The 95% level of confidence interval was used because it is the recommended Hydrometric Uncertainty Guidance ISO standard (ISO/TS 25377:2007) and the 80% prediction interval is recommended for prediction intervals empirically estimated from data with outliers or tail problems (Chatfield, 2001).

As summarised in Table 3.3, a positive treatment effect, where T_e is greater than the upper model prediction interval, indicated that the peak magnitude was reduced in the post-intervention monitoring period compared to pre-intervention response of the stream. Treatment effect below the lower model prediction interval indicated an increase in peak magnitude in the post-intervention monitoring period, and treatment effect between the prediction intervals indicated that the difference in pre and post-intervention peak stage response was too small to distinguish from model error. On the control stream a significantly negative or

positive T_e indicated that influences other than the leaky dams affected the peak stage in the post-intervention monitoring period.

Table 3.3: Classification of treatment effect (PI = Prediction Interval)

$\mathbf{T_e >}$ **Upper PI**	Positive T_e :	Peak magnitude **reduced** beyond expected model error in post-intervention monitoring period
$\mathbf{T_e <}$ **Lower PI**	Negative T_e :	Peak magnitude **increased** beyond expected model error in post-intervention monitoring period
Lower PI $\mathbf{\leq T_e \geq}$ **Upper PI**	Insignificant T_e :	Change in peak magnitude **too small** to distinguish from expected model error

Statistical test for significance

Like Gomi et al. (2006) & Dickson et al. (2012) an approximate assessment of statistical significance was given by adapting the methods of Watson et al. (2001). Because each high flow event for which T_e was calculated was assumed to be independent, every value of T_e was assumed to be independent and can be described as the model error at the peak of the event. In the previous chapter this concept was used to estimate the empirical prediction intervals of the model simulations at the peak of the event. The null hypothesis of 'no treatment effect' would be accepted if the distributions of the disturbances, or treatment effect, in the pre- and post-intervention monitoring periods were the same. To test for significance in the difference between the distribution of treatment effect in the pre-and post-intervention monitoring periods the non-parametric two-sample Kolmogorov-Smirnov test was applied (Dickson et al., 2012; Gomi et al., 2006). In the impact stream rejecting the null hypothesis of 'no treatment effect' would indicate that downstream peak magnitude was significantly different in the post-intervention monitoring period. In the control stream accepting the null hypothesis of 'no treatment effect' would increase confidence in attributing this difference to the leaky dams installed in the impact stream, rather than external impacts.

Variation in treatment effect

To obtain an initial indication of whether treatment effect varied between seasons boxplots of treatment effect were drawn grouped into seasons by the month during which the event peak occurred (spring: March-May, summer: June-August, autumn: September-November, winter: December-February). A statistical test of differences was not performed because of the low number of events in some seasons.

Association between the percentage change in peak discharge (treatment effect) and event peak magnitude was assessed by calculating the Pearson's product moment correlation (Freedman et al., 2007). Normality of the distribution of each of the characteristics was checked using the Shapiro-Wilk test (appropriate for small samples) in conjunction with QQ normality plots. The null hypothesis was that treatment effect was not correlated with each of the event characteristics. Where a significant association between treatment effect and peak magnitude was found, and residuals were normally distributed, a linear regression was performed to assess the form and strength of the relationship. The significance of the relationship was assessed by testing whether the slope of the regression line was significantly different from zero using the two-sided t-test at the 0.05 significance level. The null hypothesis was that the slope of the regression line was equal to zero. The coefficient of determination (R^2) was used to assess what proportion of the variation of treatment effect which was explained by the event characteristic. The regression relationship and its 95% prediction intervals were used to make predictions of treatment effect dependent on peak magnitude. The analysis was performed using R v. 4.0.2 (R Core Team, 2020).

Event characteristics other than peak discharge were not normally distributed and did not display a linear relationship with treatment effect, therefore, correlations for these characteristics were not calculated. Instead, differences in the median event characteristics of events which had a positive, negative or insignificant treatment effect according to the definitions in Table 3.3 were assessed. First, the events were grouped depending on whether there was a positive, negative, or insignificant treatment effect on the impact stream. The characteristics of the events in each group were then compared using boxplots to give an indication as

to whether leaky dams were more effective during some types of events than others. Differences in the median of the groups were tested for statistical significance using the non-parametric Mann Whitney U test (Mann and Whitney, 1947). The characteristics tested were peak stage (S_p) in metres, total stage (S_t) in metres, event duration (D) in hours, time since previous event, (D_a), in hours, and after Potter (1991), hydrograph rise time, T_{rise}, was used to describe the duration of the rising limb of the event. The number of days since the interventions were installed, T_{int} was also assessed.

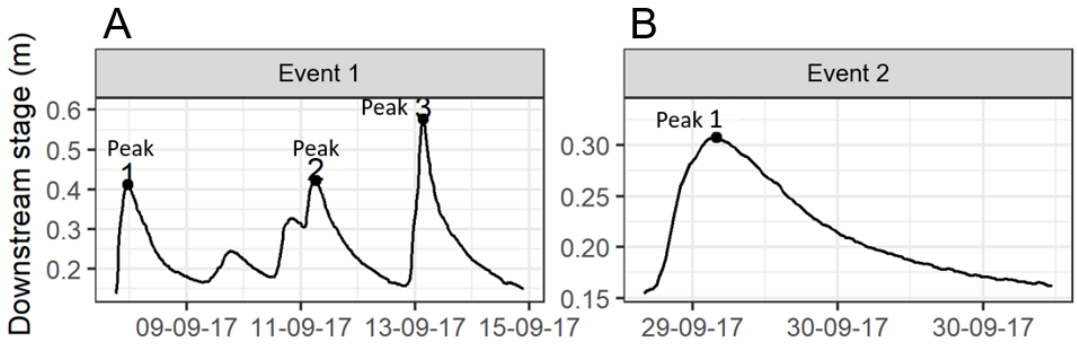


Figure 3.7: Example events to illustrate numbering of event peaks

The effect of leaky dams on event peak magnitude is thought to differ depending on whether a peak is the first peak of an event (including single-peaked events) and subsequent peaks of multi-peaked events (Metcalfe et al., 2017). As numerous multi-peaked events were observed in the post-intervention monitoring period boxplots were used to examine the effect of peak order on treatment effect. As illustrated in Figure 3.7, event hydrographs could be multi-peaked (Figure 3.7A) or single-peaked (Figure 3.7B). To examine whether the treatment effect varied between the first peak in an event and subsequent peaks, T_e was calculated for each peak in multi-peaked events and grouped by peak order; determined by whether they were the first (including single peaked events), second, third, or subsequent peak within an event, as illustrated in Figure 3.7. Individual peaks were deemed peaks of the same event if the stage fell below the level of the leaky dams (0.3 m) between peaks but did not return to within 10% of baseflow. Because the errors of multiple step ahead forecasts are additive, each peak was

simulated at an equal number of time steps (10 hours) from the beginning of the simulation to assess T_e. For example, to calculate the treatment effect for all three peaks in Figure 3.7A, the simulation was repeated three times, centred on each peak in turn. This process was repeated for all events in the post-intervention monitoring period. T_e was calculated for every peak and grouped according to whether they were the first, second, third or subsequent peak of the event and differences in the median value of T_e between the groups was assessed using the Mann-Whitney-U test (Mann and Whitney, 1947).

3.3 Results

3.3.1 Characterisation of hydrological events

The post-intervention monitoring period was generally wetter than the baseline monitoring period, with higher total and annual maximum daily rainfalls recorded at the nearest Environment Agency (EA) operated rainfall gauge (Table 3.4). The annual maximum (AMAX) flows recorded at a downstream flow gauge during the post-intervention monitoring period were the third and ninth largest events since records began in 1966. In total, the discharge exceeded the peak over threshold (POT) for the gauge 10 times during the post-intervention monitoring period, compared to three times during the baseline monitoring period.

Table 3.4: Climate data at Scar House and Kilgram gauge (*to March 2020), water year is defined as 1st October to 30th September

	Baseline		**Post-intervention**	
Water Year	2016-2017	2017-2018	2018-2019	2019-2020
Total precipitation (mm)	881	940	1129	755*
max daily rainfall (mm)	35.1	34.8	38.9	54.5*
AMAX flow (m^3/s)	3.1	3.9	4.7	5.4
AMAX rank	48	30	9	3
Number of POT events	1	2	5	5

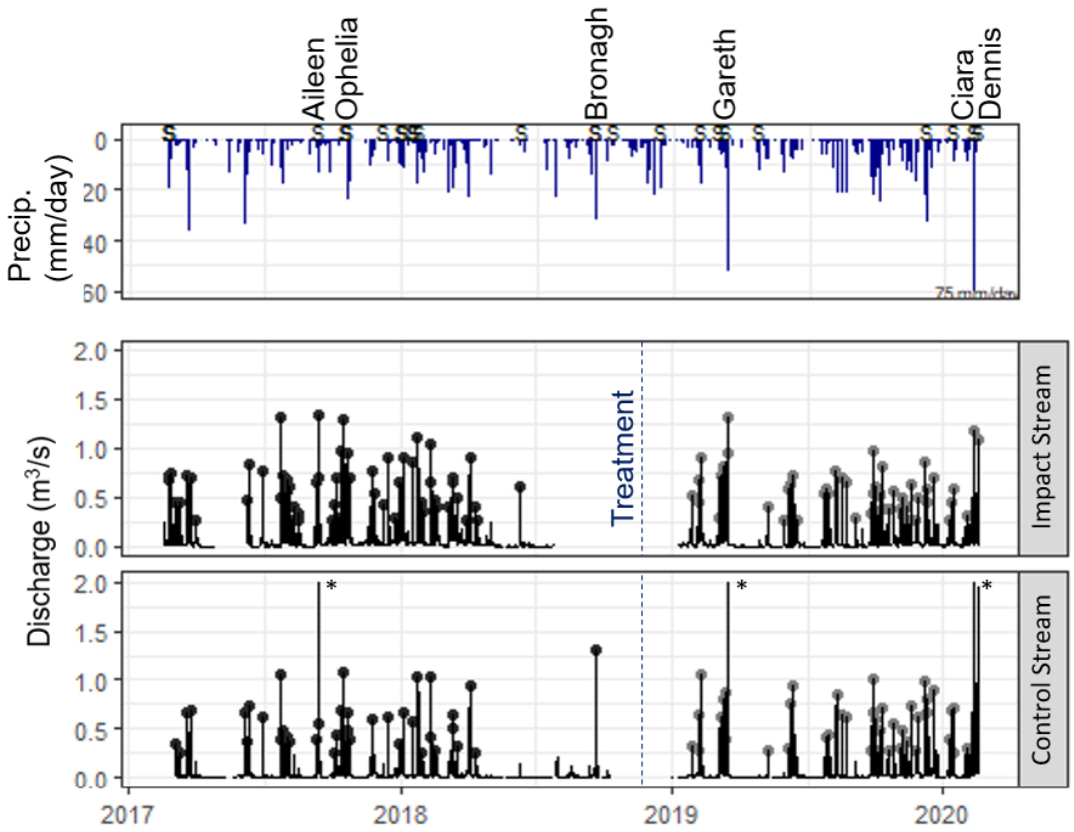


Figure 3.8: Study site rainfall (EA gauge 7426) and discharge recorded at upstream extent of study streams, 'S' indicates storms named by the UK Met Office, and those which affected the study site are labelled along the upper margin. Points indicate high flow events identified in the data, asterisks indicate discharge $>2\ m^3/s$ which are likely erroneous (Appendix B)

Table 3.5: Top ten high flow events recorded on the impact stream during the monitoring period based on peak stage. The numbers in brackets indicate the rank of the equivalent POT event on the downstream Kilgram gauge, where relevant. Total rainfall indicates the total rainfall recorded in the 72 hours up to and including the event date

Rank	Date	Storm Name	Impact Peak Stage (m)	Control Peak Stage (m)	Kilgram Peak Flow (m^3/s)	Scar House Total Rainfall (mm)
Baseline monitoring period						
1	13/09/2017	Aileen	0.58	0.50	171 (226)	35.4
2	14/10/2017	Ophelia	0.54	0.45	173 (222)	42.2
3	08/02/2018		0.52	0.40	104	13.0
4	23/01/2018	Georgina	0.51	0.41	159	31.4
5	03/04/2018		0.50	0.40	130	36.2
6	11/10/2017		0.49	0.36	162	27.2
7	21/10/2017	Brian	0.48	0.38	103	50
8	13/12/2017		0.48	0.37	92	21.2
9	03/01/2018	Eleanor	0.47	0.39	131	50.8
10	15/01/2018	Fionn	0.47	0.38	118	24.0
Post-intervention monitoring period						
1	09/02/2020	Ciara	0.63	0.62	361 (3)	113.0
2	15/02/2020	Dennis	0.60	0.73	292 (14)	56.2
3	16/03/2019	Gareth	0.58	0.59	303 (10)	72.4
4	18/12/2018	Deirdre	0.55	0.42	118	41.4
5	07/12/2018		0.55	0.38	130	37.2
6	09/02/2019	Erik	0.54	0.40	165 (251)	55.2
7	29/09/2019		0.53	0.44	179 (192)	76
8	05/12/2019	Atiyah	0.52	0.39	121	22.8
9	09/08/2019		0.51	0.42	174 (215)	36.8
10	19/12/2019		0.51	0.42	100	27

The ten events with the highest peak stage observed during the baseline and post-intervention monitoring periods on the impact and control stream are given

in Table 3.5. As expected from the downstream EA operated gauges, the largest events in the study streams were recorded during the post-intervention monitoring period, and median peak magnitude was higher overall (Figure 3.8). Rainfall in the UK is higher in the autumn and winter months (Met Office, 2020) and this is reflected in the seasonal differences in the distribution of event peak magnitude. The median peak magnitude was similar across seasons, with the exception of the summer of 2018 in the baseline monitoring period, which was exceptionally dry (Met Office, 2018), and the spring of the post-intervention monitoring period in which only five events were recorded on the control stream, although this included Storm Gareth (3 March 2019) which caused regional flooding (Flood List, 2020).

Seasonal variation in event duration (Figure 3.9B) and total stage (Figure 3.9C) were evident in the data from the control stream. Although peak stage (Figure 3.9A) was correlated with event duration and total stage (r >0.7) the same seasonal pattern was not evident in peak stage. Rising limb duration (D) was lowest in summer and highest in autumn. Furthermore, median peak stage recorded in each season was similar in the baseline and post-intervention monitoring periods, with the exception of the summer of 2018 in the baseline monitoring period, which was exceptionally dry (Met Office, 2018). Median event duration in each season differed between the baseline and post-intervention monitoring periods; in spring and winter events had a longer median duration in the baseline monitoring period, whilst in summer and autumn events with a longer median duration were recorded in the post-intervention period. Total stage differed in the same way between the baseline and post-intervention monitoring periods. The rising limb duration was lowest in summer and differed between the baseline and post-intervention monitoring periods, particularly in the autumn months.

Overall, spring events had a lower peak stage, event duration, total stage, and rising limb duration in the post-intervention monitoring period. Summer events had a higher peak stage, event duration and total stage in the post-intervention period, but rising limb duration was similar. Autumn events had a higher event duration and rising limb duration in the post-intervention monitoring period but peak stage and total stage were similar to the baseline monitoring period. Winter events had a lower event duration, total stage, and rising limb duration in the post-intervention monitoring period but peak stage was similar. Event peak stage

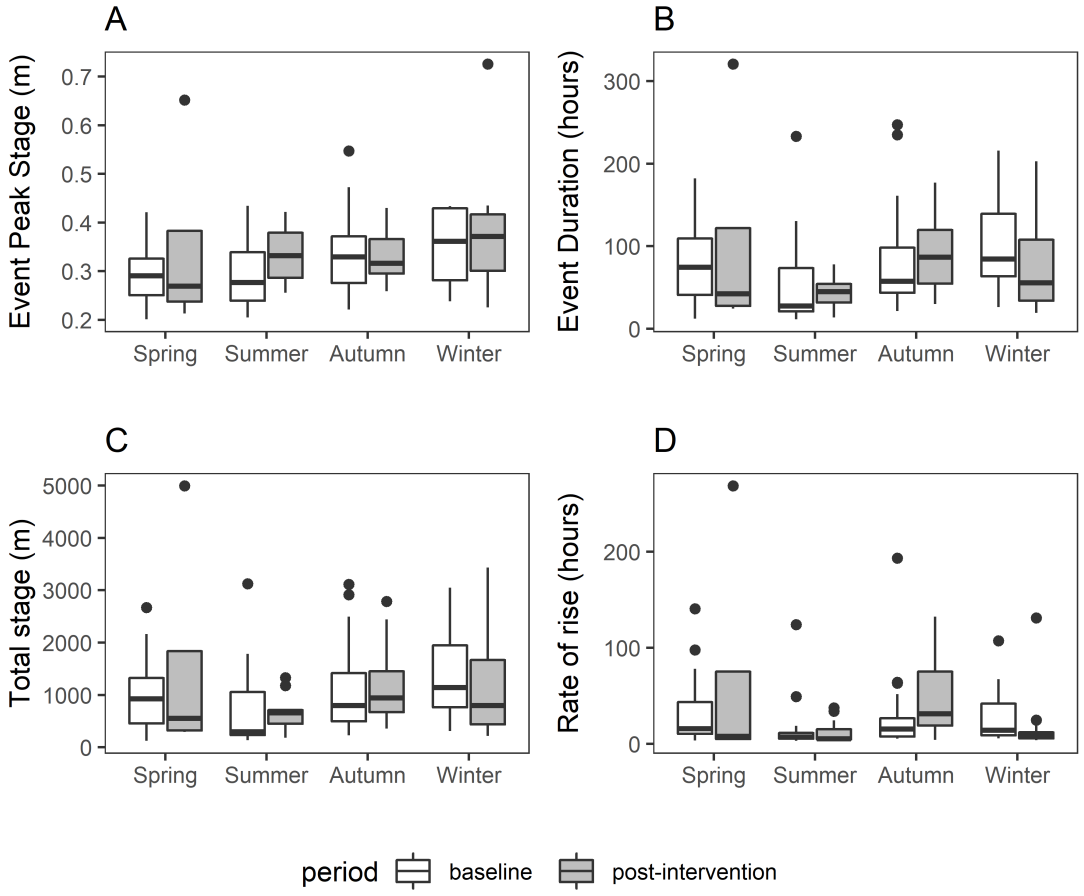


Figure 3.9: Seasonal variation of event peak stage (plot A), duration (plot B), Total stage (plot C) and Rising limb duration (D) on the control stream

was highly correlated with event duration and total stage (r >0.7) and moderately correlated with rising limb duration (r >0.5).

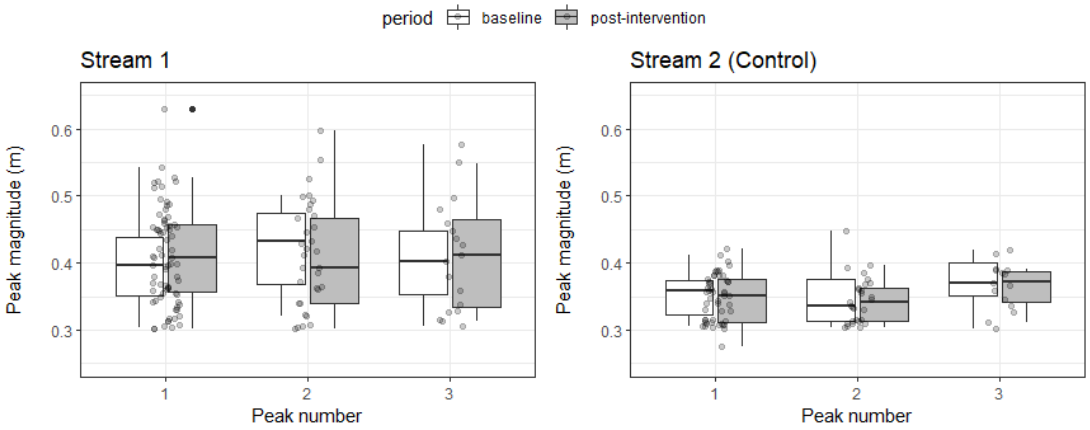


Figure 3.10: Peak stage for the first, second and third peak of high flow events in the baseline and post-intervention monitoring periods measured at the upstream extent of each study stream, jittered points illustrate the number of event peaks in each group.

During both the baseline and post-intervention monitoring periods, single, double, triple, and multi-peaked events were observed. On both the control and impact stream the distribution of upstream peak stage was similar regardless of whether the peak was the first, second or third in the event, in the baseline or post-intervention monitoring period (Figure 3.10). Events with up to eight peaks were observed but there was not enough data to include these in the analysis.

3.3.2 Event Simulations

The TFN models were fitted and validated in Chapter 2. In summary, the control stream was represented by a transfer function with five dynamic regression terms, and an ARMA noise model with four autoregressive terms and two moving average terms. The impact stream was modelled using five dynamic regression terms and an ARMA noise model with five autoregressive terms and three moving average terms. Both models performed well in validation with simulations of event peak magnitude within 2 cm across the whole range of event magnitudes for 95%

of events on the impact stream (Stream 1). On the control stream a high level of accuracy was achieved for events with a peak magnitude below 0.45 m, but errors for simulations of events with peak magnitude >0.45 m increased the 95% prediction interval to 5 cm. Nevertheless, because the model worked well for the majority of events, the 80% prediction interval was within 2 cm, similar to the accuracy achieved on the impact stream. This level of accuracy was sufficient to identify treatment effects which exceeded 2 cm to the 95% and 80% confidence level on the impact and control stream respectively.

The models were used in this chapter to simulate what the pre-intervention downstream stage response of the streams would have been for events observed in the post-intervention monitoring period. On the impact stream the observed stage was consistently lower than the simulated downstream stage (Figure 3.11) which meant that the post-intervention downstream stage response of the stream during the events in Figure 3.11 was lower than it would have been pre-intervention. On the control stream, the observed and simulated stage were similar, as expected, because no leaky dams were installed in the stream in the post-intervention monitoring period.

For all events the observed stage (with leaky dams) was similar to the simulated baseline stage up to 0.3 m (Figure 3.12, Table 3.3), the approximate height of the gap beneath the leaky dams. When the rising limb of the event exceeded a stage of 0.3 m and interacted with the leaky dams the observed stage was lower than the predicted baseline stage on the rising limb, falling limb and peak of the event. For events during which the treatment effect at the peak of the event was negligible or negative the observed hydrograph was almost identical (within the 95% prediction interval) to the simulated stage without leaky dams throughout the event (Figure 3.12).

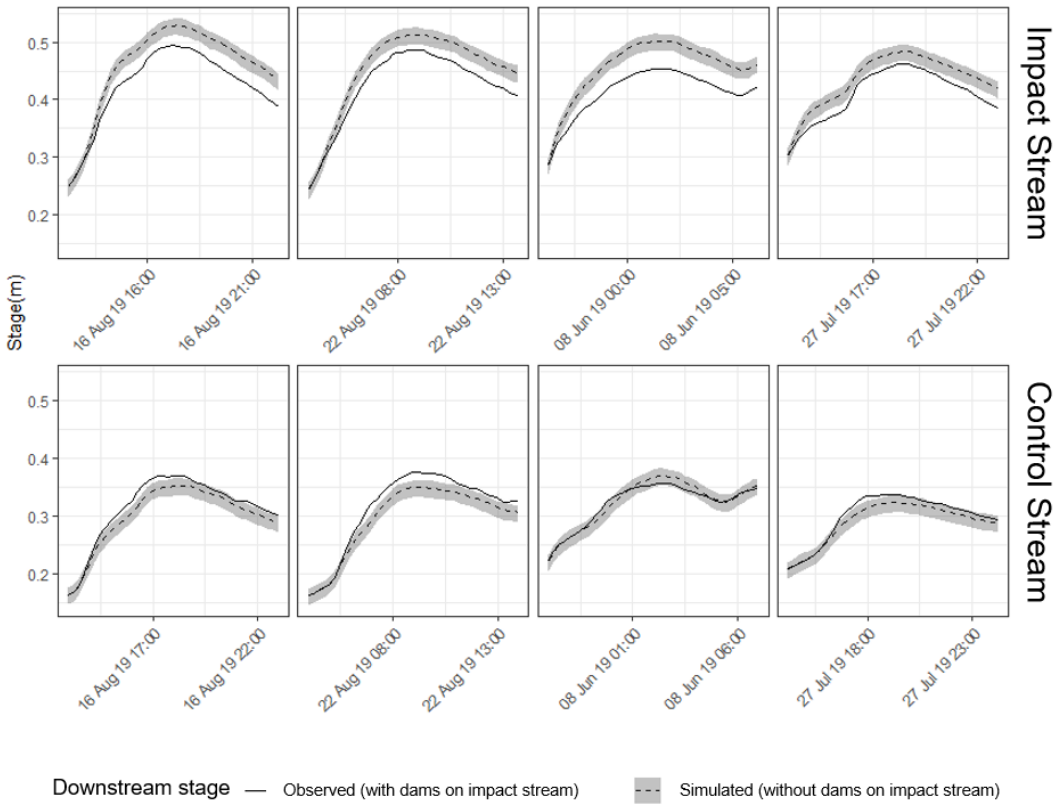


Figure 3.11: Four examples of event peaks observed in the impact and control streams after leaky dams were installed in the impact stream. The events had a return period <1 year and were ranked 12^{th},15^{th},18^{th} and 22^{nd} respectively in terms of peak stage on the impact stream. The solid line gives the observed stage (with leaky dam on the impact stream) and the dashed line gives the simulated baseline response of the streams (i.e., without leaky dams). The model prediction intervals are given by grey shading.

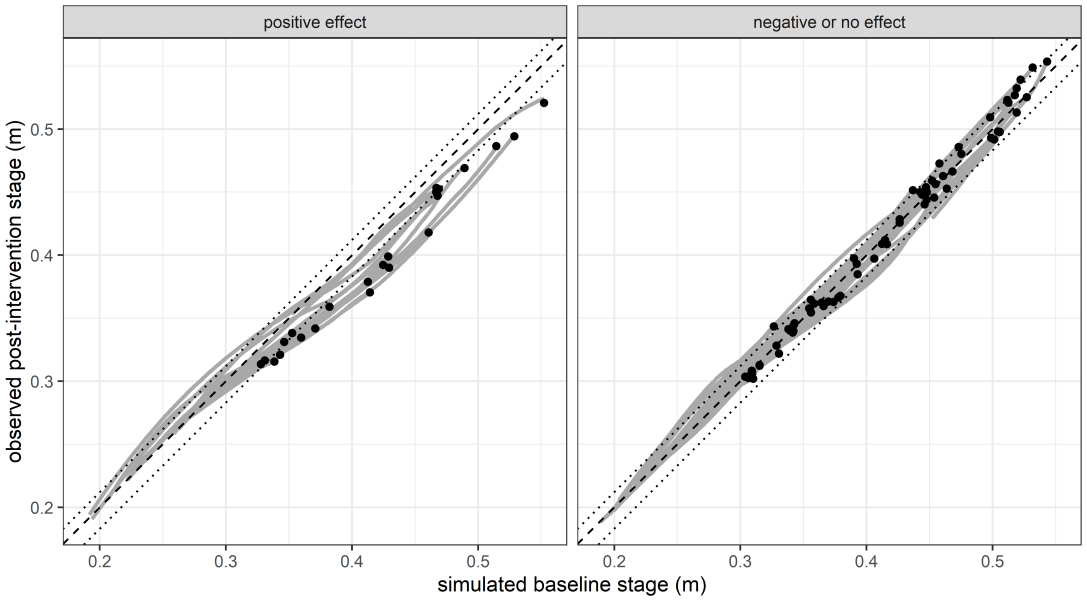


Figure 3.12: Simulated stage (without leaky dams) plotted against observed stage (with leaky dams) for the stage hydrograph of every event in the post-intervention monitoring period (solid grey lines) with its peak stage (black points) on the impact stream. The dashed line is 1-to-1-line and dotted lines are 95% model prediction intervals

3.3.3 Statistical significance of treatment effect

The treatment effect, T_e in the control stream was close to zero and almost always within the prediction limits of the model in both the pre-and post-intervention monitoring period, because no leaky dams were installed in the control stream (Figure 3.13). In the impact stream, however, the treatment effect following leaky dam installation was often larger than the 80% and 95% prediction limits of the model and was therefore attributed to the interventions, rather than model uncertainty. The two-sample Kolmogorov-Smirnov test was used to test whether the distribution of treatment effect was significantly different in the pre-intervention and post-intervention monitoring periods. The null hypothesis that the distributions of T_e came from the same population before and after the leaky dams were installed was rejected on the impact stream ($p < 0.01$) indicating that the leaky dams statistically significantly reduced event peak magnitude. On the control stream, the null hypothesis that the distribution of T_e in the pre- and post-intervention monitoring period came from the same population was not rejected ($p > 0.05$) indicating that downstream peak stage response of the control stream was not significantly different in the two monitoring periods.

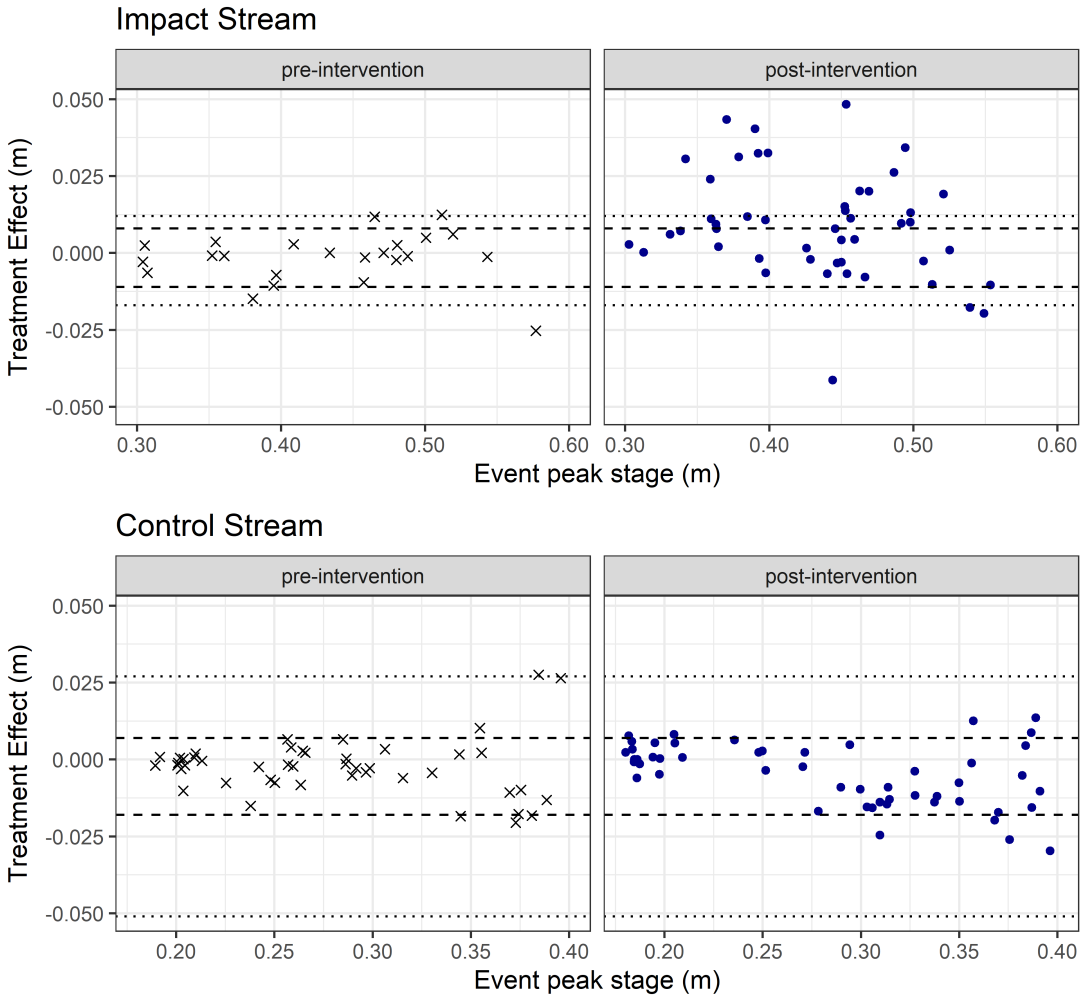


Figure 3.13: Treatment effect on peak stage in metres for high flow events observed in baseline and post-intervention monitoring periods on the impact and control stream. The dashed and dotted lines indicate the empirical 95% and 80% prediction intervals respectively.

3.3.4 Magnitude of treatment effect

High flow events with return periods between <1 and 6 years were observed on the impact stream (Appendix C). The magnitude of the treatment effect, T_{eQ} varied with the peak magnitude, (Q_p) of the high flow events (Figure 3.14). After installation of leaky dams on the impact stream, overall, peak magnitude was reduced by 8% on average for events with a peak magnitude of up to 1.2 m^3/s and by 10% on average for high flow events with a return period of up to 1 year (peak discharge 1.0 m^3/s). The linear relationship between treatment effect and peak discharge was significant (p-value = 0.02) but explained little of the variation in treatment effect (R^2=0.13, RSE = 15.9). The treatment effect was greatest for the most frequent, smaller events which interacted with the leaky dams (peak discharge ≥0.3 m^3/s) which were reduced by 16% (±8%) (Table 3.6). As the peak magnitude increased the reduction in peak magnitude decreased to 10% (±6%) for events with peak magnitude of 0.6 m^3/s, 5% (±6%) for events with a peak magnitude of 0.9m^3/s. For events with a peak magnitude of 1.2 m^3/s peak magnitude was decreased by 1% on average but increased or decreased by 11% within the 95% confidence intervals. Although they are not presented here, for events on the impact stream with a peak discharge below 0.3 m^3/s, which flowed through the gap beneath the leaky dams, the observed and simulated peak magnitudes were similar in the pre- and post-intervention monitoring periods.

Similarly to the percentage change in peak discharge, on average absolute peak discharge reduction was greater for smaller events (Figure 3.14). However, the largest reductions in absolute peak discharge (>0.2 m^3/s) for individual events were observed for three events with a peak magnitude of 0.7–0.9 m^3/s.

On the control stream there was a treatment effect of -2% on average in the pre-intervention monitoring period and -10% on average in the post-intervention monitoring period indicating that downstream peak magnitude on the control stream was higher in the post-intervention monitoring period than it would have been in the pre-intervention monitoring period. However, the two sample KS-test indicated that unlike the difference on the impact stream, the difference in the control stream was not significant (p >0.05).

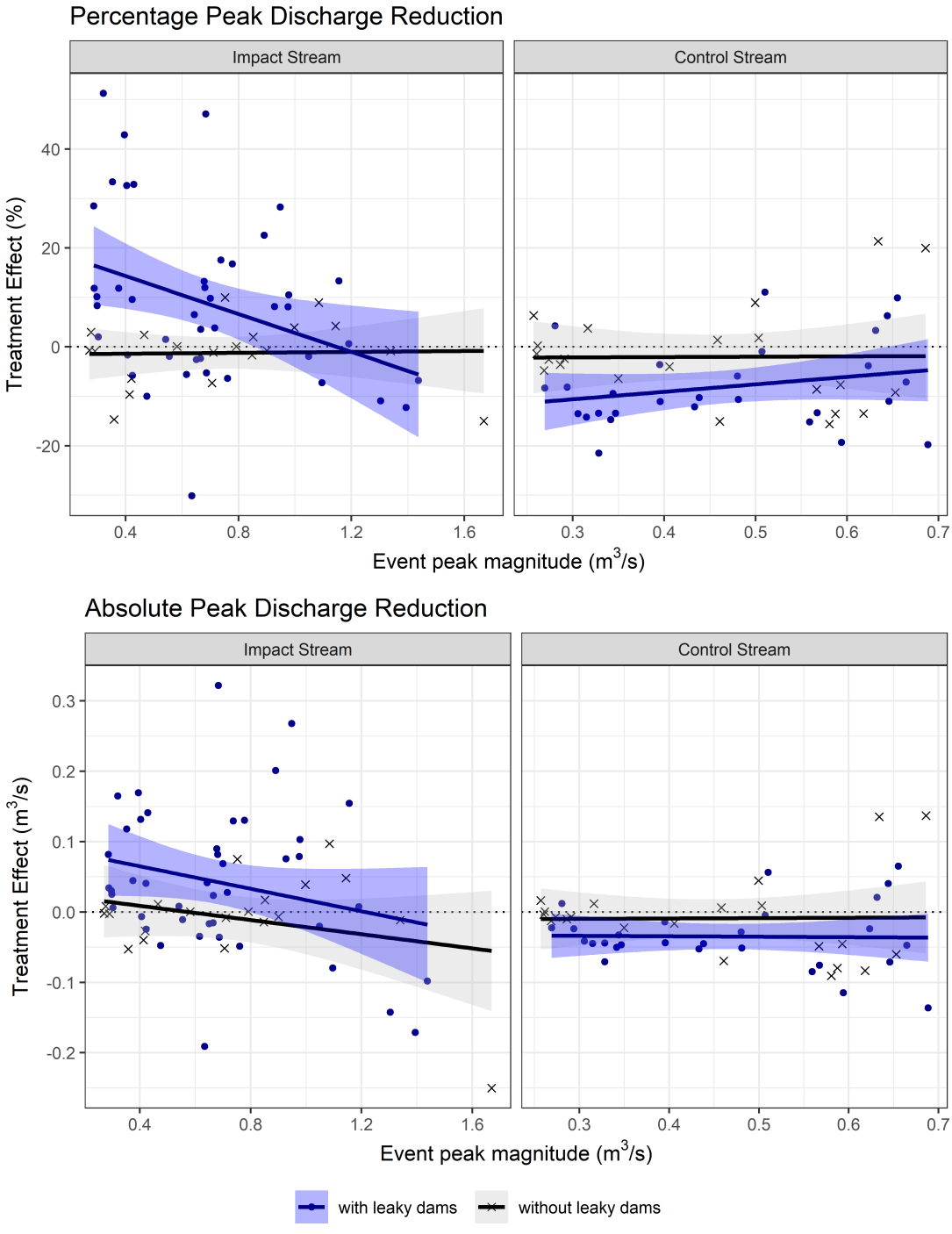


Figure 3.14: Treatment effect on peak discharge in percentage of peak discharge and absolute peak discharge for high flow events observed in baseline and post-intervention monitoring periods on the impact and control stream. A positive treatment effect indicates that the peak magnitude was reduced compared to the baseline scenario

Table 3.6: Percentage reduction in event peak discharge for events with peak discharge between 0.3 m^3/s and 1.0 m^3/s based on linear relationship between treatment effect and event peak magnitude

Peak Discharge (m^3/s)	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**	**1.1**	**1.2**	**Mean**
Reduction in peak discharge (%)	16	14	12	10	9	7	5	3	1	1	8
lower 95% confidence limit (%)	8	8	7	5	4	1	-1	-4	-7	-10	1
upper 95% confidence limit (%)	24	21	18	16	13	12	11	10	9	8	14

3.3.5 Variation of treatment effect

The treatment effect expressed as the reduction in peak discharge (T_{EQ}) varied between -30% and 51% on the impact stream (Figure 3.14) reflecting that, for the range of flows in which the interventions were observed, they could increase, decrease, or have a negligible impact on the event peak discharge compared to the baseline scenario. On the impact stream the leaky dams had a positive effect of reducing peak magnitude in almost a third (32%) of events at the 95% prediction intervals, and almost half (46%) of events at the 80% prediction interval (Table 3.7, Figure 3.13. On the control stream treatment effect of 87% of events was within the 80% prediction intervals, as expected.

Seasonal effects

The boxplots of treatment effect grouped by season (Figure 3.15) indicate that treatment effect was higher in the summer and autumn months on the impact stream (median T_e 0.014 m and 0.01 m respectively) and had negligible effects during spring and winter (median T_e -0.005 m and -0.002 m respectively). On

Table 3.7: Number of events with positive, none, or negative treatment effect (PI= Prediction interval), a positive treatment effect indicates a reduction in event peak magnitude.

Stream	Te > PI positive effect	PI < Te > PI no effect	Te < PI negative effect	Total
95% prediction interval				
Impact	16	31	3	50
Control	0	78	0	78
80% prediction interval				
Impact	23	22	5	50
Control	6	68	4	78

the control stream, treatment effect was generally within the model prediction intervals for every season. No formal statistical tests were performed because of the low number of events in some seasons, and because the peaks of multi-peaked events were not independent.

Impact of event characteristics on treatment effect

A positive treatment effect (i.e., points above the model prediction intervals) was observed for events with a shorter duration (D), lower total stage (S_t) and shorter time to rise (T_{rise}) (Figure 3.16). Time since previous event (D_a) did not affect the magnitude of treatment effect. Similarly, the change in peak stage, when expressed as an absolute value rather than a percentage of the peak magnitude, was similar for almost the whole range of stage. The number of days since the interventions were installed in the impact stream (T_{int}) shows that a positive treatment effect was observed only after 250 days of the dams being installed.

The majority of events with a positive treatment effect had short durations (D = 37-52 hours) and had a total stage between 241 m and 346 m whilst the majority of events during which the treatment effect was negligible spanned a wider range of durations (D = 32-77 hours) and total stage (S_t = 186-550 m) (Figure 3.17). However, there was no significant difference between the median of any of the event characteristics of events with a positive treatment effect or an insignificant

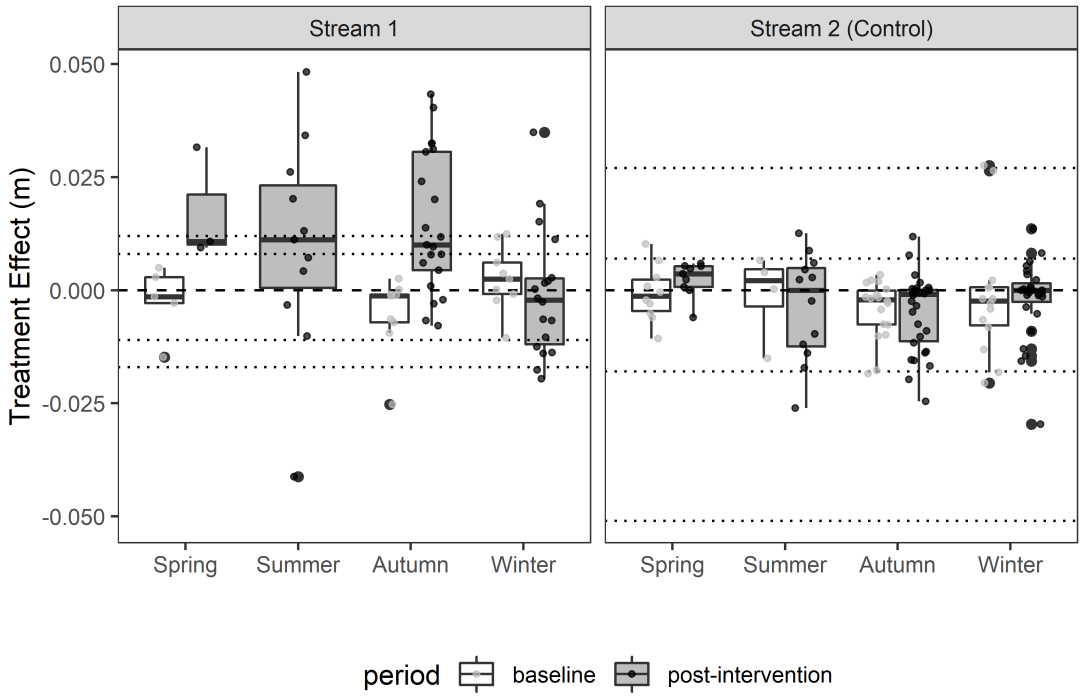


Figure 3.15: Seasonal variation of T_e for all event peaks, dotted lines indicate 80% and 95% empirical prediction intervals, dashed line indicates zero treatment effect and jittered points indicate the peaks in each group

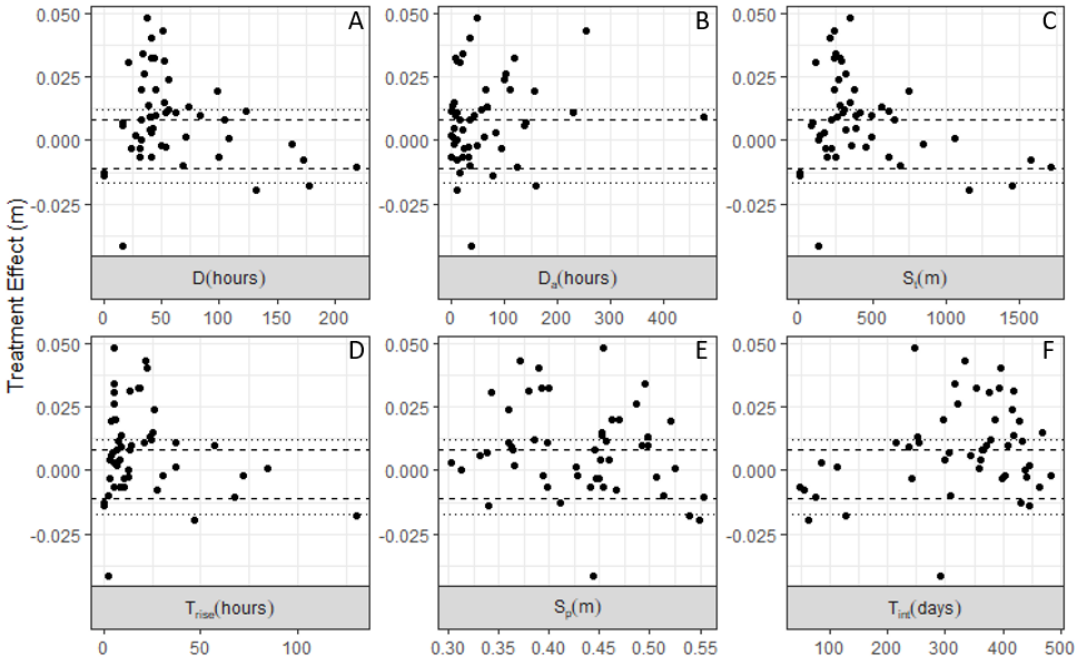


Figure 3.16: Effect of event characteristics on treatment effect, where (A) D is event duration, (B) D_a is time since previous event, (C) S_t is the total stage (as a proxy for event volume), (D) T_{rise} is the time taken for the rising limb to reach the peak of the event, (E) S_p is the peak stage and (F) T_{int} is the time since the interventions were installed in the impact stream.

treatment effect ($p > 0.05$ Wilcoxon-Mann-Whitney test). Treatment effect was negative for only three events and was therefore not included in the analysis.

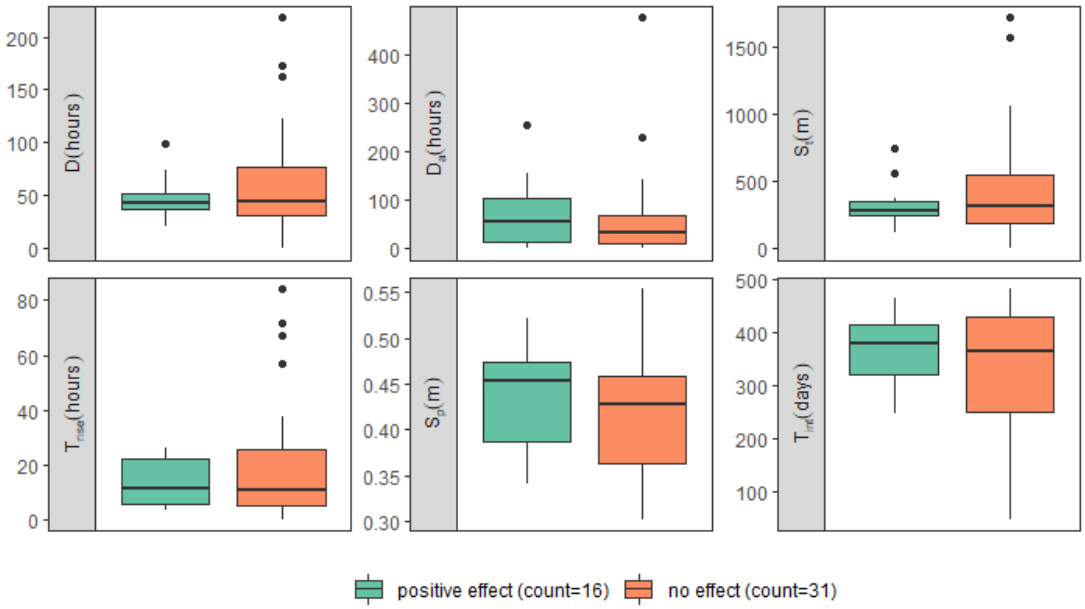


Figure 3.17: Distribution of event characteristics, duration (D), time since previous event (D_a), time to rise (T_{rise}), total stage (S_t), peak stage (S_p) and time since interventions were installed in the impact stream (T_{int})

Impact of peak order on treatment effect

During the combined baseline and post-intervention monitoring periods there were 144 peaks within 75 events on the impact stream and 115 peaks in 54 events with a peak magnitude greater than 0.3 m on the control stream. During the baseline monitoring period the observed peak magnitude and simulated peak magnitude was similar, and close to zero, for the first, second and third peak of events on both the impact and control streams (Figure 3.18), as expected. Following the installation of leaky dams on the impact stream, the treatment effect for the first peak increased significantly ($p < 0.01$ Wilcoxon-Mann-Whitney test) but remained similar for subsequent peaks ($p > 0.05$). On the control stream treatment effects in the post-intervention monitoring period were significantly

lower than in the baseline monitoring period for the first peak ($p = 0.04$), but not for subsequent peaks ($p > 0.05$).

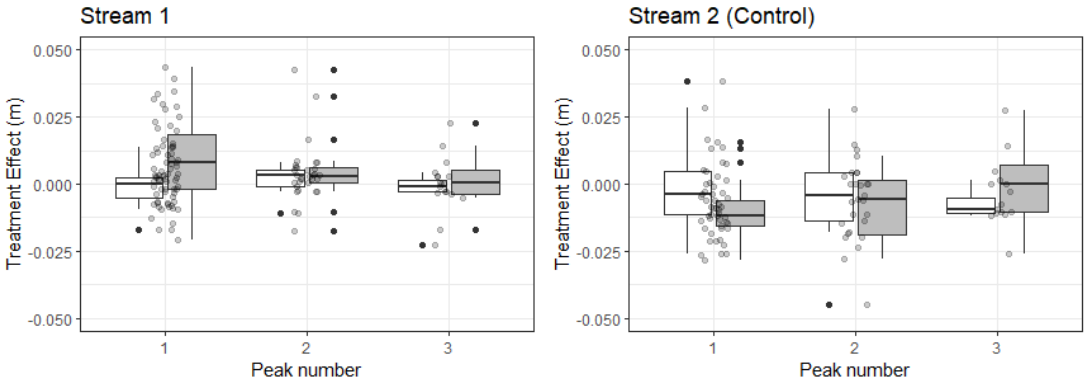


Figure 3.18: Impact of peak order on treatment effect, jittered points illustrate the number of event peaks in each group.

3.4 Discussion

Leaky dams are increasingly being installed in upland watercourses as a low-cost, sustainable solution to frequent flooding of rural communities. However, there is a paucity of information about their impact on the flood hydrograph, particularly in upland catchments, which presents a barrier to scheme design, cost-benefit analysis, and stakeholder buy-in. A BACI-style monitoring approach combined with data-based time series modelling techniques has shown how leaky dam interventions statistically significantly reduced flood peak magnitude for 32% of high flow events. Treatment effect-size was highly variable, but on average, flood peak magnitude was reduced by 10% for events up to a 1-in-1 year return period. This dataset is the first of its kind and offers insights for increases confidence in the use of leaky dams and for the design and assessment of NFM schemes.

3.4.1 Effect of leaky dams on flood peak magnitude

This study has shown, for the first time, that leaky dams installed for the purpose of NFM can statistically significantly reduce the peak magnitude of high flow

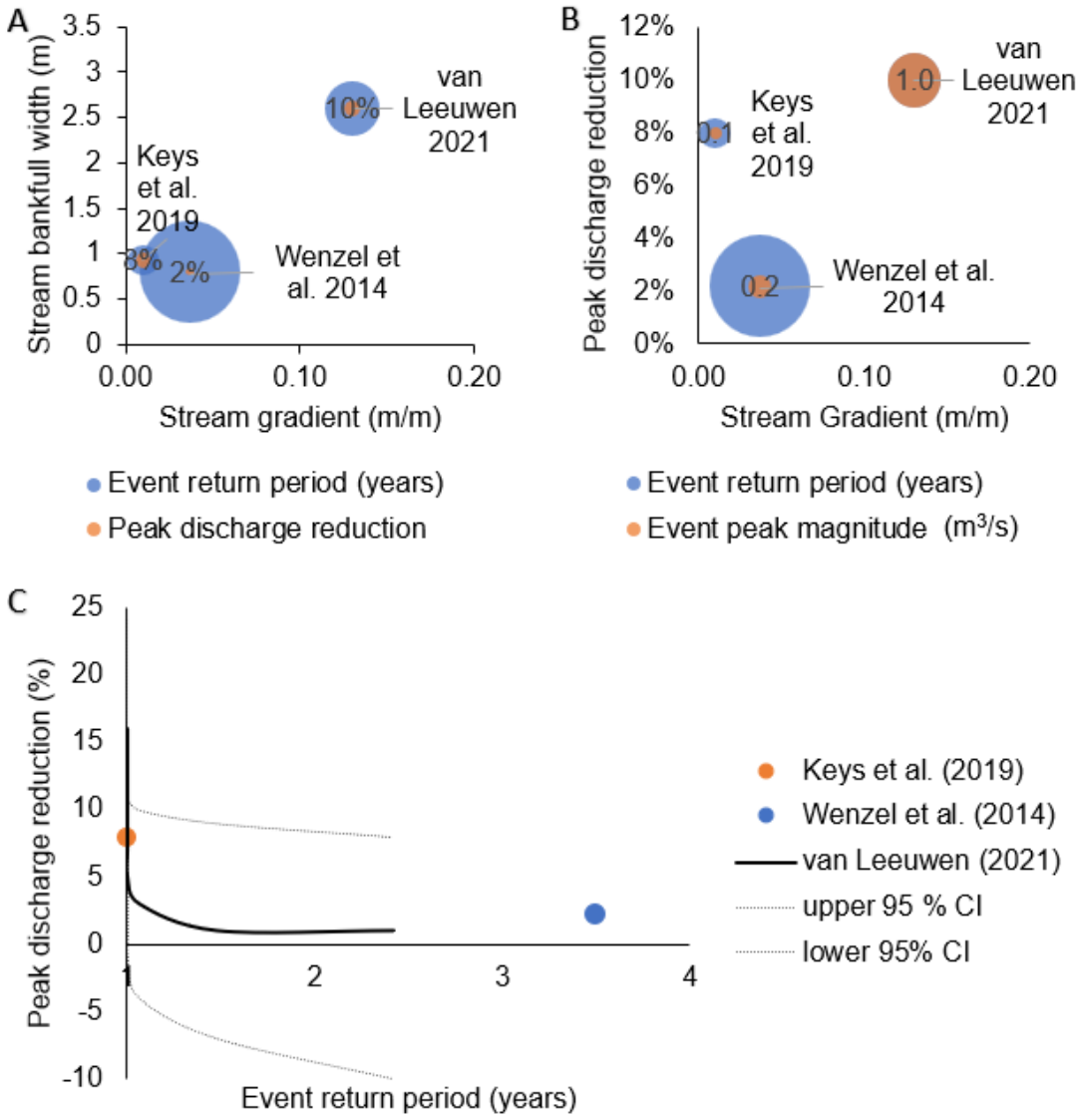


Figure 3.19: Comparison of Coverdale leaky dam impacts (denoted 'van Leeuwen 2021') to previous research (Keys et al., 2018; Wenzel et al., 2014). The size of the blue circles indicates the event return period of 3.5 years (Wenzel et al., 2014), 1 year (this study) and considerably less than 1 year (Keys et al., 2018). Points on the y-axis in figure (C) indicate events with a return period $<$ 1 year.

events at the stream scale. The study quantified the impact of leaky dams on larger, steeper streams than previous research (Figure 3.19A), which focused on in-stream wood installed for the purpose of river restoration, rather than flood risk management. The average reduction in event peak magnitude was similar to the findings of Keys et al. (2018) who found just three pieces of large wood in a 50 m section of a steep, headwater stream in the Mid-Atlantic region of the US decreased flood peak magnitude by 8% for a < 1-in-1 year event. Similarly, the findings were in agreement with Wenzel et al. (2014), who assessed the impact of nine spruce tops placed in a small, steep stream in the Ore mountains, Germany, that for a more extreme event (1-in-3.5 year) the impact on flood peak magnitude was negligible (Figure 3.19C).

Leaky dams installed for the purpose of NFM are designed to maximise flood attenuation and were therefore expected to have a larger impact on flood peak magnitude than the type of wood installed by Wenzel et al. (2014) & Keys et al. (2018). However, the comparison of the effectiveness of different types of in-stream wood is confounded by differences in structures, sites and flood events. The streams in the study site were considerably steeper and wider than the other studies (Figure 3.19A), and although the events had a similar flood return period, the peak discharge of the events tested by Wenzel et al. (2014) & Keys et al. (2018) were much smaller (<0.2 m^3/s). Nonetheless, similarity in the impact on flood peak magnitude for different types of instream wood structures, as seen between these studies, could have implications for leaky dam design. It may be possible to design instream wood for the purpose of natural flood management which optimises its geomorphological impacts for ecological (Pilotto et al., 2014; Roni et al., 2014; Wohl et al., 2015), water quality (Janes et al., 2017) or erosion control (Blanckaert et al., 2012) benefits, and has similar impacts on flood peak magnitude as the types of leaky dams tested in this study.

Contrary to the findings of modelling studies (Dixon et al., 2016; Thomas and Nisbet, 2012), comparison of the evidence from this study to existing literature suggests that the impact of leaky dams on event peak magnitude does not necessarily decrease with increasing stream gradient (Figure 3.19B). The limit of effectiveness of leaky dams in the study streams was for events with a peak discharge of 1 m^3/s, the same magnitude as the limit of effectiveness of leaky

dams at delaying flood peak timing in a low gradient (0.0075 m/m), headwater stream with similar dimensions (channel width 3 m, bank height 1.2 m) in the New Forest, UK (Gregory et al., 1985; Kitts, 2010). The limit of effectiveness equated to a 1-in-1 year event in the study site and a 1-in-2 year event in the New Forest. However, as there are few studies differences and similarities in the empirical findings could be due to a combination of site, structure, and event specific factors, rather than the stream gradient. Additionally, differences between model and empirical findings could be due to the lack of validation of the way in which leaky dams are represented in hydraulic and hydrological models (Addy and Wilkinson, 2019). Hence, without empirical data from a larger number of studies and validated representation of leaky dams in hydraulic and hydrological models it remains difficult to move on from site specific knowledge of leaky dam impacts. Crucially, this study has demonstrated that it is possible to robustly address this evidence gap using a data-based time series modelling approach.

Unlike Wenzel et al. (2014) & Keys et al. (2018) who studied just one type and magnitude of event, the data-based time series modelling approach taken in this study allowed the variability of leaky dam impacts in the stream to be studied for 50 different storm events with a range of characteristics such as peak magnitude, duration, number of event peaks and time since previous event. The variability in leaky dam effectiveness in this study indicates that it is not only site specific considerations, but also event specific consideration which confound the comparison and quantification of leaky dam impacts on flood peak magnitude. Like Gregory et al. (1985) & Kitts (2010), the current study observed decreasing impacts of leaky dams with event peak magnitude at the stream scale, supporting reviews of the NFM approach (Burgess-Gamble et al., 2017; Dadson et al., 2017; Lane, 2017) in questioning leaky dam impacts at higher event magnitudes. This conclusion could be a product of the the number and size of interventions tested in this study and elsewhere (Ellis et al., 2021), although modelling has shown that adding more leaky dams in steep streams does not necessarily mean their storage will be utilised (Hankin et al., 2020).

Field observations support the hypothesis of Dixon et al. (2016) that the limit on effectiveness is due to overtopping, or drowning out of the structures and their friction effects. However, evidence is emerging that this effect may

be stream specific; within the same catchment leaky dams have been observed to have increasing impacts on flood peak timing in some sub-catchments, and not in others (Black et al., 2021). Increased leaky dams effectiveness at higher return periods has also been observed in modelling studies (Thomas and Nisbet, 2012) and is attributed to increased flood plain connectivity and expandable field storage during higher return period events (Black et al., 2021; Hankin et al., 2020; Kay et al., 2019; Thomas and Nisbet, 2012).

Although there was a relationship between treatment effect and peak magnitude, the variability in effectiveness in the study site was not explained by peak magnitude alone. Whilst leaky dams did reduce event peak magnitude of some events, their effectiveness was highly variable, and could both decrease and increase downstream peak magnitude. Variability in leaky dam scheme effectiveness was found in a catchment scale model of leaky dam impacts (Dixon et al., 2016), but was attributed to varying interactions of sub-catchments which did not play a role at the stream-wide scale of this study. None of the flood event characteristics (event duration, time since previous event, time to rise, total event stage or peak stage), seasonal effects or peak order alone determined during which events reductions in peak magnitude were or were not observed. For example, 84% of peaks which were significantly reduced occurred in summer and autumn and 80% were the first peak of an event. However, 45% of events in which there was no discernible impact, or a negative impact also occurred in summer and autumn, and 46% were the first peak of an event. There was considerable co-linearity between event peak magnitude and many of the event characteristics, including seasonality which indicates that it is not clear which characteristic, or combination of characteristics determined whether the leaky dams significantly reduced event peak magnitude or not. Because the dataset did not span several years there was not enough data to separate seasonal effects from event characteristics or the time since the leaky dams were installed. Continued monitoring to obtain a larger number of events would provide the data required to perform a formal statistical analysis to determine which event characteristics are useful predictors of leaky dam effectiveness. This would help to ascertain, for example, whether it was the typically short, intense nature of summer storms, which made the interventions more effective during the summer months, or whether this was

down to seasonal effects, such as leaf litter reducing leaky dam porosity in the autumn months, as hypothesised by Thomas and Nisbet (2012).

The results, nevertheless, empirically showed that leaky dams were more likely to be effective during single peaked events, and the first peak of multi-peaked events, than during subsequent peaks. This indicates that the effectiveness of leaky dams in the study site was affected by the availability of a finite amount of storage volume, which could not be used if the conditions for the system to drain had not been met between event peaks. This implies that, unlike the leaky dams observed by Black et al. (2021) and modelled by Thomas and Nisbet (2012), the leaky dams were not able to utilise expandable field storage and relied on the system draining between event peaks. The importance of meeting the conditions for the system to drain between event peaks was noted in a catchment scale modelling study of the impact of 59 in-channel interventions on a double peaked flood event in Brompton, North Yorkshire (Metcalfe et al., 2017) and is an important factor to consider in the design and assessment of leaky dam effectiveness.

The variability in leaky dam effectiveness highlights the importance of assessing leaky dam impacts on a large range and variety of events, which is made possible by the data-based time series modelling approach used in this study. Comparisons of a small number of similar events monitored in the baseline and post-intervention monitoring period, or one type and magnitude of artificial flood peak can be misrepresentative and may distort expectations of leaky dam effectiveness. Similarly, this study provides empirical evidence to support the conclusion drawn from catchment scale modelling that multi-peaked events should be considered in the design and assessment of leaky dam efficacy (Metcalfe et al., 2017), even when considering their impacts at the stream scale. Had the study considered only events with well-defined peaks, as for example Gregory et al. (1985) did, it may have concluded considerably higher average leaky dam impacts despite the majority of events being multi-peaked. The assessment of leaky dam impacts on a range of events helps to manage expectations, which is crucial to avoid over-reliance of communities on NFM measures (Wells et al., 2020), and to sustain confidence in new paradigms in the long term (Vira and Adams, 2009).

3.4.2 Implications for downstream flood risk

The interventions had a negligible local impact on peak magnitude during events which caused significant flooding of properties and transport links in North Yorkshire (flood return period > 1 year). Nevertheless, whilst it remains to be seen what impact reduced peak flows in headwater streams have further downstream, the more frequent, small-scale flood events which were impacted by the interventions can cause localised flooding which can result in waterlogging, bank erosion and debris deposition on downstream agricultural land (Posthumus et al., 2008). The vulnerability of agricultural land depends on land use and the frequency, duration, depth and seasonality of the flood event (Morris and Hess, 1988). Whilst grassland used for livestock can tolerate winter flooding, summer floods can destroy an entire harvest (Morris and Brewin, 2014; Posthumus et al., 2009). The indication that leaky dams may be particularly effective in reducing event peak magnitude during summer storms is promising for the reduction of economic costs of flooding to farmers. Flooding and waterlogging of agricultural land was estimated to have an average economic cost of £12,000 per hectare, or £90,000 per farm, albeit during the extreme 2007 UK floods (Posthumus et al., 2009). Furthermore, placing leaky dams in upland watercourses reduces the reliance on flooding of productive agricultural land as temporary flood storage areas which can reduce the costs associated with compensation payments under schemes such as the UK's proposed Environmental Land Management Scheme (ELMS).

Although leaky dam impacts may be small at the stream scale, by desynchronising tributary flows, downstream flood risk during large floods can be significantly reduced (Pattison et al., 2014). Modelling has shown that leaky dams, in combination with reforestation, could have considerable impacts on catchment scale flood risk (Dixon et al., 2016). Although catchment scale impacts were not assessed in this study, it has provided a validation dataset needed to increase confidence in the representation of leaky dams in hydraulic and hydrological models which can be used to assess impacts on downstream flood risk (Addy and Wilkinson, 2019).

3.4.3 Limitations of BACI design and the data-based time series modelling approach

BACI monitoring design is proposed as the cornerstone of a research framework to collect the evidence required to increase confidence in NFM (Ellis et al., 2021). However, this study showed that high levels of uncertainty, which are typical of hydrological data (Beven, 2016), present a limitation to the use of BACI designs to detect NFM impacts. Changes in event peak magnitude or timing could not be directly assessed from the data collected on any of the streams due to frequent errors in stage datum which led to high levels of uncertainty in peak stage. This type of error is common (Guerrero et al., 2012; Searcy and Hardison, 1960; Shaw et al., 2011; Westerberg et al., 2011), and cannot necessarily be overcome by using artificial gauging structures such as flumes or weirs due to considerations of cost, drowning out or bypassing of the structure during flood flows, and potential undesirable ecological effects such as blockage to fish passage (Gordon et al., 2004; Shaw et al., 2011). The time series modelling approach taken reduced the impact of such errors and allowed remaining uncertainty to be represented by empirical prediction intervals. However, of the three streams modelled only the model for one of the impact streams was sufficiently accurate to be able to assess whether leaky dams had an impact across the whole range of observed peak stage. This means that, although leaky dam impacts were assessed for many events, they were assessed for only one impact stream. Replicates within the catchment would have increased confidence in the findings and would have allowed the transferability of findings between streams to be assessed.

A better fit of the control stream model to the largest events on the control stream would have increased confidence in the model overall but would not have affected the results of the study as the model was adequate at simulating peak stage of events in the range in which leaky dams affected peak flows on the impact stream. The control stream allowed wider impacts such as land-use change, which would impact both streams to be isolated from the impact of the interventions. However, without further replicates of the impact stream, confidence in attributing change to leaky dams relies more heavily on the assumption that there were no other changes in the impact stream or its gauging stations. This assumption

was backed up by a thorough quality assurance process (Appendix B) and local knowledge of changes which could affect the hydrology of individual streams such as stocking density or harvesting of the commercial riparian forest on the streams, gained through frequent field observations and correspondence with the landowner. However, replicates of impact streams, and/or the control stream (Underwood, 1994) would have increased confidence in attribution of change in the impact stream to the leaky dams.

This study has provided insights, for the first time, of leaky dam impacts on a large number and variety of events with a range of characteristics. Changes in peak magnitude were assessed for 50 high flow events with a return period ranging from < 1 year to 6 year and up to eight event peaks. Whilst the findings illustrated the variability of leaky dam efficacy during different types of events, a larger number of events observed over a longer time period, or from a larger number of impact streams are needed to perform a formal statistical analysis to assess the sources of variation. The visual analysis of seasonal effects, event characteristics and peak order in this study are exploratory and give an indication of avenues for further research. The data-based time series modelling approach allowed multiple data issues to be overcome, but it was not able to accurately simulate the timing of the event peak and could therefore not be used to assess the efficacy of upland leaky dams at delaying the flood peak. The impact of leaky dams on flood peak timing is arguably more important than their impact on flood peak magnitude in headwater catchments due to the potential to affect relative tributary timing, which can significantly increase or decrease catchment scale flood risk (Black et al., 2021; Dixon et al., 2016; Pattison et al., 2014).

3.4.4 Recommendations for next steps

For leaky dams to be used as a mainstream flood risk management measure evidence of their impacts on the flood hydrograph is required in a range of environments and at a range of spatial scales (Dadson et al., 2017; Ellis et al., 2021; Lane, 2017). This study has shown that it is possible to assess stream scale impacts of leaky dams using a data-based time series modelling approach. This approach overcomes difficulties associated with short lead times, high levels of

uncertainty in hydrological data and lack of comparable events before and after intervention. By applying this approach to data from other catchments, evidence could be gathered for a range of high flow events in a range of environments and leaky dam designs which would provide the evidence needed to move on from understanding of site-specific impacts of leaky dams only.

Development of hydraulic and hydrological models of the study site is recommended so that the representation of leaky dams can be assessed against the findings from this study. Hydraulic and hydrological models are used to design and assess the impacts of flood risk management interventions. However, because of the lack empirical validation data there is little guidance as to how to represent leaky dams appropriately in numerical models, leading to low confidence in their outputs (Addy and Wilkinson, 2019). This study provides one of the most comprehensive quantifications of leaky dams impacts to date, spanning a range of peak magnitudes and event types, providing a diverse validation dataset with which to assess the representation of artificial large wood in in hydraulic and hydrological models in steep upland rivers. Increased confidence in the representation of leaky dams in hydraulic and hydrological models is particularly important to be able to address questions about the impacts of NFM measures at larger spatial scales (Burgess-Gamble et al., 2017; Dadson et al., 2017; Ellis et al., 2021; Lane, 2017).

3.5 Conclusion

Leaky dams have the potential to decrease the magnitude of some frequent flood peaks (up to 1 year return period) on high gradient streams by 10% on average, but their effects are highly variable. The data-based time series modelling approach allowed the impacts of leaky dams during a large number and range of event types to be assessed, for the first time. The results have important implications for the design and assessment of leaky dam schemes. Whilst event peak magnitude is an important factor to consider when designing leaky dam schemes, the conditions required for the system to recover between event peaks are also important. Whether assessing the impact of leaky dams empirically or using numerical models, the results show that an assessment of leaky dam impacts is

not complete without considering a range of event types as well as event peak magnitudes. Leaky dam schemes which are assessed using single-peaked design storms only are likely to overpromise and underdeliver on flood risk management benefits. By supporting the BACI approach with data-based time series modelling, the challenges associated with quantifying NFM effectiveness in a range of environments can be overcome.

Chapter 4

A method for assessing the resilience of leaky dam networks

4.1 Introduction

Natural Flood Management (NFM), also known as 'Working with Natural Processes' (WwNP) to reduce flood risk, is gaining popularity, especially in communities which do not qualify for funding of traditional flood defence schemes (Wilkinson et al., 2019). NFM measures are thought to reduce and delay flood peaks alongside providing multiple benefits such as habitat creation and sediment management (Burgess-Gamble et al., 2017). A common feature of NFM schemes is the re-introduction of instream wood to watercourses in the form of channel spanning 'leaky dams'. Leaky dams interfere with high flows by increasing channel roughness and floodplain connectivity (Grabowski et al., 2019).

Whilst work is on-going to quantify the benefits of NFM schemes (e.g. Lancaster Environment Centre, 2021; University of Reading, 2021), little is known about the risks and resilience of NFM features (Environment Agency, 2018). Without knowing the resilience of NFM features to flood induced structural failure it is difficult for risk management authorities, such as the Environment Agency and lead local flood authorities in the UK, to make informed decisions about the impact of features on downstream flood risk. This presents a barrier to implementation, particularly in the case of leaky dams which are perceived to present a

potential blockage risk to downstream infrastructure (Waylen et al., 2018). Furthermore, because leaky dam benefits cannot be quantified robustly without a better understanding of their resilience, confidence regarding their use in strategic level planning and investment decisions is limited. The quantification of the system performance of leaky dams is pertinent in light of the intended Environmental Land Management scheme (ELMS) in England and Wales which will likely see increased implementation of integrated catchment management approaches such as NFM (Klaar et al., 2020).

Failures of leaky dams have been observed in various locations throughout the UK (e.g. Hankin et al., 2020). Damaged infrastructure and 13 fatalities are attributed to outburst floods caused by beaver dam failures in the US and Canada (Butler and Malanson, 2005) and exacerbated localised flooding due to large wood pieces blocking structures such as bridges and weirs is well documented in the UK and worldwide (e.g. Comiti et al., 2006, 2008; Diehl, 1997; Fenn et al., 2005; Lyn et al., 2007; Ruiz-Villanueva et al., 2014). Understanding the risks posed to infrastructure by large wood is crucial for managing wood in rivers (Ruiz-Villanueva et al., 2016), and its mobility is therefore studied for naturally occurring large wood (e.g. Dixon and Sear, 2014; Gurnell et al., 2002; Ruiz-Villanueva et al., 2014). Whilst insights from this research have been used to inform the design of leaky dams (e.g. Yorkshire Dales Rivers Trust, 2018), little is known about the resilience and mobility of engineered leaky dams. As pointed out by Dixon and Sear (2014), being able to forecast the stability of wood structures is crucial to the success of engineered leaky dams for flood risk management. Failure of leaky dams could not only provide material for downstream blockage, but could also reduce the effectiveness of the NFM scheme and potentially release a water surge (Hankin et al., 2020), similar to a traditional dam break flow albeit on a reduced scale. A better understanding of leaky dam failure rates, mechanisms and impact on the flood hydrograph would allow for mitigation of these risks by informing inspection and maintenance regimes as well as dam scheme design. Furthermore, it would lead to a more informed debate around issues of liability, which can limit the uptake of NFM measures in a similar way in which it has limited the adoption of sustainable urban drainage systems (Wingfield et al., 2019).

Hazard assessment is needed to assure flood risk managers, stakeholders and affected communities that the benefits of an NFM scheme outweigh the potential hazards in both the short and long term (Wohl et al., 2016). Robust design and understanding of risk reduction strategies is needed to account for the residual risk associated with failure of a flood defence to perform as intended and avoid inadvertently exposing communities to risk (Hankin et al., 2020). For example, network analysis of a series of eight leaky dams in West Cumbria, UK, was used to inform effective placement of dams to utilise dynamic storage and reduce the risk of breach and cascade failure (Hankin et al., 2020). The authors argued that the risk of cascade failure was more important for dam placement decisions than optimising the placement for reductions in peak magnitude. Assumptions about the conditions under which leaky dam failures occur are therefore key for network analysis and could be improved by an evidence based fragility function (Hankin et al., 2020).

Fragility functions are central to the methods used at a strategic planning level in the UK to assess the resilience of flood risk management assets in a risk based framework (Ramsbottom et al., 2005). A fragility function expresses the probability that an asset reaches or exceeds an undesirable limit state as a function of an environmental loading condition (Porter, 2020). They are used by the Environment Agency (EA) as a tool to support risk based strategic flood management planning in England (Ramsbottom et al., 2005), including the National Flood Risk Assessment (NaFRA), and in flood management, are typically based on failure mode analysis of structures. Fragility curves can also be estimated based on statistical analysis of observations of failure when failure modes are not well understood or consist of complex interactions (Schultz et al., 2010). This is a commonly used approach in earthquake engineering (Porter et al., 2007). One of the advantages of this method is that the analysis requires minimal data input. For each structure only a measurement or estimate of the largest loading condition that was withstood or that caused failure is required.

Reporting incidences of leaky dam failure forms part of the monitoring requirements of £15m Defra funded NFM projects (Arnott et al., 2018) and will be openly accessible through the Rivers Trust NFM monitoring and evaluation tool (The Rivers Trust, 2021). For asset failures related to flood events, Lamb

et al. (2019) demonstrated that the loading condition of fragility functions can be expressed as a relative measure of extremeness of the flood event. The relative extremeness of a flood event is usually expressed as its return period in years, or equivalently, its annual exceedance probability, and relates to physical drivers of failure such as water depth and velocity (Lamb et al., 2019). Estimates of a flood's return period can be made based on the probability distribution of the peak over threshold or annual maximum series (Shaw et al., 2011) which are readily available for flow or stage gauging stations around the UK (UK Centre for Ecology and Hydrology, 2021). Hence, the information needed to estimate national fragility functions for leaky dams is starting to become readily available for sites in the UK.

The aim of this chapter is to address the following research question by evaluating the resilience of NFM leaky dams installed in the UK:

Q What potential does empirical fragility analysis have for quantifying the resilience of engineered leaky dams during extreme flood events?

This will be achieved via the following specific objectives:

O1 to compile direct observations and practitioner experience of leaky dam failures and resilience;

O2 to demonstrate a novel method to infer fragility functions from failure observations of leaky dams; and

O3 to present a pilot study of leaky dam fragility based on data pooled from NFM projects around the UK.

4.2 Methods

The resilience of leaky dams was assessed by estimating a fragility function based on observations of leaky dam failure, partial failure and resilience from UK Natural Flood Management projects. Fragility functions are useful where the capacity of an asset to resist failure or damage is uncertain and lend themselves

to risk-based analysis because they represent the uncertainty using a probability distribution.

This chapter comprises two stages of analysis, representing scenarios of varying data availability. The first stage estimates a fragility function of NFM dams in a well-monitored upland catchment (Section 4.2.2). In the second stage, information from the closely monitored leaky dam study site is combined with information from other NFM sites from responses to a UK wide survey of practitioners. Estimates of the fragility function were made both by treating data from each site individually and by pooling data from all sites using, in both cases, a two-stage optimisation procedure of the lognormal fragility distribution as described in Lamb et al. (2019).

The information required to estimate the fragility function for a site is the total number of assets, the number of assets which have reached a limit state (i.e. failed or partially failed, see Section 4.2.2) and the loading condition which induced the limit state, or the highest loading condition to which the asset was resilient if it had not reached the limit state (Section 4.2.3). The fragility analysis is briefly explained in Section 4.2.1. Section 4.2.2 gives details of the Coverdale field site in which leaky dams were built and closely monitored over a three year period. It also provides details of a survey of leaky dam practitioners to obtain information about leaky dam failures across the UK. Section 4.2.3 explains how the loading conditions used to estimate the fragility functions were obtained.

4.2.1 Fragility analysis

In a fragility function, the probability that an asset fails is expressed in terms of a function of the uncertain physical loading conditions that would cause failure. The uncertain loading condition is defined as a random variable, X. A particular observation of the loading condition is $X = x$. Uncertainty about X is characterised by the probability density function $f_X(x)$. The probability that X takes a value not exceeding an observed value, $x = x_a$, is the cumulative distribution function, $F_X(x_a)$, evaluated at x_a, and is given by integrating the density function from minus infinity $(-\infty)$ to x_a, such that

$$P[X \leq x_a] = F_X(x_a) = \int_{x=-\infty}^{x_a} f_X(x)\, dx \tag{4.1}$$

An alternative and equivalent (Porter, 2020) interpretation of the fragility analysis is that X represents the uncertain "strength" of an asset, which is measured in terms of the limiting load, or some function of the load. For example, the strength of leaky dams could be expressed in terms of the limiting hydraulic forces that cause failure, which are in turn functions of the measurable quantities such as water levels or flow rates. If the uncertain limiting load (strength) is less than or equal to a particular observed load, then the leaky dam will fail. Hence, the chance of failure given an observed load, x, is the probability that the uncertain strength, X, does not exceed x. This failure probability is $P[X \leq x]$, as defined in Equation (4.1), and $F_X(x)$ is a fragility function. The fragility function for leaky dams was approximated using a lognormal distribution, which has two parameters, θ, the central tendency (median), and β, the dispersion (logarithmic standard deviation), such that

$$F_X(x) = \phi\left(\frac{ln(\frac{x}{\theta})}{\beta}\right) \tag{4.2}$$

where $\phi(.)$ is the standard normal distribution function. The lognormal distribution was chosen for reasons given by Porter (2020), namely because of its simplicity, because it has only two parameters to be estimated, because it cannot take on values below zero, and because it assumes little prior knowledge. There are several decades of precedent for using the lognormal distribution for fragility analysis in earthquake engineering, and recently in analysis of railway bridge vulnerability to flood induced scour (Lamb et al., 2019).

In the following analysis, fragility functions were inferred by the method of maximum likelihood estimation (MLE) using a two-stage optimisation procedure described by Lamb et al. (2019), conditioned on data of how many leaky dams failed and how many survived during observed storm events. Three separate sets of fragility functions were estimated in this way for different locations, one set representing complete failures of leaky dams, another being for partial failures and finally one which included both complete and partial failures, as discussed in the following section. In all three cases, the functional form was the lognormal

distribution (Equation 4.2), but different values of the parameters, θ and β, were estimated.

4.2.2 Leaky dam failures

Data about leaky dam failures were collected in two ways. The first consisted of regular, detailed condition grade assessments of 26 leaky dams in a steep, headwater catchment in Coverdale, North Yorkshire. The second involved a nationwide survey of practitioners to capture information about leaky dam failure and resilience across the UK.

Coverdale study site

The Coverdale study site is located in the headwaters of the River Cover (54.20045 N, -1.98617 E) on the Eastern flank of the Yorkshire Dales National Park, North Yorkshire, England. The climate is cool and wet, with an average annual rainfall of 1270 mm (EA rain gauging station 57426 data 1988-2018). The streams are of type A in the Rosgen classification: steep, partially entrenched and cascading with step/pool streams (Rosgen, 1994). The streams have catchment areas between 0.9 and 1.7 km^2 and stream gradient ranging from 0.14 to 0.17 m/m. Land use in the catchment is pastoral agriculture on open, unimproved grassland with small amounts of coniferous plantation whilst the moorland is managed for grouse shooting.

The leaky dam network in the Coverdale study site consisted of 26 channel-spanning leaky dams constructed in autumn and winter 2018/19 on four reaches of three steep streams (Figure 4.1). The furthest upstream tributary, Lock Gill, was split into two reaches as the upper part was forested and fenced off from livestock whilst the lower reach was used for grazing. The leaky dam scheme design was developed by balancing site constraints, such as material availability and site accessibility, with water storage potential arising from the site topography. Site constraints were identified through desk-top survey, walk-over surveys, consultation with the landowner and lead local flood authority. To fulfil the consenting requirements of the lead local flood authority a flood risk impact assessment was

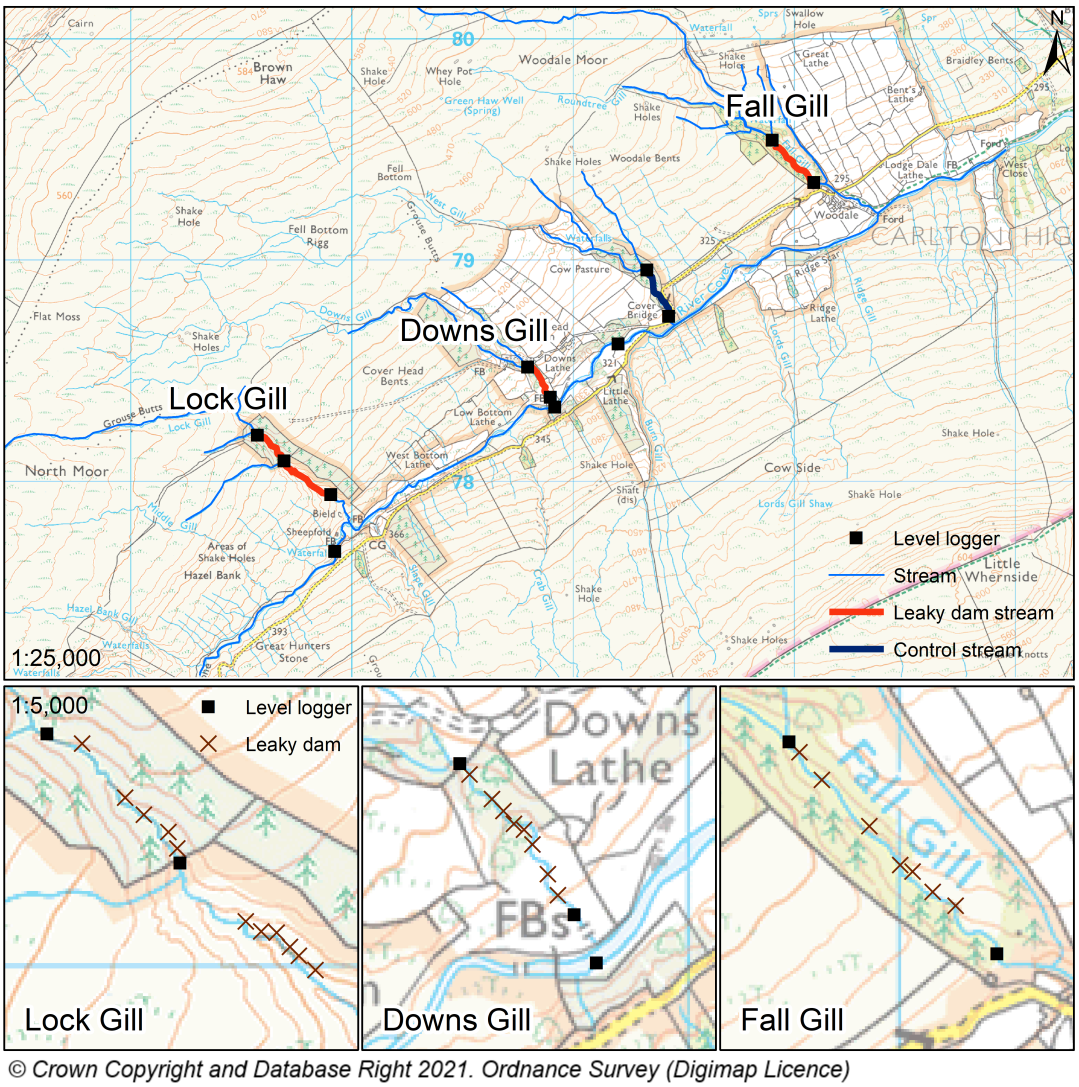


Figure 4.1: Study site map of (A) location of the study streams in the study site (B) location of leaky dams in each of the study streams

carried out which considered the risk to downstream assets of structure blockage and flood wave surge in case of failure. The final scheme in Figure 4.1 was obtained by periodically reviewing the scheme design during the build phase in response to site conditions.

Table 4.1: Average (mean) as-built dimensions (referred to in Figure 4.2) of leaky dams installed in the study streams

Stream	Lock Gill upper	Lock Gill lower	Downs Gill	Fall Gill
Count of dams	5	6	8	7
Average of W1 (m)	4.84	4.89	4.58	4.86
Average of W2 (m)	3.00	2.37	2.61	2.67
Average of H1 (m)	0.77	0.95	0.86	0.76
Average of H2 (m)	0.17	0.44	0.30	0.31
Average dam spacing (m)	33	17	25	36

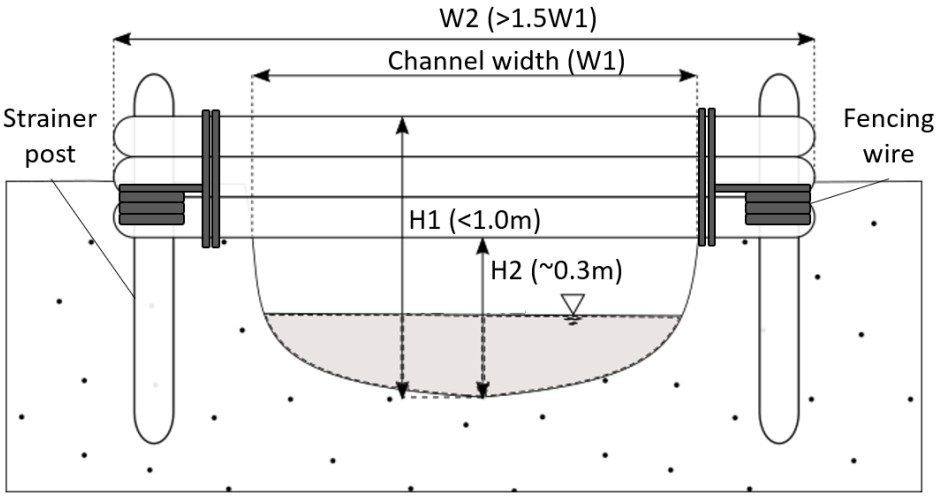


Figure 4.2: Diagram of leaky dam dimensions (referred to in Table 4.1)

The dams were built following the design guidelines of the Yorkshire Dales Rivers Trust (Yorkshire Dales Rivers Trust, 2018) by a combination of volunteers

and professionals. The design of the dams is shown in Figure 4.2, and the average as-built dimensions of the dams on each stream are given in Table 4.1. The dams were each built from 2-5 locally felled tree stems (untreated sitka spruce, beech and sycamore) with a minimum length of 1.5 times the channel width. The stems were installed to span the channel perpendicular to the direction of flow. The dams had an average height of 0.8 m above the riverbed and were installed to provide approximately 0.3 m clearance for fish passage. The dams were anchored using treated timber strainer posts and fencing wire. Brash was used to stuff the dams to enhance biodiversity impacts. The dams on Fall Gill differed from those on Lock Gill and Downs Gill because the leaky dams on Lock Gill and Downs Gill were built with the support of landscaping and forestry professionals which meant larger trunks could be felled and manoeuvred using specialist equipment.

Stage data, defined as water level above the gauge datum, was monitored upstream and downstream of the series of leaky dams on each stream at the locations shown in Figure 4.1. Stage was monitored at one-minute intervals using In-Situ Inc. (Redditch, UK) Rugged TROLL 100 non-vented pressure transducers (±0.05% full scale accuracy) in stilling wells. The pressure readings from the transducers were corrected for atmospheric pressure using an In-Situ Inc. (Redditch, UK) Rugged BaroTROLL atmospheric pressure gauge (±0.05% full scale accuracy) which was installed near the bottom of Downs Gill at a similar elevation to the non-vented pressure transducers.

The resilience of the leaky dam network was assessed through post-event walk over surveys following named storm events between September 2018 and February 2020. During the surveys, the condition of the dam was noted, including condition of anchor posts, translation or rotation of the dam, blockage of the dam by woody material, or sedimentation and scour of the bed and banks. Each dam was assigned a condition grade of good, partially failed or failed, defined as follows. A failed dam was defined as one which had been damaged to the point that it was no longer carrying out its intended purpose of spanning the channel to interfere with flood flows. A partially failed dam was a dam which had been damaged, but which was still functional and able to carry out its intended purpose. A dam which did not display any visual damage was given a condition grade of good. Blockage of a dam with woody material was not considered as failed or partially

failed as further woody material was thought to aid in interfering with flood flows. Scouring of the bed or banks was only given a condition grade of 'partially failed' or 'failed' if the erosion affected the integrity of the structure. Deposition of material, or 'regrading' of the channel upstream of the dam was classed as good because of the intended effect of increased floodplain connectivity.

Practitioner survey of UK leaky dam sites

To gather information on the frequency of dam failures across the UK, a survey was conducted asking for basic information about leaky dam projects. The survey was widely distributed to practitioners through a number of channels including social media, practitioner networks and blogs (RRC, Scottish NFM Network) and was sent out to all projects which benefited from the Environment Agency's Natural Flood Defence budget announced in the 2016 Autumn Statement. To avoid biasing responses towards practitioners who had experienced leaky dam failures the survey was positioned to relate to leaky dam resilience, rather than vulnerability. Only information about channel spanning leaky dams installed perpendicular to the flow for the purpose of natural flood management was included in the analysis.

The survey included five questions; the first three pertained to the number of dams installed, their location and the installation dates. The questions asked were::

1. Location (e.g. catchment/watercourse name/OS Grid Reference)
2. Number of leaky dams installed
3. Date of installation
4. Number of leaky dams damaged or failed, if any, and date or event during which the dam(s) were damaged.
5. Known water level (stage) or flow monitoring stations near the leaky dams (e.g. local project monitoring devices, EA gauging stations).

Survey responses were provided for fifteen projects in which leaky dams were installed (Figure 4.3). A total of 1932 dams were reported on with the number of leaky dams installed in each project ranging from 5 to 400 dams. In-depth interviews were carried out with each of the respondents who reported a leaky dam failure or partial failure to ensure the definitions of failure and partial failure used for the Coverdale study site had consistently been applied.

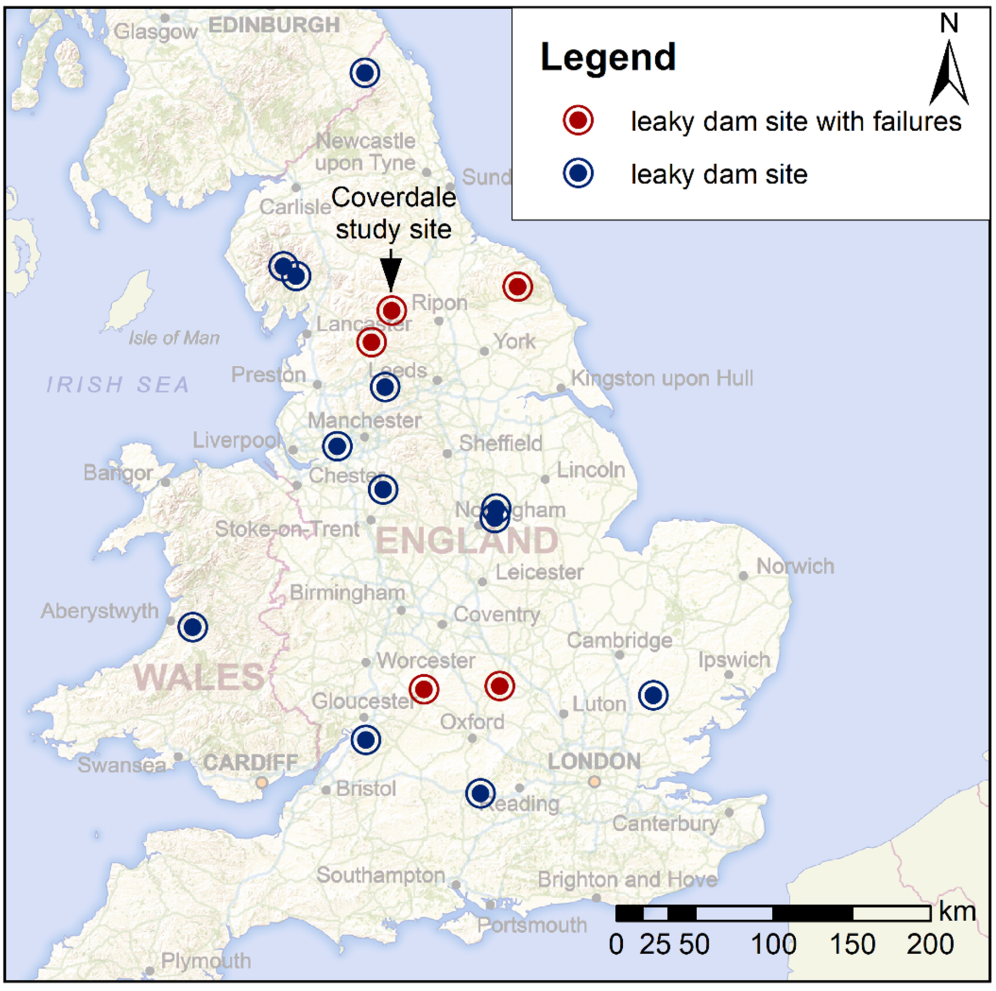


Figure 4.3: Location of leaky dam sites for which survey responses were given

4.2.3 Loading Condition

The loading condition is usually based on an observable, standardised, physical quantity that relates directly to the failure mechanism of an asset (Kim et al., 2017; Simm et al., 2009; van der Meer et al., 2008). For example, in the UK's national fragility dataset, the loading parameter for earthen fluvial levees was related to the difference between the crest level of the levee and the water level (Simm and Tarrant, 2018). However, for leaky dams, failure mechanisms are not well understood (Dixon and Sear, 2014), there is variation in the physical attributes and placement of assets, and there is a lack of structure scale measurements (Burgess-Gamble et al., 2017) which means a physical interpretation of the loading condition is not readily implemented.

After Lamb et al. (2019), the loading condition was, therefore, expressed as a relative measure of the flood event extremeness. In the closely monitored Coverdale study site it was possible to make a stream specific, normalised estimate of the severity of the flood event by calculating the z-scores of events observed over the three year study period on each of the streams. However, such site specific data was not available for the sites reported on by survey respondents and therefore, following Lamb et al. (2019), the loading condition was expressed as the flood return period in years, the inverse of annual exceedance probability. Flood return period was used because it relates to physical drivers of failure such as water depth and velocity, allowed the loading condition to be consistently estimated for all assets, and is readily interpreted and widely used in flood risk management to communicate flood frequency (Lamb et al., 2019). The availability of stream specific data from the Coverdale study site allowed for validation of the return period approach before application to the dataset obtained through the practitioner survey.

Event z-score

Stage measurements made at the upstream extent of the leaky dam streams (Figure 4.1) were used to calculate the *z-score*, a local measure of the loading condition for each stream. The peak stage for high flow events on each of the streams in the Coverdale study site was standardised by calculating the z-score

using Equation (4.3), where x_i is the event peak magnitude, $\bar{x}$ is the sample mean and σ is the sample standard deviation, all in metres. The z-scores were calculated using R v.4.0.2 (R Core Team, 2020).

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{4.3}$$

Event return period

Return period estimates were made using data from the nearest Environment Agency operated gauging station to each of the sites (Figure 4.4). The gauging stations were located between 5 and 19 km from each of the leaky dam sites (Table 4.2) and, similarly to Kay et al. (2019), were assumed to be representative of the extremeness of flows in the leaky dam sites due to their proximity.

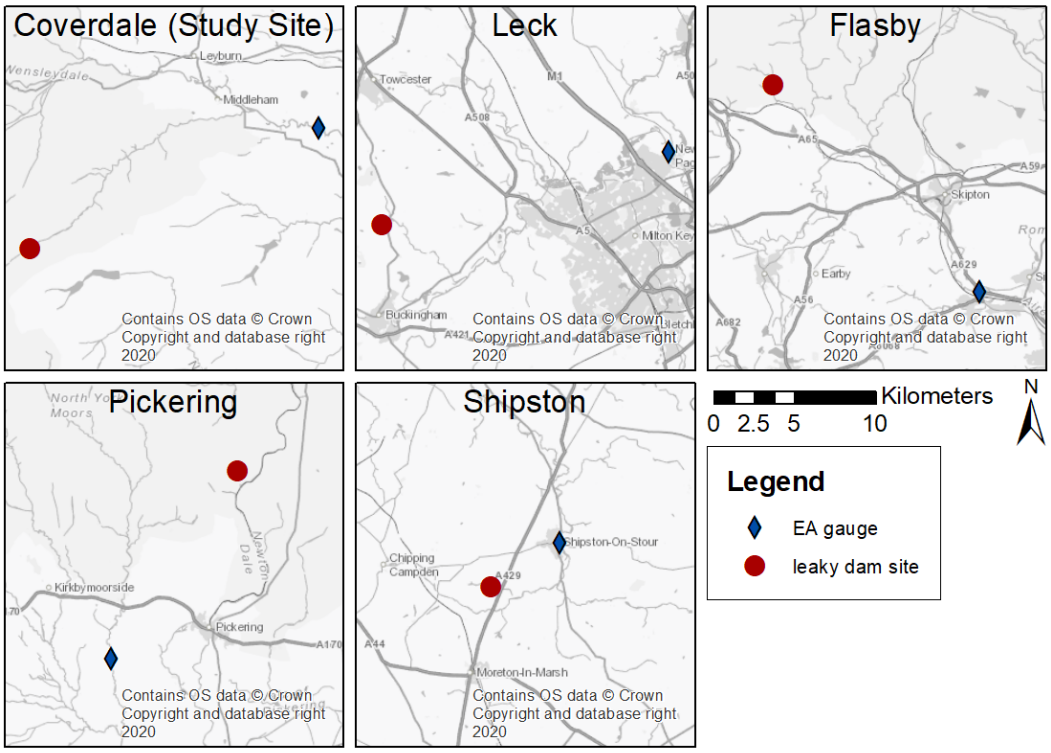


Figure 4.4: Leaky dam site locations relative to EA operated gauging stations of the five sites for which leaky dam failures were reported.

Table 4.2: Environment Agency operated gauging stations

Site	EA gauge	Distance (km)	Record length (years)	Catchment Area km^2
Coverdale	27034 Ure at Kilgram Bridge	19	52	510
Leck	33037 Bedford Ouse at Newport Pagnell	18	49	800
Flasby	27035 Aire at Kildwick Bridge	18	52	282
Pickering	27095 Pickering Beck at Pickering	14	19	66
Shipston	54106 Stour at Shipston	5	47	185

The return period estimates were made based on publicly available peak over threshold (POT) and annual maximum (AMAX) data from the UK national river flow archive (UK Centre for Ecology and Hydrology, 2021). Generalised extreme value (GEV) and Generalised Pareto distributions were fitted to the AMAX and POT series of the nearest downstream Environment Agency operated gauge with a long-term record (>30 years or longest available) using Maximum Likelihood Estimation (MLE). Following standard hydrology texts such as (Shaw et al., 2011) goodness of fit plots, goodness of fit measures (Negative log-likelihood, AIC, and BIC) and standard errors estimates were used to evaluate the fit of the distributions to the data and choose the most appropriate distribution. Although flood extremeness estimates are usually made from the annual maximum series Shaw et al. (2011) a better fit at low return periods was found for the POT series for some of the sites and the POT series was therefore identified as the most appropriate data for estimating the return period of frequent events. For the study site, where local discharge data were available, the correlation between the mean daily and maximum daily discharge measured in the study streams and at the downstream Environment Agency operated gauging station was assessed by calculating the Spearman's rank order correlation (Freedman et al., 2007). The

package extRemes v. 2.0-12 was used (Gilleland and Katz, 2016) in R v.4.0.2 (R Core Team, 2020) to fit the distributions.

4.3 Results

4.3.1 Coverdale study site

During the study period, 12 warnings of severe rain or wind weather events were given by the UK's Met Office (Met Office, 2021), five of which resulted in peak over threshold events at the nearest downstream long-term gauging station (Figure 4.5). The highest flows in the study site were recorded during Storm Gareth (16 March 2019) which was the tenth largest event on record and caused significant flooding of transport links and properties in North Yorkshire (Flood List, 2020).

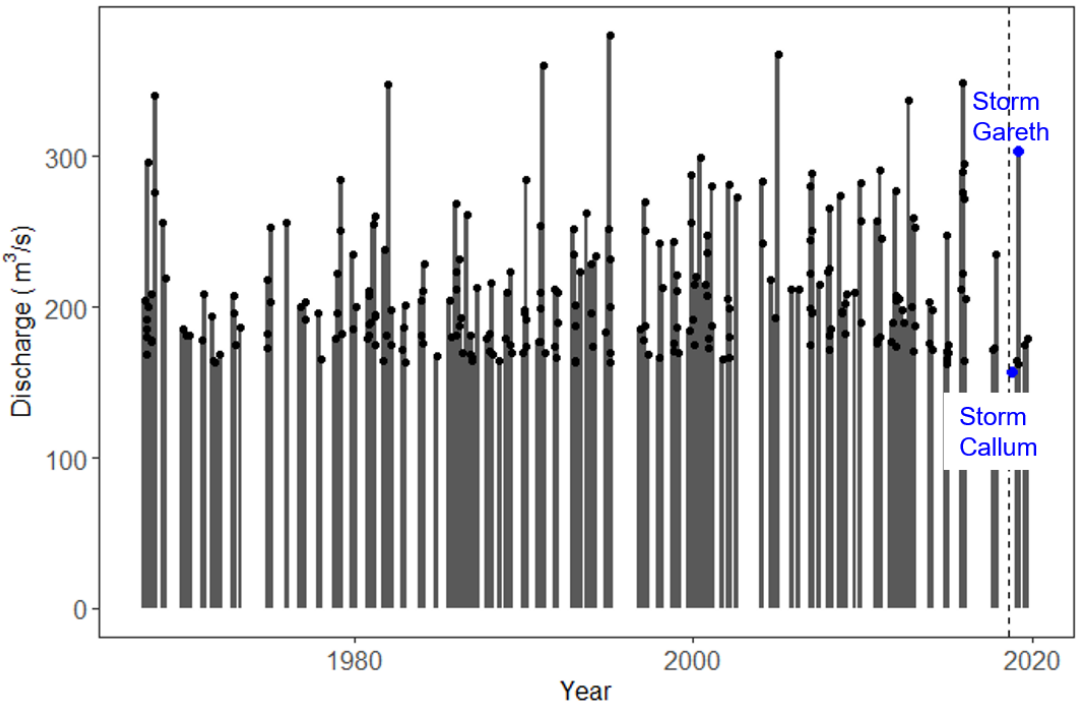


Figure 4.5: Peak over threshold series of full period of record for Kilgram gauging station, downstream of Coverdale field site. The dashed line indicates when leaky dams were installed in the study site. Two storms during which leaky dams failed in the study site are marked in blue.

Leaky dam failures

A large number of dams in the Coverdale study site were damaged to the point of partial or complete failure during named storm events in October 2018 and March 2019 (Table 4.3, Figure 4.6). In October 2018 Storm Callum damaged seven dams in total, two of which were in process of being constructed on Lock Gill. Three of the damaged dams were on Fall Gill, in which the leaky dams had been installed entirely by volunteers a month earlier. During Storm Gareth, in March 2019, a further four failures and three partial failures occurred. Two of the complete failures followed damage caused by Storm Callum.

Table 4.3: Coverdale study site leaky dam failures

		Storm Callum 1-in-1 year 31/10/2018	**Storm Gareth 1-in-6 year 16/03/2019**
Lock Gill (upper)	Failure	1	1
	Partial Failure	1	0
Lock Gill (lower)	Failure	0	0
	Partial Failure	0	1
Downs Gill	Failure	1	1
	Partial Failure	1	1
Fall Gill	Failure	0	2
	Partial Failure	3	1
Total	Failure	2	4
	Partial Failure	5	3
	Survival	19	14

Photographs of examples of failed and partially failed dams in the study site are given in Figure 4.7. Overall, two dams failed on each stream except one, the bottom reach of Lock Gill, which benefited from less clayey, more stable banks than the upper reach. All six of the completely failed dams failed by the same mechanism; failure of one of the strainer posts due to bank collapse. In each case the leaky dam remained attached to the strainer post on the opposite bank and rotated as a whole, so that the stems were more in-line with the flow, which

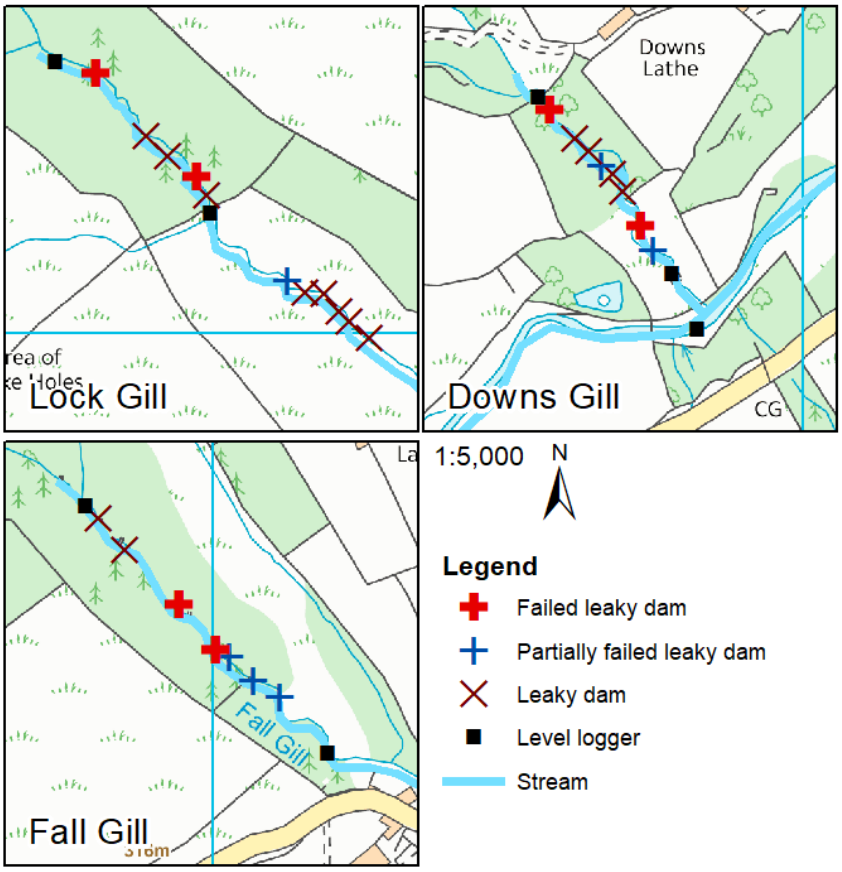


Figure 4.6: Partially failed and failed leaky dams in the three study streams of the Coverdale study site

reduced the force on the remaining strainer post. Of the eight partially failed dams, seven were classed as partially failed due to rotation or lateral movement of one of the strainer posts, and two of these later failed completely.

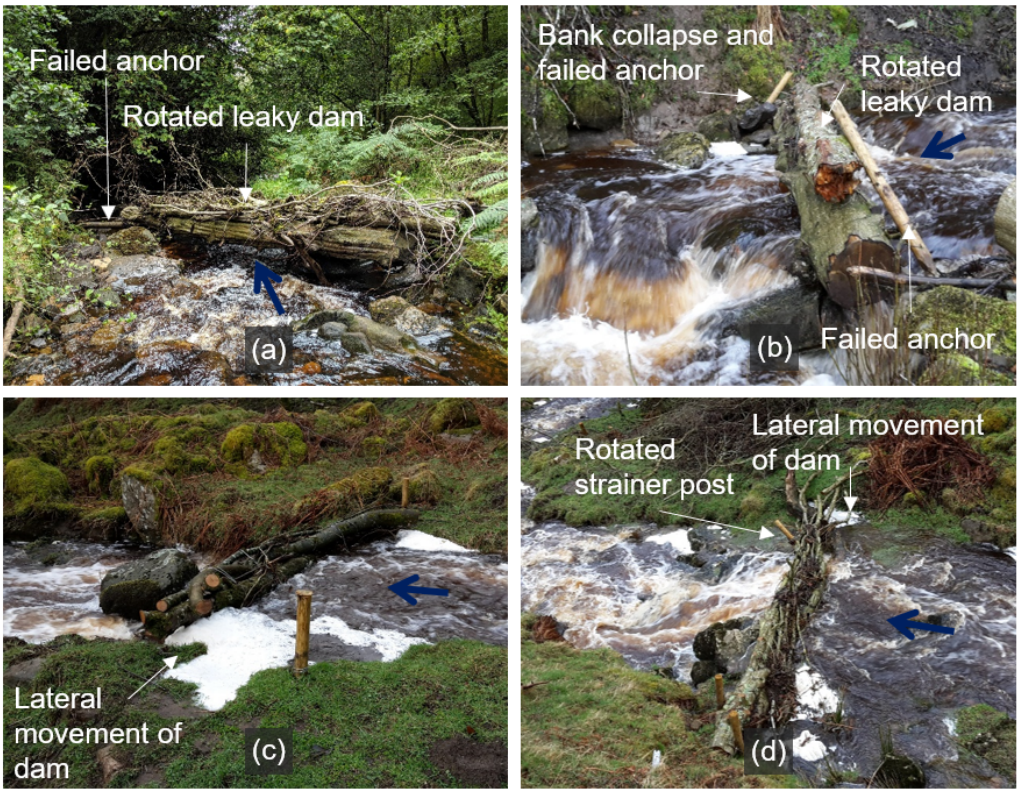


Figure 4.7: Failed (top row) and partially failed (bottom row) leaky dams in Coverdale study site. The blue arrows indicate the general direction of flow of the stream.

Z-score loading condition

The failures in the Coverdale study site happened during two events; Storm Callum in October 2018 and Storm Gareth in March 2019. Storm Callum was one of the larger events recorded during the study period but had a lower event peak than Storm Gareth which was the largest event recorded in Coverdale during the study period. Other events of a similar magnitude were recorded during the

study period on Downs gill and Fall gill, none of which damaged the leaky dam network (Figure 4.8). Z-scores for the same event differed between streams, the z-scores for Lock Gill were generally higher than those on Downs Gill and Fall Gill. On Lock Gill Storm Callum and Storm Gareth had the second highest and highest z-scores respectively.

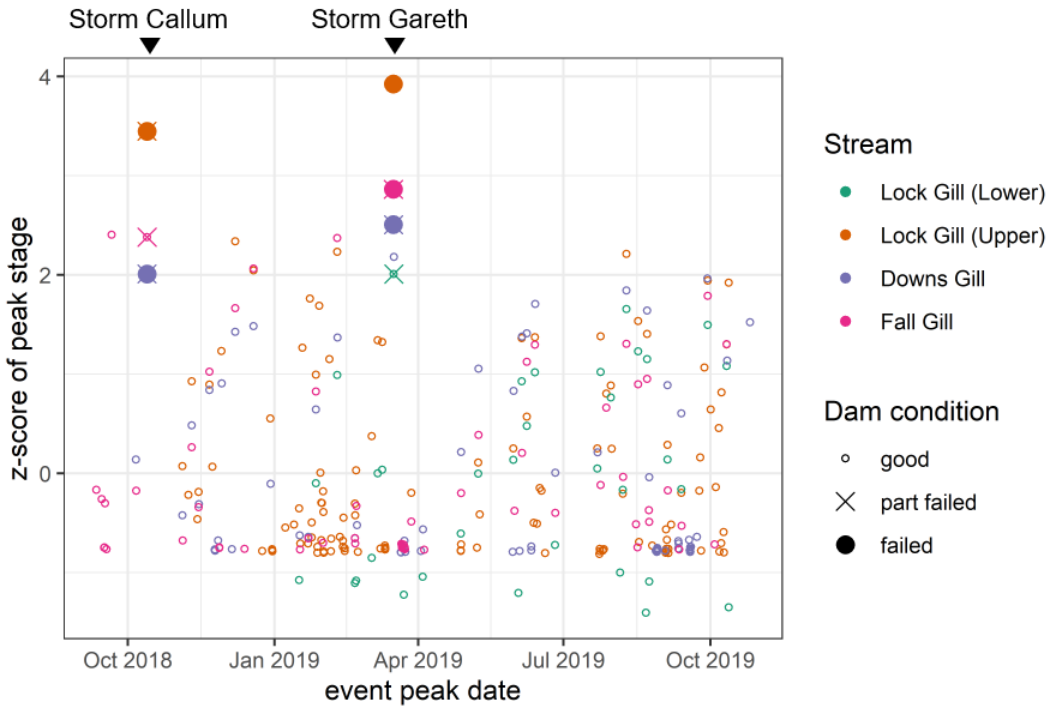


Figure 4.8: Standardised peak event stage (z-score) for each of the POT events in the post-intervention monitoring period

Return period loading condition

The downstream EA gauge record had a strong positive correlation with the discharge measured in the study streams (Spearman's correlation coefficient for control stream is 0.89, p <0.01) (Figure 4.9). Results for Lock gill were not included in Figure 4.9 because rating relationships were not available for the gauging sites on this stream. The control stream tended to have the lowest mean daily discharge and Downs Gill tended to have the highest mean daily discharge

for a given mean daily discharge at the long-term gauge. For more extreme events (Kilgram Bridge discharge >150 m^3/s) the magnitude of the maximum daily discharge on the three streams converged indicating that during Storm Callum (157 m^3/s peak flow) and Storm Gareth (303 m^3/s peak flow) the dams on different streams experienced similar loading conditions.

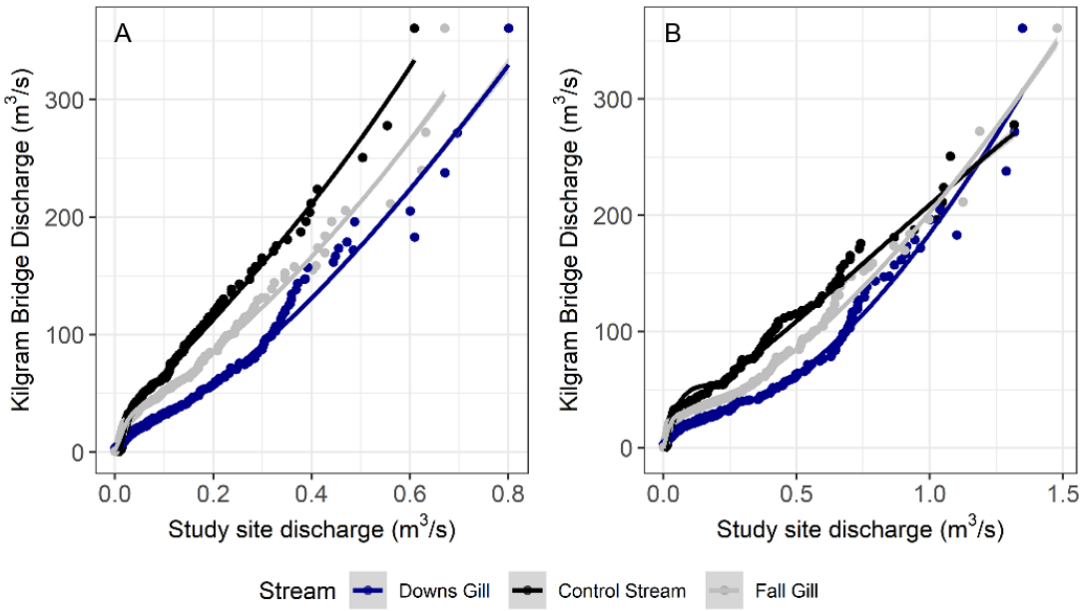


Figure 4.9: Discharge recorded at the Coverdale stream reaches compared with that of the EA gauge at Kilgram Bridge: (A) mean daily discharge, (B) maximum daily discharge with LOESS smooth.

Flood frequency estimates (Appendix C) indicate that the magnitude of the events in the POT series observed in the study streams ranged from <1 year return period event to a 6-year return period event. The return period for Storm Callum was estimated 1 year, and for Storm Gareth as 6 years from the long-term record.

4.3.2 UK leaky dam sites

One or more complete or partial failures of leaky dams was reported by five of the survey respondents (Table 4.4). In total 23 dams failed, and 32 dams partially

failed, or 1-2% of all dams. All the information required for analysis (answers to all five questions in Section 4.2.2) was provided for all of the failed dams and 24 of the 32 partially failed dams. Both complete and partial failures occurred during storms with annual probabilities ranging from 1-in-13 to almost 1-in-1. Of the 47 dams which completely or partially failed, 40 (85%) were damaged during named storm events, meaning the storm was assigned an amber or red weather warning by the UK Met Office's national severe weather warning service (Met Office, 2021). In each project the majority (94%) of dams were resilient to the event peaks they were exposed to. The most extreme events which affected the sites ranged from 1-in-4 to 1-in-33 annual probability events. Table 4.4 summarises the observations and associated return period estimate for the five sites in which partial or complete failures were reported. Details of the return period estimation are given in Appendix C.

Table 4.4: Complete and Partial Failures of leaky dams in five UK NFM projects. Where applicable, numbers in brackets in the storm name column indicate its POT ranking at the gauging station. (RP = Return Period)

	Date	Storm name	Type	Number of dams	RP (years)
Coverdale	**Oct 2018**	**-**	**Installation**	**26**	**-**
	13/10/2018	Callum	Failure	2	1.04
			Partial Failure	6	1.04
	16/03/2019	Gareth (10)	Failure	4	6.02
			Partial Failure	3	6.02
	16/03/2019	Gareth	Survival	14	6.02
Leck	**Summer 2019**	**-**	**Installation**	**24**	**-**
	16/01/2020	Brendan (42)	Failure	1	5.18
			Partial Failure	0	5.18
	16/11/2019	(45)	Failure	0	2.62
			Partial Failure	2	2.62
	26/11/2019	Gareth	Survival	21	10.20
Flasby	**Early 2019**	**-**	**Installation**	**135**	**-**
	17/03/2019	Gareth (14)	Failure	10	11.26
			Partial Failure	0	11.26
	31/10/2019	-	Survival	125	32.72
Pickering	**Jun 2010**	**-**	**Installation**	**167**	**-**
	27/11/2012	(8)	Failure	1	12.94
			Partial Failure	1	12.94
	26/12/2015	Eva (10)	Failure	3	9.54
			Partial Failure	11	9.54
	27/11/2012	(8)	Survival	151	12.94
Shipston	**August 2017**	**-**	**Installation**	**400**	**-**
	14/11/2019	(5)	Failure	2	4.12
			Partial Failure	1	4.12
	14/11/2019	(5)	Survival	397	4.12

4.3.3 Fragility Analysis

The fragility functions for the Coverdale field site estimated using normalised stream peak stage (z-scores) as the loading condition (Figure 4.10) and the fragility functions estimated using the event return period from the nearest long-term EA operated gauging station (Figure 4.11) were sufficiently similar to increase confidence in the use of return period estimated from a nearby EA gauging station as the loading condition. Fragility functions estimated based on return period loading for each site in which leaky dam failures had been reported (Figure 4.12) showed there was variability in conditional failure probability between sites. Finally, the pooled fragility function based on data from all five sites combined (Figure 4.13) serves as a proof of concept for a national estimate of leaky dam fragility which could be extended and updated in future studies to build up regional or national fragility functions. The fragility functions were estimated for complete failures, partial failures and either complete or partial failures for each site. The confidence intervals of the fragility functions overlapped considerably, therefore it was not possible to assert whether differences in the fragility functions for partial and complete failures were significant (Figure 4.10 - 4.13). Extrapolation of the fragility functions beyond the highest observed loading condition on each site was reflected in increasing width of the confidence intervals for higher loading conditions.

Coverdale study site

There was considerable overlap in the confidence intervals of the three fragility functions for the Coverdale study site, reflecting the variability in failure loads (z-score) and lack of data at higher z-scores (Figure 4.10). Although differences between the fragility functions may not be significant because of overlapping confidence intervals, there was an indication that for a given loading condition the likelihood of observing a partial failure was higher than observing a complete failure and observing either a partial or complete failure was most likely, as expected. The normalised peak stage (z-score) for the events during which leaky dams failed ranged between two and four on the streams and this is reflected in the form of the fragility functions (Figure 4.10). The probability of failure for events with a

z-score below two was close to zero, rising to >0.75 for events with a z-score higher than four. For events with a z-score of six, the probability of failure approached one, with increasing uncertainty, which reflects the lack of data for events with a z-score beyond four. This served as a 'sense check' which indicated that the fragility functions estimated from standardised event peak stage (z-scores) measured in the streams did represent the asset failure probability observed in the Coverdale study site.

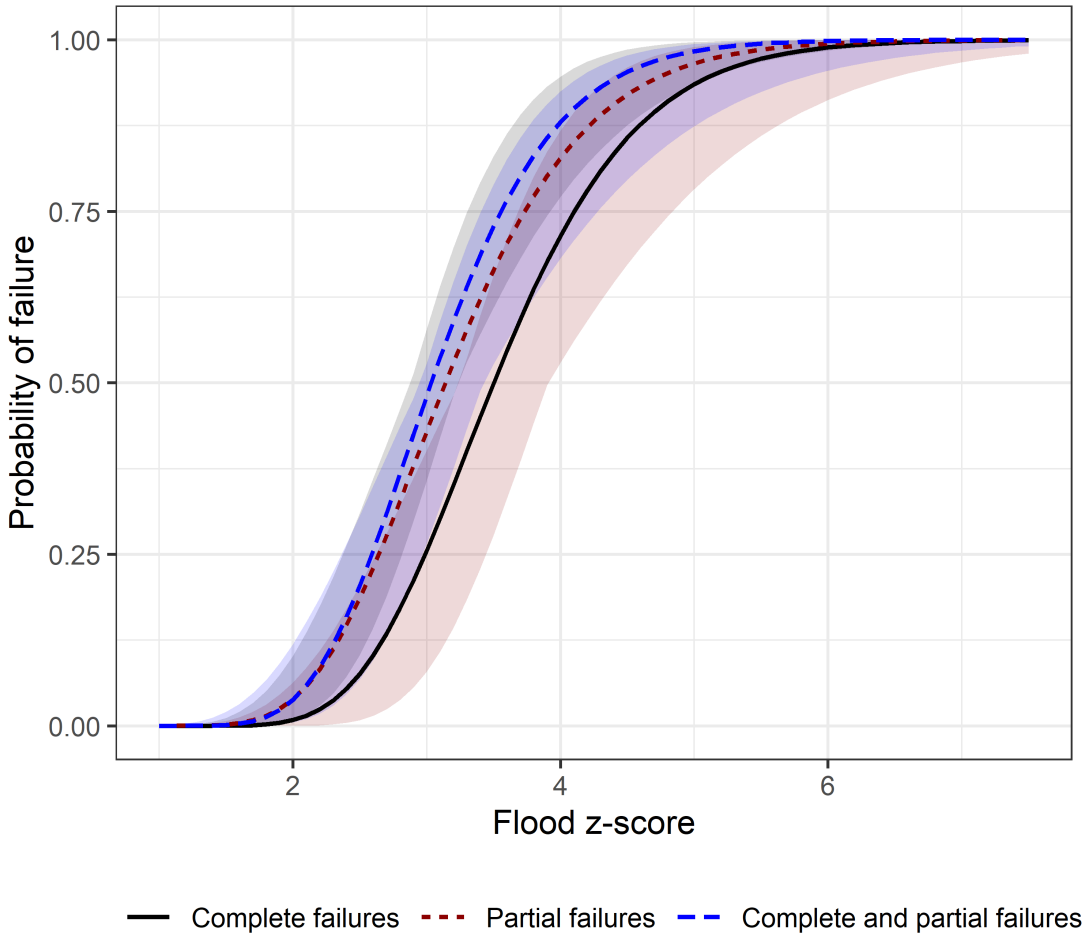


Figure 4.10: Fragility functions for Coverdale study site based on local loading condition (z-score). The shaded regions indicate the confidence interval based on the likelihood surface. The fragility curve for complete and partial failures indicates the probability of either a complete or partial failure occurring.

The fragility curves based on return period estimated from the nearest EA operated gauging station (Figure 4.11) displayed a similar inter-relationship as the fragility curves based on the locally derived z-scores (Figure 4.10). The events during which leaky dam failures occurred had an event return period of 1 year (Storm Callum) and 6 years (Storm Gareth), which is reflected in the shape of the fragility functions in Figure 4.11. The probability of failure of a dam in the network was low during frequent, higher annual probability events but increased for less frequent events, with a probability of failure of 0.47 for a 1-in-10 year event, or 0.68 for either a partial or complete failure (Table 4.5). During a 1-in-100 year event the probability of failure was close to 1, indicating that a dam in the Coverdale study site would be very likely to fail during such an extreme event. As expected, the fragility functions reflect that partial failure was more likely than complete failure for frequent events, but during more extreme events (return period >34 years) complete failure was more probable than partial failure. However, there is substantial overlap in the confidence intervals of the fragility curves for complete and partial failures and therefore it is not possible to say that these differences were significant. The sense checks increased confidence in the use of return period derived from the nearest EA operated gauge as the loading condition.

Table 4.5: Central estimates of probability of complete, partial, or either failure for a dam in the Coverdale study site

	Return Period (years)					
	1	5	10	20	50	100
Complete failure	0.02	0.25	0.47	0.70	0.90	0.97
Partial failure	0.09	0.39	0.57	0.73	0.89	0.95
Complete or Partial failure	0.09	0.47	0.68	0.85	0.96	0.99

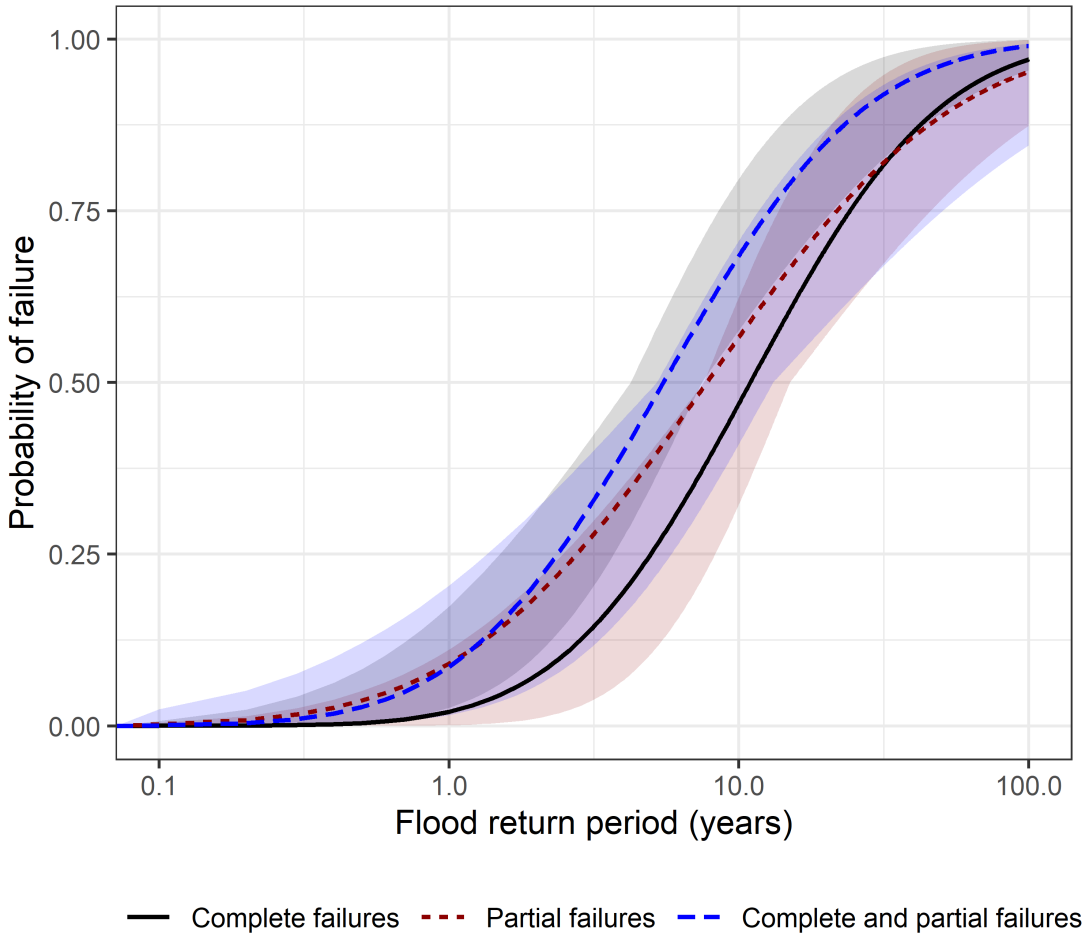


Figure 4.11: Fragility function for Coverdale study site based on nearest EA gauge data with shaded confidence region for complete, partial, and combined failures on the field site. The fragility curve for complete and partial failures indicates the probability of either a complete or partial failure occurring

UK leaky dam sites

The fragility functions were estimated for each of the sites individually (Figure 4.12) to put the failures observed in the Coverdale study site into context. In context of the other sites, the Coverdale study site dams generally had a higher probability of failure, especially during frequent events (<10 year return period) (Table 4.6). For floods with a return period below 10 years the probability of failure for the UK leaky dams sites was a magnitude lower (no more than 0.05) than on the Coverdale field site, where probability of failure was 0.5 for a 1-in-10 year event.

The dams in the other sites were more resilient (failure probability no more than 0.12) even for 1-in-20 year annual exceedance probability events, but failure probability increased considerably for 1-in-50 year events in some sites (no more than 0.67 in Pickering), but remained lower in Flasby (0.11), where the majority of dams were resilient to a 1-in-33 year probability event. For the most extreme loading condition considered, the 1-in-100 year flood, the fragility functions for Pickering and the Coverdale study site indicated a probability of failure near 1, but in Flasby and Leck this was lower at 0.19 and 0.46 respectively (Table 4.6).

The gradient of the fragility functions reflected the certainty about the loading condition at which the leaky dams were able to resist failure. The fragility function for Pickering was steeper than the other three sites because failures occurred during events of similar magnitude in Pickering (10 and 13 year return period). Although 10 failures occurred during just one, 11 year return period event, in Flasby, the majority of leaky dams were resilient to a much more extreme event (33 year return period) and therefore the fragility function was shallow.

On all sites the probability of any damage occurring (partial or complete failure) was highest, as expected, and approached 1.0 for a 1-in-100 year (or greater) event. Uncertainty due to extrapolation beyond observed events was reflected in the width of the confidence intervals. In the Coverdale study site and Leck there was a suggestion that the probability of observing a complete failure became higher than observing a partial failure at a return period of 34 and 31 years respectively, although overlap of the confidence intervals means that this effect was not conclusive. For Flasby only the fragility function for

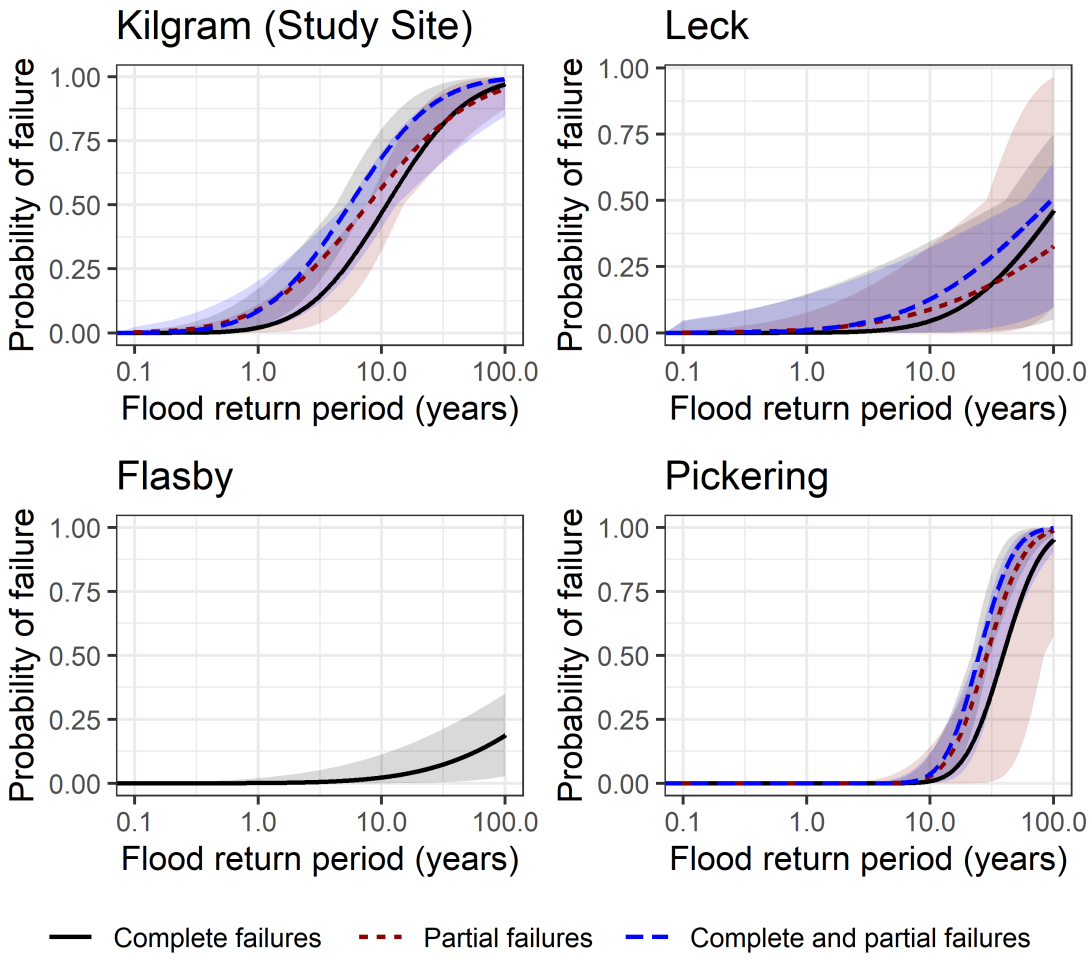


Figure 4.12: Fragility curves for each catchment. Shipston is not included because the fragility function could not be solved for only one magnitude of the loading condition. The shaded regions indicate the confidence interval based on the likelihood surface. The fragility curve for complete and partial failures indicates the probability of either a complete or partial failure occurring.

complete failures was solved because there were no partial failures on the site. The fragility function for Shipston could not be solved because the loading condition for failed and surviving dams was the same, but the data was included in the pooled analysis.

Pooled analysis

The pooled fragility functions were estimated by combining data from all five sites in which failures were reported (Figure 4.13). The fragility functions of combined data serve as a proof of concept pilot study of a pooled leaky dam failure curve, which could be extended and updated in future studies to build up regional or national fragility functions. The pooled fragility functions were less steep than the fragility functions for individual sites, reflecting the larger range of events during which leaky dams failed and survived. The pooled fragility curves indicate a considerably lower failure probability than was observed at Coverdale for all loading conditions (Table 4.6). Although there was considerable overlap of the confidence intervals, there was an indication that the probability of observing a complete failure was higher than that of observing a partial failure for events with a return period above 17 years. The fragility curves indicate that, for the pooled analysis across multiple NFM sites, the probability of complete or partial failure of a leaky dam in a 1-in-1 year event was close to zero. For a 1-in-100 year event the probability of a complete or partial failure was estimated to be 0.33 (within lower and upper confidence intervals of 0.24 and 0.40).

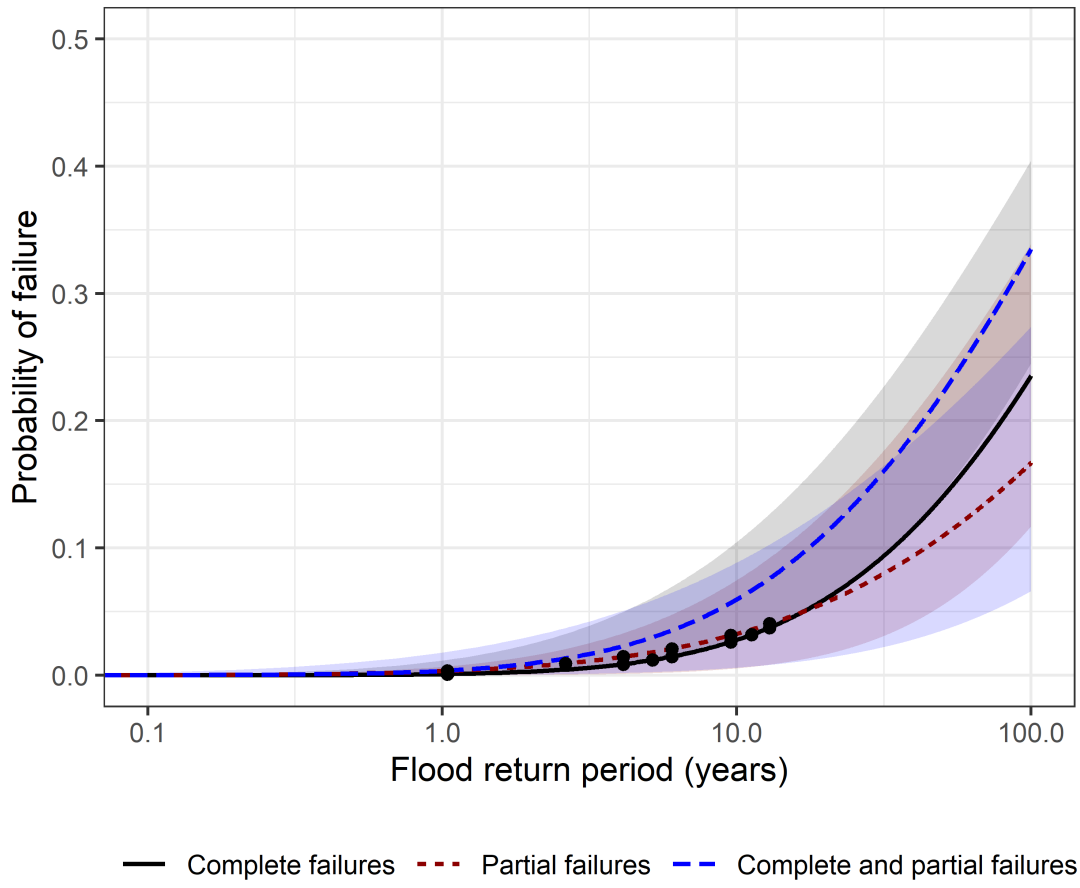


Figure 4.13: Fragility curves for failure data pooled from all sites (note that the probability of failure on the y-axis is up to 0.5), points represent the estimated loads at which complete or partial failures were observed.

Table 4.6: Central estimates of probability of complete, partial or either failure for all five UK leaky dam sites

Return period (years)	**1**	**5**	**10**	**20**	**50**	**100**
Site	**Probability of complete failure**					
Coverdale	0.02	0.25	0.47	0.70	0.90	0.97
Leck	0.00	0.01	0.05	0.11	0.28	0.46
Flasby	0.00	0.01	0.02	0.05	0.11	0.19
Pickering	0.00	0.00	0.01	0.12	0.67	0.95
Pooled	0.00	0.01	0.03	0.06	0.14	0.24
Site	**Probability of partial failure**					
Coverdale	0.09	0.39	0.57	0.73	0.89	0.95
Leck	0.01	0.05	0.09	0.14	0.24	0.33
Pickering	0.00	0.00	0.03	0.26	0.84	0.99
Pooled	0.00	0.02	0.03	0.06	0.11	0.17
Site	**Probability of complete or partial failure**					
Coverdale	0.09	0.47	0.68	0.85	0.96	0.99
Leck	0.01	0.07	0.13	0.22	0.37	0.51
Pickering	0.00	0.00	0.04	0.34	0.92	1.00
Pooled	0.00	0.03	0.06	0.11	0.22	0.33

4.4 Discussion

Leaky dams are increasingly being installed in rivers for the purpose of flood risk management (Grabowski et al., 2019). Whilst efforts are being made to quantify their benefits, little is known about the residual risks associated with their potential failure, even though some failures have been observed. Failure of leaky dams can undermine the intended performance of a flood risk management scheme and can pose hazards to property, infrastructure and people (Wohl et al., 2016). Furthermore there are concerns that cascade failure of leaky dams could lead to a surge in water level as stored water is released simultaneously (Hankin et al., 2020). A better understanding of leaky dam resilience is required to inform risk analysis, network design and investment decisions from local to national levels including the programming of maintenance. Observations of leaky dam resilience

to a range of loading conditions are emerging from leaky dam sites around the UK. This research provides a proof-of-concept for the use of empirical fragility functions to quantify system risk at both local scale and pooled across multiple sites.

4.4.1 Individual site resilience

UK practitioners reported that less than 2% of leaky dams had failed in 15 UK sites covering almost 2000 leaky dams, the first of which were installed in 2010. Complete and/or partial failures were reported for 5 sites, including the Coverdale study site which was closely monitored. The fragility functions for the failure sites showed that the loading conditions under which leaky dams failed were highly variable between and within sites, although it is not yet clear how much this reflects real differences in resilience as opposed to sampling effects. The leaky dams in the Coverdale study site had a relatively high probability of failure compared to the other sites which can be attributed to: (1) exposure of partially constructed dams to loading; (2) unstable banks presenting difficulties in anchoring leaky dams; (3) site conditions which precluded access with machinery and manoeuvring of larger stems; (4) steep stream gradient leading to high stream power per unit discharge. The failed leaky dams in the Coverdale study site all resulted from failure of one anchor. Although there was a high rate of failure the leaky dams were not mobilised as they remained tethered to the bank by the remaining anchor. The failed dams rotated in line with the flow, which reduced the forces exerted on the remaining anchor, and the likelihood of the dam being transported downstream.

A trial and error approach has been taken to the design of leaky dams for natural flood management in the UK (Hankin et al., 2020) with 'rule of thumb' guidelines about the spacing and size of leaky dams being taken from the large wood mobility literature (Herrera Environmental Consultants, 2006; Linstead and Gurnell, 1999; Thomas and Nisbet, 2007). For example, leaky dams are recommended to be 1.5 times the channel width in best practice guidelines (Saldi-Caromile et al., 2004; Yorkshire Dales Rivers Trust, 2018) based on observations of stable key pieces of natural wood accumulations (Abbe et al., 2003; Keller

and Swanson, 1979). However, observations from the Coverdale study site show that this recommendation does not necessarily prevent failure of upland leaky dams. In steep streams the length, diameter and strength of key pieces determine the stability of leaky dams (Abbe, 2000). Wood with a length 2.5 times the channel width was functionally immobile in a low gradient, headwater river in the UK (Dixon and Sear, 2014), but this ratio may be difficult to achieve in practice because of site and material limitations. For example, upland, incised streams can have steep banks and be inaccessible to machinery. The presence of a rootwad has been identified as a key factor in determining large wood stability, particularly when the rootwad faces upstream, which can stabilise the large wood by burying the rootwad over time (Abbe et al., 2003). Whilst this type of large wood is associated with river restoration activities, it has been demonstrated to increase floodplain connectivity and has a similar impact on flood peak magnitude as leaky dams in one of the steep Coverdale study site streams (Keys et al., 2018).

The observations from the Coverdale study site imply that leaky dam good practice guidance could be built upon by incorporating perspectives from the river restoration and wood mobility literature. For example, Abbe and Brooks (2011) give guidelines for geomorphic, ecological and engineering considerations of placing large wood in streams with emphasis on stability, and (Ruiz-Villanueva et al., 2014) provide a tool for assessing risks associated with reintroducing large wood to rivers, including consideration of the geomorphological and ecological risks associated with not restoring it.

4.4.2 Pooled site resilience

Pooling information from geographically disperse data has been done previously to obtain empirical fragility functions for assets such as railway bridges (Lamb et al., 2019), overhead line components of electricity distribution networks (Dunn et al., 2018), and expressway structures (Yamazaki et al., 1997). In this study, fragility analysis was applied successfully to pooled data of leaky dam complete failures, partial failures and resilience to assess the probability of failure during flood events with a return period of up to 100 years. Failure probability varied

between sites but, by pooling the data, an estimate of the resilience of dams across the five sites was made. For a 10 year return period event the probability of partial or complete failure for the five sites combined was 6% increasing to 11%, 22% and 33% for the 20, 50 and 100 year return period events.

The fragility functions were conservative because only data from sites in which failures occurred were included in the analysis. Including data for sites in which no failures occurred, such as the 10 sites included in the survey responses, would likely considerably decrease the failure probabilities. Including this data was beyond the scope of this proof of concept study but is recommended for future work. This means that the estimates are conditional fragility functions for the probability of failure at sites where at least one failure has already occurred. If most leaky dam sites never experience a leaky dam failure these fragility functions will over-estimate leaky dam failure probability. However, given the recent installation of leaky dams for NFM, the stochastic nature of flood events and the return periods at which leaky dam failures have been observed, it is plausible that leaky dam failures will occur in other leaky dam sites. By repeating the analysis with data from more sites as data emerges, the fragility functions will stabilise over time and the number of sites needed to obtain a representative sample will become apparent. However, even though this was a pilot study, responses to the informal survey already encompassed observations from 2000 leaky dams (failed and non-failed), demonstrating both the feasibility of the analysis at scale and the practitioner community's interest. By including data from more UK leaky dam sites, a national estimate of leaky dam fragility could readily be made using the methodology demonstrated in this study, and progressively improved as observations of more flood events materialise over time. Crucially, this improvement should include better quantification of the number of leaky dams that do not fail.

It is inconsequential to compare the vulnerability of leaky dams with that of traditional, engineered flood defence assets because engineered structures have been designed to withstand higher loads. For example, the probability of scour induced failure of a railway bridge exposed to a 1-in-100 year event is small, and less than 0.1 even for a 1-in-1000 year flood event (Lamb et al., 2019). Furthermore, fragility functions for flood defence assets in the UK are developed using an analytical, rather than empirical approach (Schultz et al., 2010). The

analytical approach involves identifying and modelling all possible failure modes and performing a series of reliability analyses (Simm and Tarrant, 2018). The loading condition is related to the failure process, such as asset overtopping rate, rather than the flood extremeness estimate (Simm and Tarrant, 2018) and so the fragility functions are not directly comparable.

The failure rates of leaky dams are, instead, briefly compared to rates of large wood mobility in small streams. Failure of leaky dams is related to large wood mobility firstly in that the dams were mobilised to induce failure, and secondly, because a number of the hazards associated with leaky dam failure, namely downstream blockage, require the leaky dams to be mobilised and transported after failure. In small streams, log jams typically form which consist of at least one key piece which is longer than the channel width (Abbe et al., 2003) and mobility is influenced by extreme flood events, rather than being part of a continuous process (Jochner et al., 2015), which means they are rarely transported (Bilby and Ward, 1989; Nakamura and Swanson, 1993). In a 1-in-50 year flood event all large wood was removed from a steep, headwater stream in the Swiss pre-alps, whilst a 1-in-20 year event removed only some of the 9 log jams which had formed since. The residence time of large wood in the stream was between 1 and 13 years with a mean of 4 years. Similarly, an incised mountain channel in Poland retained only 12.5% of large wood in a 1-in-20 year event (Wyżga et al., 2017), and 75% of large wood was found to be mobile in a low gradient headwater stream in the UK (Dixon and Sear, 2014). The rates of mobilisation are similar to leaky dam failure rates observed on the Coverdale study site (failure probability 0.7 for a 1-in-20 year event, and 0.97 for a 1-in-50 year event) but pooled analysis suggests that leaky dams are typically considerably more resilient, with a failure probability of <0.01 for a 1-in-4 year event, 0.06 for a 1-in-20 year event and only 0.14 in a 1-in-50 year event. Hence, as expected, in general the fragility analysis implied that although leaky dams can be expected to fail, they are less likely to fail or be mobilised than natural log jams.

4.4.3 Implications

The fragility functions developed for pooled site data may be a reasonable first indication of failure probabilities for this type of leaky dam across the UK, although that estimate should be refined and progressively updated by adding data from more sites and events as it emerges. The relative failure probabilities can be used by practitioners to increase confidence in budgeting and planning of inspection and maintenance of new and existing leaky dam schemes. As more resilience data from leaky dam sites across the UK emerges, the fragility functions can be extended and updated to form a robust regional or national picture of leaky dam resilience.

Fragility functions have been used in flood risk assessment in the UK since 2002 (Simm and Tarrant, 2018). Rather than just having one fragility function for each class of flood defence asset the fragility functions have been estimated for five condition grades of each flood defence asset class. This allows the fragility functions to be used to quantify the flood risk of the system in its current, improved or deteriorated state to inform investment decisions and programming of maintenance at both local and national scales (Simm and Tarrant, 2018). As leaky dam resilience data emerges, empirical fragility functions for different leaky dam types, condition grades and site conditions can be estimated, which would allow them to be integrated in the cost, benefit and risk assessment frameworks of traditional, engineered defences, a crucial step in aligning them to flood risk management practice (Hankin et al., 2020). Fragility functions can also be estimated using a hybrid approach in which understanding of failure mechanisms is used to develop fragility functions with an analytical approach and then validated using empirical observations (Pregnolato et al., 2015).

Currently, probabilistic reliability assessment methods are only used to prioritise investment (Simm and Tarrant, 2018), but the large uncertainty associated with NFM presents challenges for deterministic design methods (Addy and Wilkinson, 2019) and, therefore, the probabilistic framework is a useful tool for design and optioneering (Hankin et al., 2020). For example, Hankin et al. (2020) showed that cascade failure of leaky dams was more likely in some configurations than others using a system performance model. Although the authors concluded

that failure risk can be a more important factor to consider than impact on the flood hydrograph for placement decisions, the fragility of the leaky dams in the model was based on a critical state assumption, rather than empirical data. The empirical fragility functions developed in this study, and national estimates for different types of leaky dams, would help to constrain such an analysis so that the flood risk and resilience of leaky dam systems can be assessed with more confidence.

4.4.4 Limitations and further work

Generic, national fragility curves developed for flood defence assets are not intended to be used to assess the performance of individual assets, but rather to inform national scale investment decisions (Simm and Tarrant, 2018). However, empirical fragility functions from various disciplines have been criticised for being specific to structures and situations (Pregnolato et al., 2015; Schultz et al., 2010; Simm and Tarrant, 2018) because their use requires the assumption that past damage of an asset type is representative of future damage of that asset type exposed to the same loading (Pregnolato et al., 2015). This can lead to difficulties in generalising between sites and does not consider variability in the resilience of leaky dams within a site. In the Coverdale study site, Flasby, and Pickering, for example, the leaky dam design was tweaked following failures, and dams were re-built stronger (T Nisbet, 2020, personal communication, 8 January; D Vine, 2020, personal communication, 17 January). As more data emerges it may be possible to estimate fragility functions for different types of leaky dams, site conditions and age, or condition of the leaky dams. Variation in the resilience of leaky dams over time due to deterioration of the wood, for example, could be accounted for by developing fragility functions for different condition grades, in the same way they are for traditional flood defence assets (Simm and Tarrant, 2018). Collecting data about the streams, such as the catchment area and stream gradient, would provide the opportunity to contextualise the observed variability in leaky dam resilience. A hybrid approach in which analytical fragility functions are validated using empirical data may be needed to reach this level of detail. Once developed, the fragility functions can be incorporated in a hydraulic based

system analysis model which will provide information about the reliability of the system and allow dam failure to be incorporated in cost-benefit and risk analysis.

To increase confidence at higher return periods, data from more extreme events should be included in the analysis. The fragility functions were extrapolated beyond the observed data, particularly in the Coverdale study site where the most extreme event was a 1-in-6 year event, but network failure probability was estimated for a loading condition up to a 1-in-100 year event. Uncertainty associated with the extrapolation was captured in the confidence intervals of the fragility functions and would be reduced by including more data. Resilience of the majority of leaky dams to a more extreme event (1-in-33 year) in Flasby reduced the failure probability for a given load predicted by the fragility function considerably compared to the other sites. More data from a wide range of loading conditions is therefore required to increase confidence in the conditional failure probabilities. Due to the stochastic nature of flood events more time may be needed before these data are available. The study demonstrated that fragility functions can be estimated for pooled leaky dam sites but to make generalisable national estimates of the fragility functions, data from more sites across the UK is needed. For the fragility functions not to be biased these data should include a representative sample of all sites in which no failures have occurred, as well as a representative sample of sites in which failures have occurred.

4.5 Conclusion

Leaky dams installed for the purpose of NFM have been observed to fail in the Coverdale study site and elsewhere (Hankin et al., 2020). However, the number of leaky dams which have completely or partially failed is relatively low compared with the number of leaky dams which have been installed in the UK. This study demonstrates that observations of leaky dam failure and resilience can be used to derive quantitative fragility functions which describe the failure probability of leaky dams conditional on flood event return period. The probability of complete or partial failure of leaky dams was 0.03 in a frequent (5 year return period) event and increased to 0.33 for a 1-in-100 year event. The estimates were conservative because only data from sites in which failures were observed were included, but

could be readily improved using the same methods by including data from more leaky dam sites and more flood events. This data will become available in due course as it forms part of the monitoring requirements of all Defra funded NFM projects (Arnott et al., 2018). Although challenges remain regarding generalisability of fragility functions and data availability, this study provides a proof of concept for the use of empirical fragility functions to assess leaky dam system vulnerability and resilience to extreme flood events at a regional and national scale.

Chapter 5

Discussion & Conclusion

5.1 Research Summary

The management of wood in rivers requires the benefits provided for downstream flood risk to be weighed up against the potential for large wood to become mobile and present hazards (Wohl et al., 2016). However, there is little empirical evidence of the impact of leaky dams on downstream flood peak magnitude (Addy and Wilkinson, 2019; Burgess-Gamble et al., 2017; Dadson et al., 2017; Ellis et al., 2021; Lane, 2017), and no empirical quantification of their likelihood to fail during flood events at all (Hankin et al., 2020). Uncertainty in the effects of leaky dams and other NFM measures on downstream flood risk undermines confidence in their uptake and limits its adoption (Bark et al., 2021; Waylen et al., 2018; Wingfield et al., 2019). A robust evidence base is required to increase confidence in the implementation of NFM measures, such as leaky dams, for flood risk management (Cook et al., 2016; Dadson et al., 2017; Ellis et al., 2021; Iacob et al., 2017; Lane, 2017; McLean et al., 2013; Waylen et al., 2018).

This study has increased understanding of both the benefits for flood risk management in upland catchments and the resilience of engineered leaky dams. The aim of the study was addressed by the novel application of cross-disciplinary methods to observations from a UK upland, headwater catchment in which leaky dams were installed. The use of these methods not only allowed the research questions to be answered, but also provides a proof of concept for application of the approaches to data from other environments, NFM interventions, and

spatial scales, which is crucial if NFM is to be used as a mainstream flood risk management measure (Ellis et al., 2021).

5.1.1 Q1: What role could data-based time series modelling techniques play in quantifying NFM impacts from short and uncertain BACI data?

The lack of empirical quantification of the impacts of NFM on the flood hydrograph is attributed to difficulties in isolating the effects of NFM measures, short monitoring periods, the low frequency of extreme flood events and the influence of context and scale (Connelly et al., 2020; Wingfield et al., 2019). Furthermore, uncertainty in empirical hydrological data can mask the signal of NFM interventions (Lane, 2017), particularly for implementation at small spatial study extents (Ellis et al., 2021).

Data-based time series modelling techniques have long been used in hydrology (e.g. Dooge, 1959; Hipel and McLeod, 1994; Young, 1986). Transfer function noise (TFN) models, in particular, have been used to model the rainfall-runoff relationship (Young, 2003), detect the signal of land use change (e.g. Katimon et al., 2013; Watson et al., 2001) and real time modelling of river stage for flood warning (e.g. Leedal et al., 2010; Young, 2002).

In this study, the applicability of linear TFN models to modelling the baseline relationship between upstream and downstream stage of three steep, upland streams was assessed. Where the underlying processes were linear, the models were able to simulate the peaks of flood events to within ±2 cm accuracy for 95% of events.

Although the linear TFN family of models was chosen based on the statistical properties of the data, such a good fit to the data was not found for all three streams. Data-based time series modelling is an iterative process (Hipel and McLeod, 1994), given the necessary resources, and the insights gained from this study, it may be possible to find an appropriate family of time series models to fit the data from these streams with a similar level of accuracy. There are many other types of data-based time series models (Beven, 2001), including non-linear variations of the TFN approach (Young, 2003), which could be trialled, given

the necessary resources. Nevertheless, by achieving a high level of simulation accuracy on one of the streams, the study demonstrated that data-based time series modelling can be applied to extract valuable information for assessing leaky dam impacts on flood peak magnitude.

5.1.2 Q2: In upland streams, what is the impact of leaky dams on the flood peak magnitude of a range of flood events?

Fifty flood events with a return period between <1 year and 6 years were observed in the study site after the installation of seven leaky dams in a 280 m reach of a 0.13 m/m gradient stream. Both single and multi peaked events were recorded, with up to eight event peaks and durations ranging from hours to days.

Leaky dams reduced peak stage by more than the expected simulation error of a third of the flood peaks recorded, most of which were single peaked events, or the first peak of multi-peaked events. The biggest percentage reductions (16±8%) in peak discharge were observed for small events (peak discharge of 0.3 m^3/s, return period <1 year), but were negligible (3 ±7 %) for events >1.0 m^3/s, which had a 1 year return period. Absolute peak discharge reductions were, similarly, greatest for smaller events on average, but the largest individual reductions in absolute peak discharge were observed for events with a peak magnitude of 0.7–0.9 m^3/s. Peaks recorded in the stream during events which caused downstream flooding of infrastructure and properties were higher than 1 m^3/s and were not reduced by the leaky dams. On average for all events with a return period up to 1 year, peak discharge was reduced by 10%.

The results were in agreement with the impacts observed by Wenzel et al. (2014) & Keys et al. (2018) in smaller, lower gradient streams but did not necessarily support conclusions from modelling studies that leaky dams in high gradient streams are less effective than leaky dams in low gradient streams (Dixon et al., 2016; Thomas and Nisbet, 2012). The limit of effectiveness of leaky dams (1.0 m^3/s, 1 year return period event) in the study stream was similar to the limit of effectiveness of leaky dams at delaying flood peak magnitude in a low gradient, headwater stream (1.0 m^3/s, 2 year return period event) (Gregory et al., 1985;

Kitts, 2010). Comparison of findings between studies is confounded, however, by site, flood event and structure specific factors.

Leaky dam impacts were highly variable, which emphasises the importance of assessing NFM impacts for events with a range of characteristics which include not just different peak magnitudes, but also event durations, total volumes and events with more than one peak. Failure to do so could lead to overpromising and underdelivering on the efficacy of leaky dam NFM schemes, which could lead to over-reliance of communities on NFM measures (Wells et al., 2020) and ultimately lead to abandonment of the paradigm shift towards integrating NFM (Vira and Adams, 2009).

5.1.3 Q3: What potential does empirical fragility analysis have for quantifying the resilience of engineered leaky dams during extreme flood events?

There are concerns that failure of leaky dams installed for the purpose of NFM can lead to release of flood surges, scheme under-performance and supply mobile wood which could block downstream structures such as culverts and bridges (Hankin et al., 2020; Ruiz-Villanueva et al., 2014; Wohl et al., 2016). Network risk models are able to take into account leaky dam failure to inform strategic placement of leaky dams but are based on assumptions of leaky dam resilience during flood events, rather than observations (Hankin et al., 2020).

A survey of 15 UK leaky dam sites, covering almost 2000 in-stream leaky dams, showed that for dams installed between 2010 and 2020, 2% had failed. Based on observations of leaky dam resilience and failure from the five sites in which leaky dam failures had occurred, the probability of failure of leaky dams during flood events was estimated conditional on the flood return period. This fragility analysis approach is more commonly used to estimate the resilience of buildings during earthquakes, but is also used by the Environment Agency to make investment decisions about flood defence assets at a strategic level (Simm and Tarrant, 2018).

The fragility analysis showed that there was considerable variability in the probability of failure for a given return period event between leaky dam sites,

reflecting differences in site conditions, structure type and structure condition, as well as the loading experienced by the dams. Data from the five sites was pooled to estimate combined fragility functions for partial and complete leaky dam failures. Probability of complete failure was 3% for a 10 year return period event, increasing to 14% and 24% for a 50 and 100 year return period event respectively. By updating the fragility functions as data emerges from UK leaky dam sites, the fragility functions are likely to stabilise and present a national picture of leaky dam fragility.

5.1.4 Interrelationship of findings

The aim of the research was to quantify the risks and benefits for flood risk management of installing engineered leaky dams in steep, upland streams. The first two research questions addressed the quantification of the intended flood risk management benefits of leaky dams whilst the third addressed the quantification of the probability of leaky dam failure. The highly variable effects of the leaky dams on flood peak magnitude showed that, to have realistic expectations of the effectiveness of leaky dams, analysis needs to take into account a range of event types and magnitudes. Furthermore, the relatively high probability of failure of leaky dams, compared to traditional engineered structures, found in this study highlights the importance of taking into account resilience when assessing the effectiveness of a proposed scheme.

Because the failed leaky dams were rebuilt following failure in this study, it is not possible to say whether a reduction in the number of leaky dams due to failure would have reduced the overall impact on flood peak magnitude in the stream. The indication that the effectiveness of the leaky dams was linked to their in-channel storage volume suggests that it would, because the in-channel storage provided in the stream would be reduced. A reduced number of leaky dams due to failure could lead to greater forces being exerted on downstream leaky dams which, in turn, could increase their probability of failure. However, the findings suggest that failure probability was relatively low for the types of events during which the leaky dams were effective, and therefore this feedback

effect is unlikely to have a pronounced impact on leaky dam failures in the study site.

Leaky dams were likely to fail compared to traditional engineered defences and therefore it is recommended that redundancy is built into the system to maintain the intended level of protection. Although modelling suggests that adding a greater number of leaky dams to a steep stream reach does not necessarily increase its impact on flood peak magnitude due to underutilisation of storage (Hankin et al., 2020), the additional leaky dams could become valuable when leaky dam failures occur. Network risk models provide a method by which leaky dam failure probability can be accounted for when assessing leaky dam effectiveness (Hankin et al., 2020). Both the leaky dam impacts on flood peak magnitude, and failure probability found in this study could be used to inform and validate such a model to increase confidence in its results.

For high flow events with a return period up to one year the leaky dams in the study site had a low probability of complete failure (0.02) and reduced flood peak magnitude by 10% on average. However, for high flow events during which downstream flooding of infrastructure and properties was recorded the leaky dams did not reduce peak magnitude on the stream and were increasingly likely to fail (0.25 for a 5 year return period event). The benefits of leaky dams are therefore unlikely to outweigh the risks in the study site. However, the pooled fragility analysis has shown that this is unlikely to be the case for leaky dam sites around the UK as leaky dam failure probability was generally considerably lower than in the study site. Moreover, emerging empirical and modelling evidence indicates that leaky dams can be effective during more extreme events where extendable field storage is available (Black et al., 2021; Hankin et al., 2020). It is therefore recommended that to increase leaky dam effectiveness in the study site their design could be adapted to extend further onto the floodplain to encourage overland flow to offline storage areas. Increasing the length of the large wood pieces used to build leaky dams in such unconfined sections of the stream is also likely to decrease the risk of the wood being mobilised during flood events (Dixon and Sear, 2014). As only large wood pieces with a length greater than 2.5 times the channel width are functionally immobile (Dixon and Sear, 2014),

the recommendation is made that leaky dams should be built to be 2.5 times the channel width to increase both their resilience and floodplain connectivity.

As leaky dam design is adapted to become more effective at reducing flood peak stage, the signal of the interventions in monitored hydrological data will become stronger, and its detection will therefore be less sensitive to uncertainty in the data. As highlighted by Ellis et al. (2021) the impact on flood peak magnitude, and therefore problems of detecting change from uncertain data, is also related to the scale of implementation of NFM features. As uptake is limited by lack of evidence of leaky dam effectiveness (Bark et al., 2021; Waylen et al., 2018; Wingfield et al., 2019), and uncertainty about their risks (Waylen et al., 2018), results such as those obtained in this study are valuable to break this cycle.

5.2 Significance and implications of findings

This study empirically quantified, for the first time, the impact of leaky dams on flood peak magnitude for a range of events. It adds to the evidence base required to increase confidence in the implementation of leaky dams for flood risk management (Cook et al., 2016; Dadson et al., 2017; Ellis et al., 2021; Iacob et al., 2017; Lane, 2017; McLean et al., 2013; Waylen et al., 2018). It has provided completely new insights by considerably extending the empirical quantification of Wenzel et al. (2014) & Keys et al. (2018) of the impact of instream wood on two, artificial single peaked events in small streams, to the quantification of the impacts of leaky dams installed for the purpose of NFM in a larger, steeper stream on the peak magnitude of 50 high flow events.

The leaky dams did not reduce peak magnitude during flood events which caused flooding of downstream infrastructure and property. The results from the study site, therefore, support the conclusions of Dadson et al. (2017) that small floods may be significantly reduced by leaky dams, but will not have a major effect during more extreme events. These findings were in agreement with previous research (Keys et al., 2018; Wenzel et al., 2014) from the US and Germany that leaky dams could reduce peak magnitude of events with a return period less than one year, but not during more extreme events. However, recent research in the UK (Black et al., 2021) has empirically shown that the impact of leaky dams on

flood peak timing can increase with event peak magnitude, illustrating the site specific nature of these findings. Particularly, the availability of extendable field storage appears to be a key factor (Hankin et al., 2020).

By quantifying the impact of leaky dams on single and multi-peaked events with a range of durations, antecedent conditions and return periods ranging from <1 year to 6 years, the study showed that the effectiveness of leaky dams was highly variable, even for events with a similar peak magnitude. Whether the event was single or multi-peaked had a particularly pronounced impact; generally, leaky dams reduced the magnitude of single peaked events and the first peak of multi peaked events, but not subsequent event peaks. This supports the conclusions drawn from modelling of leaky dams at the catchment scale, that the time for the system to recover between flood peaks is an important design consideration (Metcalfe et al., 2017). The study has shown that simplifying the assessment of leaky dam impacts by considering single peaked events alone, as is often the case in modelling studies (Hankin et al., 2020), but also in empirical studies (e.g. Gregory et al., 1985), is likely to lead to misrepresentation of their effectiveness. Realistic quantification of leaky dam effectiveness is paramount to avoid over reliance of communities on leaky dams as flood risk management measures (Wells et al., 2020). Appreciation of the complexities of the effectiveness of leaky dams, and other NFM measures, is needed because unrealistic expectations of new paradigms, such as the shift towards greater working with natural processes, can lead to disappointment and premature abandonment (Vira and Adams, 2009).

Crucially, this study showed that the difficulties associated with the stochastic nature of flood events and high levels of uncertainty in hydrological data which have previously precluded empirical quantification of leaky dam impacts on flood peak magnitude (Black et al., 2021; Kitts, 2010; National Trust, 2015; Nisbet et al., 2015b), can be overcome by taking a data-based time series modelling approach. It therefore presents the opportunity to provide evidence of the impacts of leaky dams in a range of environments, in a similar way as there is for traditional flood defences (Ellis et al., 2021). Ellis et al. (2021) argued that the problems associated with NFM quantification can be overcome by integrated modelling and field observations at nested spatial scales. This study has shown

that this recommendation could be extended to include a data-based time series modelling step which allows the signal of NFM measures to be detected from field observations.

This study presents the first dataset of UK leaky dam failures and quantification of the vulnerability of leaky dams to complete or partial failure during flood events. By being based on observations of failure and resilience, the fragility functions developed improve on assumptions made about the failure probability of leaky dams (Hankin et al., 2020). Quantification of their failure probability allows the hazards of introducing large wood to streams in the form of engineered leaky dams to be weighed up against the benefits, which is crucial for their management (Wohl et al., 2016) and success (Dixon and Sear, 2014). The study provides a proof of concept for the application of fragility analysis to observations of leaky dam vulnerability and resilience from UK leaky dam sites. The proof of concept means that, as data emerges from Defra funded NFM projects across the UK, it can be incorporated in the fragility analysis to provide a national estimate of the probability of failure of leaky dams during flood events. Such national level fragility curves are used to make strategic investment decisions about flood risk management assets in the UK (Simm and Tarrant, 2018).

5.3 Limitations of the study

The impact of leaky dams on flood peak magnitude was quantified for one type of catchment and one type of instream wood. There are many site specific factors which are likely to affect leaky dam effectiveness, ranging from landscape scale differences, such as the underlying geology (Burgess-Gamble et al., 2017), catchment area and landcover (Black et al., 2021), stream specific differences such as slope and geomorphology (Dixon et al., 2016; Kitts, 2010), to structure specific differences such as the complexity of the leaky dam (Wilcox and Wohl, 2006), the availability of expandable field storage (potential for floodplain connectivity) (Black et al., 2021; Hankin et al., 2020; Thomas and Nisbet, 2012), and the gap beneath the dam (Gippel et al., 1996; Manners and Doyle, 2008; Shields and Alonso, 2012). The results from this study can, therefore, not be generalised to

other catchments and instream wood types. For leaky dams to become mainstream flood risk management measures, evidence from a range of environments is needed in the same way there is for traditional engineered defences (Ellis et al., 2021). Crucially, the study has shown that it is possible to quantify leaky dam impacts on flood peak magnitude by taking a data-based time series modelling approach, which can be applied to data from other catchments.

Considerable expertise is required to apply data-based time series modelling techniques because the type of model used depends on the statistical properties of the data. As found in this study, data from adjacent streams with similar hydrological properties in the same catchment may require different modelling approaches. In this case, a linear TFN model was found to be able to simulate event peak magnitude of the full range of events to within ±2 cm on one impact stream, but systematically over and under predicted peak stage on another impact stream, indicating that a non-linear model may have been more appropriate for this stream. Hence, the data-based modelling approach requires expertise in the application of different types of time series models. Whilst some hydrologists have specialised in applying this type of approach to hydrological data (e.g. Hipel and McLeod, 1994), expertise in data-based time series modelling is more commonplace in the fields of systems engineering and econometrics (Okiy et al., 2015) and could therefore be enhanced by cross-disciplinary collaborations.

The timeframe of this doctoral research was a limiting factor in the amount of baseline and post-intervention data that could be collected. By building leaky dams specifically for the purpose of this monitoring study it was possible to have an 18 month baseline monitoring period, but this limited the length of the post-intervention monitoring period and the time available for the analysis of the complete dataset. A longer post-intervention monitoring period, covering several flood seasons could provide the data required to conduct a formal statistical analysis of the variability of leaky dam efficacy during different types of events. Better understanding of the potential of seasonal effects, for example, could provide valuable insights about the potential of leaky dams to reduce flooding during the summer season, when agricultural land is most vulnerable (Morris and Brewin, 2014; Posthumus et al., 2009).

The fragility functions developed in this study were limited by the amount of data which was available. Whilst the empirical fragility functions improve on assumptions about leaky dam resilience (Hankin et al., 2020), there was not enough data to take into account variability in site conditions, leaky dam designs, exposure to loading conditions and condition of the leaky dams. These factors were, instead, represented in the width of the uncertainty bounds of the fragility functions. Although generic, national scale fragility functions are not intended to assess individual assets, but rather represent an average asset to inform national scale decisions (Simm and Tarrant, 2018), some level of differentiation would be informative. As data emerges, it will be possible to estimate fragility functions for different leaky dam designs, site and leaky dam conditions. As for traditional engineered flood defence assets (Simm and Tarrant, 2018), this will inform maintenance requirements and allow investment decisions to be made at local and national scales.

This study focused on the potential of leaky dams for flood risk management, but it is the potential to deliver multiple benefits and ecosystem services which presents opportunities for their use alongside traditional engineered defences (Connelly et al., 2020; Ellis et al., 2021; Klaar et al., 2020; Wilkinson et al., 2019). Whilst the focus on their impact on flood risk management is needed (Schanze, 2017), and their multiple benefits are being researched Deane et al. (e.g. 2021); Lo et al. (e.g. 2021), approaching NFM measures as if they were engineered flood defences is limiting. Rather than approaching NFM from the point of view of 'working with natural processes', it may be beneficial for the sustained success of NFM to approach it from the point of view of 'working for natural processes'. In such a framework the objective becomes restoring the natural hydrological and geomorphological functioning of wood in rivers, rather than a target reduction in flood peak magnitude. This is more likely to lead to reasonable expectations of leaky dams and successful monitoring objectives, and therefore sustained implementation of NFM measures (Nesshöver et al., 2016; Schanze, 2017; Wells et al., 2020).

5.4 Questions for further research

Further research is needed to assess the effectiveness and resilience of different types of engineered leaky dams in a range of environments and at larger spatial scales (Burgess-Gamble et al., 2017; Dadson et al., 2017; Ellis et al., 2021; Lane, 2017; Schanze, 2017; Wilkinson et al., 2019). Particularly, the ability of leaky dams, and other NFM measures, to deliver flood risk management benefits at the catchment scale, where NFM should preferably be implemented (Connelly et al., 2020), is questioned (Dadson et al., 2017; Lane, 2017). To address this, three avenues of further research are proposed:

1. Empirically quantify NFM impacts at the catchment scale using a data-based time series modelling approach

Monitoring of leaky dam impacts at the catchment scale are proceeding (e.g. Lancaster Environment Centre, 2021; University of Reading, 2021). Whilst such efforts have led to successful quantification of leaky dam impacts on flood peak timing, uncertainty in the hydrological data meant that quantification of leaky dam impacts on flood peak magnitude was avoided (Black et al., 2021). As demonstrated in this study, at the stream scale certain data uncertainty issues can be overcome by taking a data-based time series modelling approach. This approach may also present a solution to quantifying catchment scale impacts of NFM measures. The underlying data generating processes are more complex at the catchment scale than the stream scale because the impacts of heterogeneity on hydrological processes cause uncertainty (Black et al., 2021), interactions between different parts of the catchment play a role (Dixon et al., 2016; Pattison et al., 2014), and dominant hydrological processes change with scale (Ellis et al., 2021). However, the ability of data-based time series modelling, such as neural networks (Piotrowski and Napiorkowski, 2013), and the DBM approach (Young, 2003) to represent the complex processes of the rainfall-runoff relationship at the catchment scale is promising. It is therefore recommended that the data-based time series modelling approach is applied to detecting the impacts of leaky dams, and other NFM measures, on flood peak magnitude at the catchment scale.

2. Validate the representation of leaky dams in hydraulic and hydrological models using the results of data-based time series modelling studies

Questions of upscaling NFM interventions may also be addressed by using hydraulic and hydrological models (Dadson et al., 2017). Addy and Wilkinson (2019) highlight the need for guidance about the representation of leaky dams in hydraulic and hydrological models to increase confidence in their outputs. This research presents a dataset of leaky dam impacts, and an approach by which to quantify leaky dam impacts in other environments, which provides the data needed to validate the representation of leaky dams in such models. For example, leaky dam impacts on the flood peak magnitude of the 50 events observed on the impact stream in the Coverdale study site could be used to validate the representation of the leaky dams in a hydraulic model of the stream. Given appropriate consideration of the model's ability to extrapolate beyond the validated scale (Dadson et al., 2017), this representation of leaky dams could be used to assess the impact of installing leaky dams in all of the hydraulically similar headwater streams in the Coverdale catchment. This approach, particularly if applied using data from a range of leaky dam types and environments, could considerably increase confidence in the results of catchment scale modelling studies.

3. Develop national scale fragility functions for leaky dams

As a result of the monitoring requirements of Defra funded NFM projects (Arnott et al., 2018), observations of leaky dam failures and resilience during extreme flood events are becoming available (The Rivers Trust, 2021). Such observations can be used to calculate the failure probability of leaky dams during a flood event, as demonstrated in this study, conditional on the flood return period. It is recommended that, as data becomes available, national scale estimates of the fragility functions of different leaky dam designs, environments, and condition are made. National scale fragility functions inform programming of maintenance and investment decisions at regional and national scales (Simm and Tarrant, 2018), and could be used for leaky dams in the same way they are for traditional flood defences. Furthermore, decisions about the design of leaky dam schemes

could be made with greater confidence by incorporating empirically based fragility functions in network models (Hankin et al., 2020).

5.5 Concluding remarks

Leaky dams reduced flood peak magnitude of events with a return period up to one year by 10% on average, but their efficacy was highly variable and only a third of the event peaks observed in the study site were significantly reduced. Observations from UK leaky dam sites indicated that the probability of complete failure of leaky dams during frequent flood events was low (0.01 for a 1-in-5 year event) but increased for more extreme events (0.24 for a 1-in-100 year event).

Unrealistic expectations are unlikely to lead to sustained adoption of new paradigms (Adams and Hulme, 2001). By empirically demonstrating the variability of leaky dam impacts on downstream flood peak magnitude, and quantifying the vulnerability of leaky dams to partial and complete failure during flood events, this research presents a critical evaluation of the benefits and hazards of leaky dams for flood risk management. By doing so, this study helps to manage expectations which ultimately supports the sustained adoption of working with natural processes in flood risk management.

References

Abbe, T. (2000). Patterns, mechanics and geomorphic effects of wood debris accumulations in a forest river system. Doctoral thesis, University of Washington, Seattle.

Abbe, T., Brooks, A. (2011). Geomorphic, engineering, and ecological considerations when using wood in river restoration. *Geophysical Monograph Series*, **194**, pp. 419–451.

Abbe, T.B., Brooks, A.P., Montgomery, D.R. (2003). Wood in River Rehabilitation and Management. In S.V. Gregory, K.L. Boyer, A.M. Gurnell (Eds.), *The ecology and management of wood in world rivers*. American Fisheries Society, pp. 1–26.

Adams, W.M., Hulme, D. (2001). Conservation and communities: changing narratives, policies and practices in African conservation. In D. Hulme, M. Murphree (Eds.), *African wildlife and livelihoods: the promise and performance of community conservation*. London: James Currey, pp. 9–23.

Addy, S., Wilkinson, M. (2016). An assessment of engineered log jam structures in response to a flood event in an upland gravel-bed river. *Earth Surface Processes and Landforms*, **1670**(April), pp. 1658–1670.

Addy, S., Wilkinson, M.E. (2019). Representing natural and artificial in-channel large wood in numerical hydraulic and hydrological models. *WIREs Water*, **6**(6), pp. 1–20.

Agriculture Act (2020). (c.21) London: The Stationary Office.
[Online] Available from: *https://www.legislation.gov.uk/*

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. Petrov, F. Csaki (Eds.), *International Symposium on Information Theory*. Budapest, pp. 267–281.

Alderson, D.M., Evans, M.G., Shuttleworth, E.L., Pilkington, M., Spencer, T., Walker, J., Allott, T.E. (2019). Trajectories of ecosystem change in restored blanket peatlands. *Science of the Total Environment*, **665**, pp. 785–796.

Alila, Y., Hudson, R., Kuraś, P.K., Schnorbus, M., Rasouli, K. (2010). Reply to comment by Jack Lewis et al. on "Forests and floods: A new paradigm sheds light on age-old controversies". *Water Resources Research*, **46**(5).

Alila, Y., Kuras, P.K., Schnorbus, M., Hudson, R. (2009). Forests and floods: A new paradigm sheds light on age-old controversies. *Water Resources Research*, **45**(8), pp. 1–24.

Andrews, F., Guillaume, J. (2018). hydromad: Hydrological Model Assessment and Development.
[Online] Available from: *http://hydromad.catchment.org/*

Archer, D.R. (2007). The use of flow variability analysis to assess the impact of land use change on the paired Plynlimon catchments, mid-Wales. *Journal of Hydrology*, **347**(3-4), pp. 487–496.

Armstrong, J., Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, **8**(1), pp. 69–80.

Arnott, S., Burgess-Gamble, L., Dunsford, D., Webb, L., Johnson, D., Andison, E., Slaney, A., Vaughan, M., Ngai, R., Rose, S., Maslen, S. (2018). Monitoring and evaluating the DEFRA funded Natural Flood Management projects. Technical Report July, Environment Agency.
[Online] Available from: *http://www.gov.uk/government/publications*

Astrup, R., Coates, K.D., Hall, E. (2008). Finding the appropriate level of complexity for a simulation model: An example with a forest growth model. *Forest Ecology and Management*, **256**(10), pp. 1659–1665.

Bailey, D.H., Ger, S., Lopez de Prado, M., Sim, A. (2015). *Statistical Overfitting and Backtest Performance*. Elsevier.

Bark, R.H., Martin-Ortega, J., Waylen, K.A. (2021). Stakeholders' views on natural flood management: Implications for the nature-based solutions paradigm shift? *Environmental Science & Policy*, **115**, pp. 91–98.

Bartolomei, S.M., Sweet, A.L. (1989). A note on a comparison of exponential smoothing methods for forecasting seasonal series. *International Journal of Forecasting*, **5**(1), pp. 111–116.

Bell, V.A., Carrington, D.S., Moore, R.J. (2001). Comparison of Rainfall-Runoff Models for Flood Forecasting Part 2: Calibration and evaluation of models. Technical report, Environment Agency, Bristol.

Bergmeir, C., Benítez, J.M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, **191**, pp. 192–213.

Bernhardt, E.S., Palmer, M.A., Allan, J.D., Alexander, G., Barnas, K., Brooks, S., Carr, J., Clayton, S., Dahm, C., Galat, D., Gloss, S., Goodwin, P., Hart, D., Hassett, B., Jenkinson, R., Katz, S., Kondolf, G.M., Lake, P.S., Lave, R., Meyer, J.L., Donnell, T.K.O., Pagano, L., Powell, B., Sudduth, E. (2005). Synthesizing U.S. River Restoration Efforts. *Science*, **308**, pp. 636–637.

Beschta, R.L. (1983). The effects of large organic debris upon channel morphology: A flume study. In *D.B. Simons Symposium on Erosion and Sedimentation.* Ft. Collins, CO: Simon, Li & Assoc., pp. 8.63–8.78.
[Online] Available from: *http://www.fsl.orst.edu/*

Beven, K. (1989). Changing ideas in hydrology- the case of physically-based models. *Journal of Hydrology*, **105**, pp. 157–172.

Beven, K. (2001). *Rainfall-runoff modelling: the primer*. 2 edition. Chichester: Wiley-Blackwell.

Beven, K. (2016). Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, **61**(9), pp. 1652–1665.

Beven, K., Westerberg, I. (2011). On red herrings and real herrings: Disinformation and information in hydrological inference. *Hydrological Processes*, **25**(10), pp. 1676–1680.

Beven, K., Young, P., Leedal, D., Romanowicz, R. (2008). Computationally efficient flood water level prediction (with uncertainty). *Flood Risk Management: Research and Practice*, pp. 281–289.

Bezak, N., Horvat, A., Šraj, M. (2015). Analysis of flood events in Slovenian streams. *Journal of Hydrology and Hydromechanics*, **63**(2), pp. 134–144.

Bilby, R.E., Ward, J.W. (1989). Changes in Characteristics and Function of Woody Debris with Increasing Size of Streams in Western Washington. *Transactions of the American Fisheries Society*, **118**(4), pp. 368–378.

Black, A., Peskett, L., MacDonald, A., Young, A., Spray, C., Ball, T., Thomas, H., Werritty, A. (2021). Natural flood management, lag time and catchment scale: Results from an empirical nested catchment study. *Journal of Flood Risk Management*, **14**(3), pp. 1–16.

Blanckaert, K., Duarte, A., Chen, Q., Schleiss, A.J. (2012). Flow processes near smooth and rough (concave) outer banks in curved open channels. *Journal of Geophysical Research*, **117**, pp. 1–17.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R.A., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G.T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G.B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T.R., Kohnová, S., Koskela, J.J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J.L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, **573**, pp. 108–111.

Boon, P.J. (1998). River restoration in five dimensions. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **8**(1), pp. 257–264.

Bowerman, B.L., Koehler, A.B. (1989). The appropriateness of Gardner's simple approach and Chebychev prediction intervals. In *The 9th International Symposium on Forecasting*. Vancouver, BC.

Box, G.E.P., Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. revised edition. Oakland, California: Holden-Day.

Braccia, A., Batzer, D.P. (2008). Breakdown and invertebrate colonization of dead wood in wetland, upland, and river habitats. *Canadian Journal of Forest Research*, **38**(10), pp. 2697–2704.

Bradshaw, C.J., Sodhi, N.S., Peh, K.S., Brook, B.W. (2007). Global evidence that deforestation amplifies flood risk and severity in the developing world. *Global Change Biology*, **13**(11), pp. 2379–2395.

Broadmeadow, S., Nisbet, T. (2013). Yorkshire and North East England Woodland for Water Project - Phase 1 Opportunity mapping. Technical report, Forest Research, Cheshire.

Brown, J.D., Damery, S.L. (2002). Managing flood risk in the UK: Towards an integration of social and technical perspectives. *Transactions of the Institute of British Geographers*, **27**(4), pp. 412–426.

Brummer, C.J., Abbe, T.B., Sampson, J.R., Montgomery, D.R. (2006). Influence of vertical channel change associated with wood accumulations on delineating channel migration zones, Washington, USA. *Geomorphology*, **80**(3-4), pp. 295–309.

Brunner, G.W. (2000). Common Model Stability Problems When Performing an Unsteady Flow Analysis.
[Online] Available from: *https://www.nws.noaa.gov/* [Accessed: 18 May 2020]

Buffington, J.M., Montgomery, D.R. (1999). Effects of hydraulic roughness on surface textures of gravel-bed rivers. *Water Resources Research*, **35**(11), pp. 3507–3521.

Buffington, J.M., Montgomery, D.R., Greenberg, H.M. (2004). Basin-scale availability of salmonid spawning gravel as influenced by channel type and hydraulic roughness in mountain catchments. *Canadian Journal of Fisheries and Aquatic Sciences*, **61**(11), pp. 2085–2096.

Burgess-Gamble, L., Ngai, R., Wilkinson, M., Nisbet, T., Pontee, N., Harvey, R., Kipling, K., Addy, S., Rose, S., Maslen, S., Jay, H., Nicholson, A., Page, T., Jonczyk, J., Quinn, P. (2017). Working with Natural Processes - Evidence Directory. Technical report, Environment Agency, Bristol.
[Online] Available from: *https://assets.publishing.service.gov.uk/*

Butler, D.R., Malanson, G.P. (2005). The geomorphic influences of beaver dams and failures of beaver dams. *Geomorphology*, **71**(1-2), pp. 48–60.

Calderdale Council (2016). Significant progress made since 2015 Boxing Day floods but 90,000 homes still at risk without further investment.
[Online] Available from: *http://news.calderdale.gov.uk/* [Accessed: 6 March 2017]

Caner, M., Kilian, L. (2001). Size distortions of tests of the null hypothesis of stationarity: Evidence and implications for the PPP debate. *Journal of International Money and Finance*, **20**(5), pp. 639–657.

Carrick, J., Abdul Rahim, M.S.A.B., Adjei, C., Ashraa Kalee, H.H.H., Banks, S.J., Bolam, F.C., Campos Luna, I.M., Clark, B., Cowton, J., Domingos, I.F.N., Golicha, D.D., Gupta, G., Grainger, M., Hasanaliyeva, G., Hodgson, D.J., Lopez-Capel, E., Magistrali, A.J., Merrell, I.G., Oikeh, I., Othman, M.S., Ranathunga Mudiyanselage, T.K.R., Samuel, C.W.C., Sufar, E.K., Watson, P.A., Zakaria, N.N.A.B., Stewart, G. (2019). Is planting trees the solution to reducing flood risks? *Journal of Flood Risk Management*, **12**(S2), pp. 1–10.

Castellarin, A., Di Baldassarre, G., Bates, P.D., Brath, A. (2009). Optimal Cross-Sectional Spacing in Preissmann Scheme 1D Hydrodynamic Models. *Journal of Hydraulic Engineering*, **135**(2), pp. 96–105.

CBEC Eco Engineering (2017). Natural Flood Management Toolbox Guidance for working with natural processes in flood management schemes. Technical Report June, Environment Agency, Bristol, UK.

Chatfield, C. (2001). Prediction Intervals for Time-Series Forecasting. In J. Armstrong (Ed.), *Principles of Forecasting. International Series in Operations Research & Management Science, vol 30*. Boston, MA: Springer, pp. 475–494.

Checkland, P.B. (1981). Systems Thinking. In *Systems Thinking, Systems Practice*. Chichester: John Wiley, p. 330.

Chow, V.T. (1959). *Open-channel hydraulics*. New York: McGraw-Hill Book Co.

CIRIA (2021). Guidance on natural flood management RP1094.
[Online] Available from: *https://www.ciria.org/* [Accessed: 1 October 2021]

Clausen, J.C., Spooner, J. (1993). Paired Watershed Study Design. Technical report, Office of Wetlands, Oceans and Watersheds, Environmental Protection Agency, Washington DC,.
[Online] Available from: *https://nepis.epa.gov/*

Coe, H.J., Kiffney, P.M., Pess, G.R. (2006). A comparison of methods to evaluate the response of periphyton and invertebrates to wood placement in large Pacific coastal rivers. *Northwest Science*, **80**(4), pp. 298–307.

Collentine, D., Futter, M.N. (2018). Realising the potential of natural water retention measures in catchment flood management: trade-offs and matching interests. *Journal of Flood Risk Management*, **11**(1), pp. 76–84.

Collins, B.D., Montgomery, D.R. (2002). Forest Development, Wood Jams, and Restoration of Floodplain Rivers in the Puget Lowland, Washington. *Restoration Ecology*, **10**(2), pp. 237–247.

Collins, B.D., Montgomery, D.R., Fetherston, K.L., Abbe, T.B. (2012). The floodplain large-wood cycle hypothesis: A mechanism for the physical and biotic structuring of temperate forested alluvial valleys in the North Pacific coastal ecoregion. *Geomorphology*, **139-140**, pp. 460–470.

Comiti, F., Andreoli, A., Lenzi, M., Mao, L. (2006). Spatial density and characteristics of woody debris in five mountain rivers of the Dolomites (Italian Alps). *Geomorphology*, **78**(1-2), pp. 44–63.

Comiti, F., Andreoli, A., Mao, L., Aristide Lenzi, M. (2008). Wood storage in three mountain streams of the Southern Andes and its hydromorphological effects. *Earth Surface Processes and Landforms*, **33**, pp. 244–262.

Commission of the European Communities (2009). White Paper - Adapting to climate change - towards a European framework for action. Technical report, European Commission, Brussels.
[Online] Available from: *https://eur-lex.europa.eu/legal-content/*

Committee on Climate Change (2017). UK Climate Change Risk Assessment 2017 Synthesis report: priorities for the next five years. Technical report, Committee on Climate Change, London.
[Online] Available from: *https://www.theccc.org.uk/wp-content/*

Connelly, A., Snow, A., Carter, J., Lauwerijssen, R. (2020). What approaches exist to evaluate the effectiveness of UK-relevant natural flood management measures? A systematic map protocol. *Environmental Evidence*, **9**(1), pp. 1–13.

Cook, B., Forrester, J., Bracken, L., Spray, C., Oughton, E. (2016). Competing paradigms of flood management in the Scottish/English borderlands. *Disaster Prevention and Management*, **25**(3), pp. 314–328.

Correa-Metrio, A., Urrego, D.H., Cabrera, K.R., Bush, M.B. (2012). paleoMAS: Paleoecological Analysis.
[Online] Available from: *https://cran.r-project.org/package=paleoMAS*

Costanza, R., D'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, **387**(6630), pp. 253–260.

Coz, J.L. (2012). A literature review of methods for estimating the uncertainty associated with stage-discharge relations. Technical report, WMO, Lyon, France. [Online] Available from: *http://www.wmo.int/*

Cramer, M.L. (2012). Stream Habitat Restoration Guidelines. Technical report, Washington State Aquatic Habitat Guidelines Program, Washington DC.

Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., Arheimer, B. (2020). Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, **65**(5), pp. 699–711.

Curran, J.H., Wohl, E.E. (2003). Large woody debris and flow resistance in step-pool channels, Cascade Range, Washington. *Geomorphology*, **51**(1-3), pp. 141–157.

Dadson, S.J., Hall, J.W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., Heathwaite, L., Holden, J., Holman, I.P., Lane, S.N., O'Connell, E., Penning-Rowsell, E., Reynard, N., Sear, D., Thorne, C., Wilby, R. (2017). A restatement of the natural science evidence concerning catchment-based 'natural' flood management in the UK. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **473**(2199), p. 20160706.

David, G.C., Wohl, E., Yochum, S.E., Bledsoe, B.P. (2011). Comparative analysis of bed resistance partitioning in high-gradient streams. *Water Resources Research*, **47**(7), pp. 1–22.

Davis, S.R., Brown, A.G., Dinnin, M.H. (2007). Floodplain connectivity, disturbance and change: A palaeoentomological investigation of floodplain ecology from south-west England. *Journal of Animal Ecology*, **76**(2), pp. 276–288.

Deane, A., Norrey, J., Coulthard, E., McKendry, D.C., Dean, A.P. (2021). Riverine large woody debris introduced for natural flood management leads to rapid improvement in aquatic macroinvertebrate diversity. *Ecological Engineering*, **163**(106197), pp. 1–10.

Deasy, C., Brazier, R.E., Heathwaite, A.L., Hodgkinson, R. (2009). Pathways of runoff and sediment transfer in small agricultural catchments. *Hydrological Processes*, **23**(1), pp. 1349–1358.

Defra (2005). Making Space for Water:Taking Forward a New Government Strategy for Flood and Coastal Management in England. Technical report, DEFRA, London.

Defra (2009). Appraisal of flood and coastal erosion risk management. A Defra policy statement. Technical Report June, Department for Environment, Food and Rural Affairs, London.
[Online] Available from: *https://www.gov.uk/government/publications/*

Defra (2011a). The Natural Choice: securing the value of nature. Technical Report June, HM Government, London.
[Online] Available from: *http://www.official-documents.gov.uk/document/*

Defra (2011b). Uplands Policy Review. Technical report, Department for Environment, Food and Rural Affairs, London.
[Online] Available from: *www.defra.gov.uk*

Defra (2013). Catchment Based Approach: Improving the quality of our water environment. Technical report, London.
[Online] Available from: *https://www.gov.uk/government/publications/catchment-based-approach-improving-the-quality-of-our-water-environment*

Defra (2015). UK Terrestrial and Freshwater Habitat types: Uplands Habitat descriptions.
[Online] Available from: *https://webarchive.nationalarchives.gov.uk/* [Accessed: 6 October 2021]

Defra (2018). A Green Future: Our 25 Year Plan to Improve the Environment. Technical report, HM Government, London.
[Online] Available from: *https://assets.publishing.service.gov.uk/government/*

Defra (2020). The Environmental Land Management scheme: public money for public goods. Technical report, Department for Environment, Food and Rural Affairs, London.
[Online] Available from: *https://www.gov.uk/government/publications/*

Defra, Dbeis, Met Office Hadley Centre, Environment Agency (2018). UKCP18 Factsheet : Precipitation. Technical report, Met Office, London.
[Online] Available from: *https://www.metoffice.gov.uk/binaries/*

Defra, EAFRD, Natural England (2016). Countryside Stewardship - How can it help reduce flooding?
[Online] Available from: *https://www.gov.uk/government/uploads/* [Accessed: 10 March 2017]

Di Baldassarre, G., Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, **13**(6), pp. 913–921.

Dickson, N.E., Carrivick, J.L., Brown, L.E. (2012). Flow regulation alters alpine river thermal regimes. *Journal of Hydrology*, **464-465**, pp. 505–516.

Diehl, T. (1997). Potential Drift Accumulation at Bridges FHWA-RD-97-028. Technical Report April, FHWA, Virginia.

Dixon, S.J., Sear, D.A. (2014). The influence of geomorphology on large wood dynamics in a low gradient headwater stream. *Water Resources Research*, **50**(12), pp. 9194–9210.

Dixon, S.J., Sear, D.A., Odoni, N.A., Sykes, T., Lane, S.N. (2016). The effects of river restoration on catchment scale flood risk and flood hydrology. *Earth Surface Processes and Landforms*, **41**(7), pp. 997–1008.

Dodd, J.A., Newton, M., Adams, C.E. (2016). The effect of natural flood management in-stream wood placements on fish movement in Scotland, CD2015_02. Technical report, CREW.
[Online] Available from: *crew.ac.uk/publications*

Dooge, J. (1959). A general theory of the unit hydrograph. *Journal of Geophysical Research*, **64**, pp. 241–256.

Dorofki, M., Elshafie, A.H., Jaafar, O., Karim, O.A., Mastura, S. (2012). Comparison of Artificial Neural Network Transfer Functions Abilities to Simulate Extreme Runoff Data. In *International Conference on Environment, Energy and Biotechnology*, volume 33. Singapore: IACSIT Press, pp. 39–44.
[Online] Available from: *www.ipcbee.com/vol33/008-ICEEB2012-B021.pdf*

Dunn, S., Wilkinson, S., Alderson, D., Fowler, H., Galasso, C. (2018). Fragility Curves for Assessing the Resilience of Electricity Networks Constructed from an Extensive Fault Database. *Natural Hazards Review*, **19**(1), pp. 04017019.1–10.

Dust, D., Wohl, E. (2012). Characterization of the hydraulics at natural step crests in step-pool streams via weir flow concepts. *Water Resources Research*, **48**(9), pp. 1–14.

Ellis, N., Anderson, K., Brazier, R. (2021). Mainstreaming natural flood management: A proposed research framework derived from a critical evaluation of current knowledge. *Progress in Physical Geography: Earth and Environment*, pp. 1–23.

England, J., Skinner, K.S., Carter, M.G. (2008). Monitoring, river restoration and the Water Framework Directive. *Water and Environment Journal*, **22**(4), pp. 227–234.

Environment Agency (2007a). Geomorphological Monitoring Guidelines for River Restoration Schemes. Technical Report Revision R01 February 07, Environment Agency, Bristol.

Environment Agency (2007b). Tree management work in Valency Valley, Boscastle - Factsheet: Work aims to reduce flood risk. Technical report, Environment Agency, Bristol.
[Online] Available from: *http://www.boscastlecornwall.org.uk/national_trust/*

Environment Agency (2010a). Flood and Coastal Erosion Risk Management appraisal guidance. Technical report, Environment Agency, Bristol. [Online] Available from: *https://www.gov.uk/government/uploads/*

Environment Agency (2010b). Water Framework Directive (WFD) Mitigation Measures Online Manual - Management and use of large wood. [Online] Available from: *http://evidence.environment-agency.gov.uk/FCERM/* [Accessed: 26 January 2017]

Environment Agency (2018). Natural Flood Management Minimising the Risks - Quick guide. Technical Report 337_18, Environment Agency, Bristol, UK. [Online] Available from: *https://catchmentbasedapproach.org/wp-content/*

Environment Agency (2019). Natural Flood Management Programme : Interim Lessons Learnt. Technical Report January, Environment Agency, Bristol.

Environment Bill (2021). (HL Bill 53). London: The Stationary Office. [Online] Available from: *https://bills.parliament.uk/publications/*

Evrard, O., Persoons, E., Vandaele, K., van Wesemael, B. (2007). Effectiveness of erosion mitigation measures to prevent muddy floods: A case study in the Belgian loam belt. *Agriculture, Ecosystems and Environment*, **118**(1-4), pp. 149–158.

Fenn, C., Bettess, R., Golding, B., Farquharson, F., Wood, T. (2005). The Boscastle flood of 16 August 2004: Characteristics, causes and consequences. In *Proceedings of the 40th Defra Flood and Coastal Management Conference*. HR Wallingford, p. 19.

Flood List (2020). Flood Data and API. [Online] Available from: *http://floodlist.com/* [Accessed: 6 November 2020]

Forbes, H., Ball, K., McLay, F. (2015). Natural Flood Management Handbook. Technical report, SEPA, Stirling.

Freedman, D., Pisani, R., Purves, R. (2007). *Statistics (international student edition)*. 4th edition. New York: WW Norton and Company.

Fuller, W.A. (1996). *Introduction to Statistical Time Series*. 2 edition. New York: John Wiley and Sons.

Gapinski, C.M., Hermes, J., von Haaren, C. (2021). Why people like or dislike large wood in rivers- a representative survey of the general public in Germany. *River Research and Applications*, **37**(2), pp. 187–197.

Gebrehiwot, S.G., Di Baldassarre, G., Bishop, K., Halldin, S., Breuer, L. (2019). Is observation uncertainty masking the signal of land use change impacts on hydrology? *Journal of Hydrology*, **570**, pp. 393–400.

Gilleland, E., Katz, R.W. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, **72**(8), pp. 1–39.

Gippel, C.J., O'Neill, I.C., Finlayson, B.L., Schnatz, I. (1996). Hydraulic Guidelines for the Re-Introduction and Management of Large Woody Debris in Lowland Rivers. *Regulated Rivers: Research & Management*, **12**(2-3), pp. 223–236.

Glendell, M., Extence, C., Chadd, R., Brazier, R.E. (2014). Testing the pressure-specific invertebrate index (PSI) as a tool for determining ecologically relevant targets for reducing sedimentation in streams. *Freshwater Biology*, **59**(2), pp. 353–367.

Godfrey, L.G. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, **46**(6), p. 1293.

Gomi, T., Moore, R.D., Dhakal, A.S. (2006). Headwater stream temperature response to clear-cut harvesting with different riparian treatments, coastal British Columbia, Canada. *Water Resources Research*, **42**(8), pp. 1–11.

Gordon, N.D., McMahon, T.A., Finlayson, B.L., Gippel, C.J., Nathan, R.J. (2004). *Stream Hydrology An Introduction for Ecologists*. 2 edition. Chichester: John Wiley and Sons.

Government, H. (2016). National Flood Resilience Review. Technical Report September, Cabinet Office, London.
[Online] Available from: *https://assets.publishing.service.gov.uk/government/*

Grabowski, R.C., Gurnell, A.M., Burgess-Gamble, L., England, J., Holland, D., Klaar, M.J., Morrissey, I., Uttley, C., Wharton, G. (2019). The current state of the use of large wood in river restoration and management. *Water and Environment Journal*, **33**, pp. 366–377.

Grand-Clement, E., Anderson, K., Smith, D., Luscombe, D., Gatis, N., Ross, M., Brazier, R.E. (2013). Evaluating ecosystem goods and services after restoration of marginal upland peatlands in South-West England. *Journal of Applied Ecology*, **50**(2), pp. 324–334.

Gregory, K.J., Gurnell, a.M., Hill, C.T. (1985). The permanence of debris dams related to river channel processes. *Hydrological Sciences Journal*, **30**(3), pp. 371–381.

Guerrero, J.L., Westerberg, I.K., Halldin, S., Xu, C.Y., Lundin, L.C. (2012). Temporal variability in stage–discharge relationships. *Journal of Hydrology*, **446-447**, pp. 90–102.

Gurnell, A.M., Piegay, H., Swanson, F.J., Gregorys, S.V. (2002). Large wood and fluvial processes. *Freshwater Biology*, **47**, pp. 601–619.

Hahn, G.J., Meeker, W.Q. (1991). *Statistical Intervals: A Guide for Practitioners*. Wiley Series in Probability and Statistics. New York: Wiley.

Hammond, D., Mant, J., Holloway, J., Elbourne, N., Janer, M. (2011). Practical River Restoration Appraisal Guidance for Monitoring Options (PRAGMO). *The River Restoration Centre (RRC)*, p. 330.

Han, D., Chan, L., Zhu, N. (2007). Flood forecasting using support vector machines. *Journal of Hydroinformatics*, **9**(4), pp. 267–276.

Hankin, B., Hewitt, I., Sander, G., Danieli, F., Formetta, G., Kamilova, A., Kretzschmar, A., Kiradjiev, K., Wong, C., Pegler, S., Lamb, R. (2020). A risk-based network analysis of distributed in-stream leaky barriers for flood risk management. *Natural Hazards and Earth System Sciences*, **20**(10), pp. 2567–2584.

Hannaford, J. (2015). Climate-driven changes in UK river flows: A review of the evidence. *Progress in Physical Geography*, **39**(1), pp. 29–48.

Hastie, T., Tibshirani, R., Friedman, J. (2017). *The Elements of Statistical Learning*. 2 edition. Springer-Verlag.

Herrera Environmental Consultants (2006). Conceptual Design Guidelines - Application of Engineered Logjams. Technical report, SEPA, Stirling.

Hess, T.M., Holman, I.P., Rose, S.C., Rosolova, Z., Parrott, A. (2010). Estimating the impact of rural land management changes on catchment runoff generation in England and Wales. *Hydrological Processes*, **24**(10), pp. 1357–1368.

Hester, N., Rose, S., Worrall, P. (2016). Case study 3 - Holnicote demonstration project - Somerset. Technical report, Environment Agency, Bristol. [Online] Available from: *https://www.gov.uk/government/publications/*

Hipel, K.W., McLeod, A.I. (1994). *Time Series Modelling of water resouces and environmental systems*. Amsterdam: Elsevier.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, **3**(9), pp. 816–821.

HM Treasury (2016). Autumn Statement 2016: Philip Hammond's speech. [Online] Available from: *https://www.gov.uk/government/speeches/*

Hogan, D. (1986). Channel morphology of unlogged, logged and debris torrented streams in the Queen Charlotte Islands. British Columbia Ministry of Forests and Lands, Land Management Report 49. Technical report, Victoria BC: Research Branch, Ministry of Forests and Lands.

Holden, J., Evans, M.G., Burt, T.P., Horton, M. (2006). Impact of Land Drainage on Peatland Hydrology. *Journal of Environmental Quality*, **35**(5), pp. 1764–1778.

Horner, I., Renard, B., Le Coz, J., Branger, F., McMillan, H.K., Pierrefeu, G. (2018). Impact of Stage Measurement Errors on Streamflow Uncertainty. *Water Resources Research*, **54**(3), pp. 1952–1976.

Horritt, M.S., Bates, P.D. (2002). Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology*, **268**(1-4), pp. 87–99.

House of Commons Environment Food and Rural Affairs Committee (2016). Future flood prevention - Second Report of Session 2016-17. Technical Report October, House of Commons, London.

Hughes, A.O., Davies-Colley, R., Bellingham, M., van Assema, G. (2020). The stream hydrology response of converting a headwater pasture catchment to Pinus radiata plantation. *New Zealand Journal of Marine and Freshwater Research*, **54**(3), pp. 308–328.

Hurvich, C.M., Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**(2), pp. 297–307.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F. (2020). forecast: Forecasting functions for time series and linear models.
[Online] Available from: *http://pkg.robjhyndman.com/forecast*

Hyndman, R.J., Athanasopoulos, G. (2018). *Forecasting Principles and Practice*. 2 edition. OTexts.
[Online] Available from: *https://otexts.com/fpp2/*

Hyndman, R.J., Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, **27**(3), p. 22.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, **18**(3), pp. 439–454.

Iacob, O., Brown, I., Rowan, J. (2017). Natural flood management, land use and climate change trade-offs: the case of Tarland catchment, Scotland. *Hydrological Sciences Journal*, **62**(12), pp. 1931–1948.

Interreg North Sea Region Programme (2017). Study finds nature-based solutions add flood resilience across large areas.
[Online] Available from: *https://northseablog.eu/study-finds-nature-based-solutions-add-resilience-against-floods/* [Accessed: 4 October 2021]

Isengildina-Massa, O., Irwin, S., Good, D.L., Massa, L. (2011). Empirical confidence intervals for USDA commodity price forecasts. *Applied Economics*, **43**(26), pp. 3789–3803.

Jackson, B., Wheater, H., Mcintyre, N., Chell, J., Francis, O., Frogbrook, Z., Marshall, M., Reynolds, B., Solloway, I. (2008). The impact of upland land management on flooding: insights from a multiscale experimental and modelling programme. *Journal of Flood Risk Management*, **1**(2), pp. 71–80.

Jakeman, A.J., Hornberger, G.M. (1993). How much complexity is warranted in a rainfall- runoff model? *Water Resources Research*, **29**(8), pp. 2637–2649.

Jakeman, A.J., Littlewood, I.G., Whitehead, P.G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology*, **117**(1-4), pp. 275–300.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2014). *An introduction to statistical learning: With applications in R.* New York: Springer.

Janes, V.J., Grabowski, R.C., Mant, J., Allen, D., Morse, J.L., Haynes, H. (2017). The Impacts of Natural Flood Management Approaches on In-Channel Sediment Quality. *River Research and Applications*, **33**(1), pp. 89–101.

Jeffries, R., Darby, S.E., Sear, D.A. (2003). The influence of vegetation and organic debris on flood-plain sediment dynamics: case study of a low-order stream in the New Forest, England. *Geomorphology*, **51**(1-3), pp. 61–80.

Jochner, M., Turowski, J.M., Badoux, A., Stoffel, M., Rickli, C. (2015). The role of log jams and exceptional flood events in mobilizing coarse particulate organic matter in a steep headwater stream. *Earth Surface Dynamics*, **3**(3), pp. 311–320.

Kail, J. (2003). Influence of large woody debris on the morphology of six central European streams. *Geomorphology*, **51**(1-3), pp. 207–223.

Kail, J., Hering, D., Muhar, S., Gerhard, M., Preis, S. (2007). The use of large wood in stream restoration: Experiences from 50 projects in Germany and Austria. *Journal of Applied Ecology*, **44**(6), pp. 1145–1155.

Katimon, A., Shahid, S., Abd Wahab, A.K., Ali, M.H. (2013). Hydrological behaviour of a drained agricultural peat catchment in the tropics. 1: Rainfall, runoff and water table relationships. *Hydrological Sciences Journal*, **58**(6), pp. 1297–1309.

Kay, A.L., Crooks, S.M., Pall, P., Stone, D.A. (2011). Attribution of Autumn/Winter 2000 flood risk in England to anthropogenic climate change: A catchment-based study. *Journal of Hydrology*, **406**(1-2), pp. 97–112.

Kay, A.L., Old, G.H., Bell, V.A., Davies, H.N., Trill, E.J. (2019). An assessment of the potential for natural flood management to offset climate change impacts. *Environmental Research Letters*, **14**(4).

Keller, E.A., Swanson, F.J. (1979). Effects of Large Organic Debris on Channel Form and Fluvial Processes in the Coastal Redwood Environment. *Earth Surface Processes*, **4**, pp. 361–380.

Keys, T.A., Govenor, H., Jones, C.N., Hession, W.C., Hester, E.T., Scott, D.T. (2018). Effects of large wood on floodplain connectivity in a headwater Mid-Atlantic stream. *Ecological Engineering*, **118**(March), pp. 134–142.

Kiang, J.E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I.K., Belleville, A., Sevrez, D., Sikorska, A.E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., Mason, R. (2018). A Comparison of Methods for Streamflow Uncertainty Estimation. *Water Resources Research*, **54**(10), pp. 7149–7176.

Kim, H., Sim, S.H., Lee, J., Lee, Y.J., Kim, J.M. (2017). Flood fragility analysis for bridges with multiple failure modes. *Advances in Mechanical Engineering*, **9**(3), pp. 1–11.

Kitts, D.R. (2010). The hydraulic and hydrological performance of large wood accumulation in a low-order forest stream. Ph.D. thesis, University of Southampton.
[Online] Available from: *http://eprints.soton.ac.uk/185791/*

Klaar, M.J., Carver, S., Kay, P. (2020). Land management in a post-Brexit UK: An opportunity for integrated catchment management to deliver multiple benefits? *Wiley Interdisciplinary Reviews: Water*, **7**(5), pp. 1–6.

Koffler, D., Gauster, T., Laaha, G. (2016). lfstat.
[Online] Available from: *https://cran.r-project.org/package=lfstat*

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics*, **54**, pp. 159–178.

Lacombe, G., Ribolzi, O., De Rouw, A., Pierret, A., Latsachak, K., Silvera, N., Dinh, R.P., Orange, D., Janeau, J.L., Soulileuth, B., Robain, H., Taccoen, A., Sengphaathith, P., Mouche, E., Sengtaheuanghoung, O., Duc, T.T., Valentin, C. (2016). Contradictory hydrological impacts of afforestation in the humid tropics evidenced by long-term field monitoring and simulation modelling. *Hydrology and Earth System Sciences*, **20**(7), pp. 2691–2704.

Lagasse, P., Clopper, P., Pagan-Ortiz, J., Zevenbergen, L., Arneson, L., Schall, J., Girard, L. (2009). Bridge Scour and Stream Instability Countermeasures: Experience, Selection, and Design Guidance-Third Edition. *Hydraulic Engineering Circular*, **2**(23), p. 376.

Lamb, R., Garside, P., Pant, R., Hall, J.W. (2019). A Probabilistic Model of the Economic Risk to Britain's Railway Network from Bridge Scour During Floods. *Risk Analysis*, **39**(11), pp. 2457–2478.

Lamb, R., Zaidman, M., Archer, D., Marsh, T., Lees, M. (2003). River Gauging Station Data Quality Classification (GSDQ). Technical report, Environment Agency, Bristol, UK.

Lancaster Environment Centre (2021). NERC Quantifying the likely magnitude of nature-based flood mitigation effects across large catchments (Q-NFM). [Online] Available from: *https://www.lancaster.ac.uk/lec/sites/qnfm/* [Accessed: 16 September 2021]

Lane, S.N. (2017). Natural flood management. *WIREs Water*, **4**(3), pp. 1–14.

Lane, S.N., Landström, C., Whatmore, S.J. (2011). Imagining flood futures: Risk assessment and management in practice. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **369**(1942), pp. 1784–1806.

Lee, Y.S., Scholtes, S. (2014). Empirical prediction intervals revisited. *International Journal of Forecasting*, **30**(2), pp. 217–234.

Leedal, D., Neal, J., Beven, K., Young, P., Bates, P. (2010). Visualization approaches for communicating real-time flood forecasting level and inundation information. *Journal of Flood Risk Management*, **3**(2), pp. 140–150.

Lin, J.Y., Cheng, C.T., Chau, K.W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, **51**(4), pp. 599–612.

Linstead, C. (2001). The effects of large woody debris accumulations on river hydraulics and implications for physical habitat. *Hydro-Ecology: Linking Hydrology and Aquatic Ecology*, **266**(266), pp. 91–99.

Linstead, C., Gumell, a.M. (1999). Large Woody Debris in British Headwater Rivers - Physical Habitat Role and Management Guidelines. Technical report, Environment Agency, Bristol.

Linstead, C., Gurnell, A.M. (1999). Large Wood Debris in British Headwater Rivers - Summary Report. Technical report, Environment Agency, Bristol.

Lo, H.W., Smith, M., Klaar, M., Woulds, C. (2021). Potential secondary effects of in-stream wood structures installed for natural flood management: A conceptual model. *Wiley Interdisciplinary Reviews: Water*, **8**(5), pp. 1–22.

Lyn, D.A., Cooper, T.J., Condon, C.A., Gan, L. (2007). Factors in Debris Accumulation at Bridge Piers. Technical report, Joint Transportation Research Program, Indiana Publication FHWA/IN/JTRP-2006/36 Department of Transportation and Purdue University, West Lafayette, Indiana.

Maddala, G.S., Lahiri, K. (2009). *Introduction to Econometrics*. 4 edition. Chichester: Wiley.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**(2), pp. 111–153.

Makridakis, S., Winkler, R.L. (1989). Sampling Distributions of Post-Sample Forecasting Errors. *Journal of the Royal Statistical Society*, **38**(2), pp. 331–342.

Mann, H.B., Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, **18**(1), pp. 50–60.

Manners, R.B., Doyle, M.W. (2008). A mechanistic model of woody debris jam evolution and its application to wood-based restoration and management. *River Research and Applications*, **24**(8), pp. 1104–1123.

Mao, L., Comiti, F. (2010). The effects of large wood elements during an extreme flood in a small tropical basin of Costa Rica. *WIT Transactions on Engineering Sciences*, **67**, pp. 225–236.

Marshall, M.R., Ballard, C.E., Frogbrook, Z.L., Solloway, I., Mcintyre, N., Reynolds, B., Wheater, H.S. (2014). The impact of rural land management changes on soil hydraulic properties and runoff processes: Results from experimental plots in upland UK. *Hydrological Processes*, **28**(4), pp. 2617–2629.

Mazzorana, B., Hübl, J., Zischg, A., Largiader, A. (2011). Modelling woody material transport and deposition in alpine rivers. *Natural Hazards*, **56**(2), pp. 425–449.

McCuen, R.H., Knight, Z., Cutter, A.G. (2006). Evaluation of the Nash- Sutcliffe Efficiency Index. *Journal of Hydrologic Engineering*, **11**(6), pp. 597–602.

McLean, L., Beevers, L.C., Pender, G., Haynes, H., Wilkinson, M. (2013). Natural Flood Management in the United Kingdom: Developing a Conceptual Management Tool. In *35th IAHR World Congress*. Chengdu, China: IAHR, pp. 4996–5004.

Met Office (2018). Summer 2018. Technical report, Met Office.

Met Office (2020). Met Office: Climate averages 1971- 2000. [Online] Available from: *http://www.metoffice.gov.uk/climate/* [Accessed: 20 May 2021]

Met Office (2021). UK Storm Centre. [Online] Available from: *https://www.metoffice.gov.uk/weather/* [Accessed: 5 July 2021]

Metcalfe, P., Beven, K., Hankin, B., Lamb, R. (2017). A modelling framework for evaluation of the hydrological impacts of nature- based approaches to flood risk management, with application to in-channel interventions across a 29-km2 scale catchment in the United Kingdom. *Hydrological Processes*, **31**(9), pp. 1734–1748.

Metcalfe, P., Beven, K., Hankin, B., Lamb, R. (2018). A new method, with application, for analysis of the impacts on flood risk of widely distributed enhanced hillslope storage. *Hydrology and Earth System Sciences*, **22**(4), pp. 2589–2605.

Millington, C.E., Sear, D.A. (2007). Impacts of river restoration on small-wood dynamics in a low-gradient headwater stream. *Earth Surface Processes and Landforms*, **32**(8), pp. 1204–1218.

Montgomery, D.R., Buffington, J.M. (1997). Channel-reach morphology in mountain drainage basins. *Bulletin of the Geological Society of America*, **109**(5), pp. 596–611.

Montgomery, D.R., Collins, B.D., Buffington, J.M., Abbe, T.B. (2003). Geomorphic effects of wood in rivers. *Ecology and Management of Wood in World Rivers*, **37**, pp. 21–47.

Moore, R.D. (2005). Introduction to salt dilution gauging for streamflow measurement: Part 1. *Streamline*, **8**(4), pp. 1–6.

Moors for the Future Partnership (2015). An appraisal of the Defra Multi-Objective Flood Management Projects , December 2015. Technical Report December, Moors for the Future Partnership.
[Online] Available from: *http://www.uplandhydrology.org.uk/wp-content/*

Morris, J., Brewin, P. (2014). The impact of seasonal flooding on agriculture: The spring 2012 floods in Somerset, England. *Journal of Flood Risk Management*, **7**(2), pp. 128–140.

Morris, J., Hess, T.M. (1988). Agricultural flood alleviation benefit assessment - a case study. *Journal of Agricultural Economics*, **39**(3), pp. 402–412.

Mott, N. (2006). Managing Woody Debris in Rivers, Streams & Floodplains. Technical report, Staffordshire Wildlife Trust, Stafford.
[Online] Available from: *www.staffs-wildlife.org.uk*

Mott, N. (2011). Fish Live in Trees Too! River Rehabilitation and Large Woody Debris. Technical report, The Wildlife Trusts.

Mudelsee, M., Börngen, M., Tetzlaff, G., Grünewald, U. (2003). No upward trends in the occurrence of extreme floods in central Europe. *Nature*, **425**(6954), pp. 166–169.

Nakamura, F., Swanson, F.J. (1993). Effects of coarse woody debris on morphology and sediment storage of a mountain stream system in western Oregon. *Earth Surface Processes and Landforms*, **18**(1), pp. 43–61.

National Trust (2015). From source to sea - Natural Flood Management - The Holnicote Experience. Technical report, National Trust.

Nesshöver, C., Assmuth, T., Irvine, K.N., Rusch, G.M., Waylen, K.A., Delbaere, B., Haase, D., Jones-Walters, L., Keune, H., Kovacs, E., Krauze, K., Külvik, M., Rey, F., van Dijk, J., Vistad, O.I., Wilkinson, M.E., Wittmer, H. (2016). The science, policy and practice of nature-based solutions: An interdisciplinary perspective. *Science of The Total Environment*, **579**, pp. 1215–1227.

Ng, S., Perron, P. (1995). Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Association*, **90**(429), p. 268.

Nisbet, T., Marrington, S., Thomas, H., Broadmeadow, S., Valantin, G. (2011). Project RMP5455: Slowing the Flow at Pickering - Phase 1. Technical report, Department for Environment, Food & Rural Affairs, London.

Nisbet, T., Row, P., Marrington, S., Thomas, H., Broadmeadow, S., Valatin, G. (2015a). Project RMP5455: Slowing the Flow at Pickering - Phase 2. Technical report, Department for Environment, Food & Rural Affairs.

Nisbet, T., Thomas, H., Roe, P. (2015b). Case study 12 . Slowing the Flow at Pickering. Technical report, Defra, Environment Agency.

North York Moors National Park (2015). The official blog for the North York Moors National Park - Slowing down.
[Online] Available from: *https://northyorkmoorsnationalpark.wordpress.com/tag/pickering/*
[Accessed: 3 February 2017]

Noymanee, J., Theeramunkong, T. (2019). Flood Forecasting with Machine Learning Technique on Hydrological Modeling. *Procedia Computer Science*, **156**, pp. 377–386.

O'Connell, E., Ewen, J., O'Donnell, G., Quinn, P. (2007). Is there a link between agricultural land-use management and flooding? *Hydrology and Earth System Sciences*, **11**(1), pp. 96–107.

Odoni, N.A., Lane, S.N. (2010). Assessment of the Impact of Upstream Land Management Measures on Flood Flows in Pickering Beck using Overflow. Technical report, Durham University, Durham.

O'Driscoll, C., O'Connor, M., Asam, Z., de Eyto, E., Brown, L.E., Xiao, L. (2016). Forest clearfelling effects on dissolved oxygen and metabolism in peatland streams. *Journal of Environmental Management*, **166**, pp. 250–259.

Okiy, S., Chukwuemeka Nwobi-Okoye, C., Clement Igboanugo, A. (2015). Transfer Function Modelling: A Literature Survey. *Research Journal of Applied Sciences, Engineering and Technology*, **11**(11), pp. 1265–1279.

Padiyedath Gopalan, S., Kawamura, A., Amaguchi, H., Takasaki, T., Azhikodan, G. (2019). A bootstrap approach for the parameter uncertainty of an urban-specific rainfall-runoff model. *Journal of Hydrology*, **579**(August), p. 124195.

Pant, P.N., Starbuck, W.H. (1990). Innocents in the Forest: Forecasting and Research Methods. *Journal of Management*, **16**(2), pp. 433–460.

Parliamentary Office of Science and Technology (2011). Natural Flood Management. Technical report, Houses of Parliament.

Pastorello, G., Agarwal, D., Papale, D., Samak, T., Trotta, C., Ribeca, A., Poindexter, C., Faybishenko, B., Gunter, D., Hollowgrass, R., Canfora, E. (2014). Observational data patterns for time series data quality assessment. *Proceedings - 2014 IEEE 10th International Conference on eScience, eScience 2014*, **1**(October), pp. 271–278.

Pattison, I., Lane, S.N., Hardy, R.J., Reaney, S.M. (2014). The role of tributary relative timing and sequencing in controlling large floods. *Water Resources Research*, **50**(7), pp. 5444–5458.

Peters, R., Missildine, B., Low, D. (1998). Seasonal fish densities near river banks stabilized with various satbilization methods. Technical report, U.S. Fish and Wildlife Service, Washington DC.
[Online] Available from: *https://www.fws.gov/wafwo/*

Piegay, H., Gurnell, A.M. (1997). Large woody debris and river geomorphological pattern : examples from S. E. France and S. England. *Geomorphology*, **19**, pp. 99–116.

Pilotto, F., Bertoncin, A., Harvey, G.L., Wharton, G., Pusch, M.T. (2014). Diversification of stream invertebrate communities by large wood. *Freshwater Biology*, **59**(12), pp. 2571–2583.

Pinter, N., Van der Ploeg, R.R., Schweigert, P., Hoefer, G. (2006). Flood magnification on the River Rhine. *Hydrological Processes*, **20**(1), pp. 147–164.

Pinto, C., Ing, R., Browning, B., Delboni, V., Wilson, H., Martyn, D., Harvey, G.L. (2019). Hydromorphological, hydraulic and ecological effects of restored wood: findings and reflections from an academic partnership approach. *Water and Environment Journal*, **33**(3), pp. 353–365.

Piotrowski, A.P., Napiorkowski, J.J. (2013). A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *Journal of Hydrology*, **476**, pp. 97–111.

Pitt, M. (2008). The Pitt review: Learning lessons from the 2007 floods. *Floods Review*, **19**, pp. 1–205.

Pollock, D. (1992). Lectures in time-series analysis and forecasting. [Online] Available from: *https://www.le.ac.uk/* [Accessed: 19 March 2021]

Porter, K. (2020). A Beginner ' s Guide to Fragility, Vulnerability, and Risk. Technical report, University of Colorado Boulder. [Online] Available from: *https://www.sparisk.com/pubs*

Porter, K., Kennedy, R., Bachman, R. (2007). Creating fragility functions for performance-based earthquake engineering. *Earthquake Spectra*, **23**(2), pp. 471–489.

Posthumus, H., Hewett, C.J., Morris, J., Quinn, P.F. (2008). Agricultural land use and flood risk management: Engaging with stakeholders in North Yorkshire. *Agricultural Water Management*, **95**(7), pp. 787–798.

Posthumus, H., Morris, J., Hess, T.M., Neville, D., Phillips, E., Baylis, A. (2009). Impacts of the summer 2007 floods on agriculture in England. *Journal of Flood Risk Management*, **2**(3), pp. 182–189.

Potter, K. (1991). Hydrological impacts of changing land management practices in a moderately sized agricultural catchment. *Water Resources Research*, **27**(5), pp. 845–855.

Pregnolato, M., Galasso, C., Parisi, F. (2015). A compendium of existing vulnerability and fragility relationships for flood: Preliminary results. In *12th International Conference on Applications of Statistics and Probability in Civil Engineering*. Vancouver, pp. 1–8.

Purseglove, J. (2015). *Taming the flood - rivers, wetlands and the centuries-old battle against flooding*. Harper Collins UK.

Quinn, P., O'Donnell, G., Nicholson, A., Wilkinson, M., Owen, G., Jonczyk, J., Barber, N., Hardwick, M., Davies, G. (2013). Potential Use of Runoff Attenuation Features in Small Rural Catchments for Flood Mitigation. (June), p. 34.

R Core Team (2020). A Language and Environment for Statistical Computing. [Online] Available from: *http://www.r-project.org*

Rahm, E., Do, H.H. (2000). Data Cleaning: Problems and Current Approaches. In *IEEE Techn. Bulletin on Data Engineering*, 31. pp. 1–11.

Ramsbottom, D., Sayers, P., Wicks, J. (2005). Development of a Decision Support system for a risk based approach to catchment, estuary and coastal flood management planning MDSF2. Technical report, Defra/Environment Agency, Bristol.

Ramsbottom, D.M., Whitlow, C.D. (2003). Extension of Rating Curves at Gauging Stations Best Practice Guidance. Technical report, Environment Agency, Bristol.

Ratto, M., Young, P.C., Romanowicz, R., Pappenberger, F., Saltelli, A., Pagano, A. (2007). Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrology and Earth System Sciences*, **11**(4), pp. 1249–1266.

Rayer, S., Smith, S.K., Tayman, J. (2009). Empirical prediction intervals for county population forecasts. *Population Research and Policy Review*, **28**(6), pp. 773–793.

Reistad, K.S., Petersen-Øverleir, A., Bogetveit, L. (2007). Setting up rating curves using HEC-RAS. *Nor. Water Assoc,*, **3**, pp. 20–30.

Rhea, L., Jarnagin, T., Hogan, D., Loperfido, J.V., Shuster, W. (2015). Effects of urbanization and stormwater control measures on streamflows in the vicinity of Clarksburg, Maryland, USA. *Hydrological Processes*, **29**(20), pp. 4413–4426.

River Restoration Centre (2017). UK Projects Map. [Online] Available from: *http://www.therrc.co.uk/uk-projects-map* [Accessed: 31 May 2017]

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **40**(8), pp. 913–929.

Robinson, M. (1986). Changes in catchment runoff following drainage and afforestation. *Journal of Hydrology*, **86**(1-2), pp. 71–84.

Robinson, M., Cognard-Plancq, A.L., Cosandey, C., David, J., Durand, P., Führer, H.W., Hall, R., Hendriques, M.O., Marc, V., McCarthy, R., McDonnell, M., Martin, C., Nisbet, T., O'Dea, P., Rodgers, M., Zollner, A. (2003). Studies of the impact of forests on peak flows and baseflows: A European perspective. *Forest Ecology and Management*, **186**(1-3), pp. 85–97.

Robinson, M., Dupeyrat, A. (2005). Effects of commercial timber harvesting on streamflow regimes in the Plynlimon catchments, mid-Wales. *Hydrological Processes*, **19**(6), pp. 1213–1226.

Robison, E., Beschta, R. (1990). Coarse woody debris and channel morphology interactions for undisturbed streams in southeast Alaska, USA. *Earth Surface Processes and Landforms*, **15**(2), pp. 149–156.

Romanowicz, R.J., Young, P.C., Beven, K.J., Pappenberger, F. (2008). A data based mechanistic approach to nonlinear flood routing and adaptive flood level forecasting. *Advances in Water Resources*, **31**(8), pp. 1048–1056.

Roni, P., Beechie, T., Pess, G., Hanson, K. (2014). Wood placement in river restoration: Fact, fiction, and future direction. *Canadian Journal of Fisheries and Aquatic Sciences*, **72**(3), pp. 466–478.

Rosgen, D.L. (1994). A classification of natural rivers. *Catena*, **22**(3), pp. 169–199.

RRC (2009). Modifying River Bed Levels, Water Levels and Flows. In *Manual of River Restoration Techniques*, volume 5, chapter 5.6. Cranfield, pp. 1–5.

RRC (2013). Revetting and Supporting River Banks. In *Manual of river restoration techniques*, volume 4, chapter 4.8. Cranfield, pp. 1–4.

Ruiz-Villanueva, V., Diez-Herrero, A., Bodoque, J.M., Ernest Blade (2014). Large wood in rivers and its influence on flood hazard. *Cuadernos de Investigación Geográfica*, **40**(1), pp. 229–246.

Ruiz-Villanueva, V., Piégay, H., Gurnell, A.M., Marston, R.A., Stoffel, M. (2016). Recent advances quantifying the large wood dynamics in river basins: New methods and remaining challenges. *Reviews of Geophysics*, **54**(3), pp. 611–652.

Saldi-Caromile, K., Bates, K., Skidmore, P., Barent, J., Pineo, D. (2004). Stream habitat restoration guidelines. Technical report, Washington Departments of Fish and Wildlife and Ecology and the U.S. Fish and Wildlife Service, Olympia, Washington.

Samuels, P. (1989). Backwater lengths in rivers. In *Proceedings — Institution of Civil Engineers, Part 2, Research and Theory*, 87. pp. 571–582.

Sauer, V.B., Turnipseed, D.P. (2010). Stage Measurement at Gaging Stations. In *U.S. Geological Survey Techniques and Methods book 3*, chapter A7. Reston, VA: U.S. Geological Survey, p. 45.
[Online] Available from: *http://pubs.usgs.gov/tm/tm3-a7/*

Sayers, P.B., Hall, J.W., Meadowcroft, I.C. (2002). Towards risk-based flood hazard management in the UK. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, **150**(5), pp. 36–42.

Schalko, I., Lageder, C., Schmocker, L., Weitbrecht, V., Boes, R.M. (2019). Laboratory Flume Experiments on the Formation of Spanwise Large Wood Accumulations: I. Effect on Backwater Rise. *Water Resources Research*, **55**(6), pp. 4854–4870.

Schaller, N., Kay, A.L., Lamb, R., Massey, N.R., Van Oldenborgh, G.J., Otto, F.E., Sparrow, S.N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S.M., Haustein, K., Huntingford, C., Ingram, W.J., Jones, R.G., Legg, T., Miller, J., Skeggs, J., Wallom, D., Weisheimer, A., Wilson, S., Stott, P.A., Allen, M.R. (2016). Human influence on climate in the 2014 southern England winter floods and their impacts. *Nature Climate Change*, **6**(6), pp. 627–634.

Schanze, J. (2017). Nature-based solutions in flood risk management – Buzzword or innovation? *Journal of Flood Risk Management*, **10**(3), pp. 281–282.

Schmocker, L., Weitbrecht, V. (2013). Driftwood : Risk Analysis and Engineering Measures. *Journal of Hydraulic Engineering*, **139**(7), pp. 683–695.

Schultz, M., Gouldby, B., Simm, J., Wibowo, J. (2010). Beyond the Factor of Safety - Developing Fragility Curves to Characterize System Reliability. Technical Report July, US Army Corps of Engineers, Washington DC. [Online] Available from: *http://deltacouncil.ca.gov/sites/*

Schwert, G.W. (2002). Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, **20**(1), pp. 5–17.

Sear, D.A., Millington, C.E., Kitts, D.R., Jeffries, R. (2010). Logjam controls on channel:floodplain interactions in wooded catchments and their role in the formation of multi-channel patterns. *Geomorphology*, **116**(3-4), pp. 305–319.

Searcy, J.K., Hardison, C.H. (1960). Double-Mass Curves. In *Manual of Hydrology: Part 1. General Surface-Water Techniques*. Washington: US Government

Printing Office, pp. 1–66.
[Online] Available from: *http://dspace.udel.edu:8080/dspace/*

Sene, K., Tilford, K. (2004). Review of Transfer Function Modelling for Fluvial Flood Forecasting. Technical report, Defra/ Environment Agency, Bristol.
[Online] Available from: *https://www.gov.uk/government/publications/*

SEPA (2009). Engineering the Water Environment Good Practice Guide: Riparian Vegetation Management. Technical report, SEPA, Stirling.

Sharma, S., Swayne, D.A., Obimbo, C. (2015). Automating the Smoothing of Time Series Data. *Journal of Environmental & Analytical Toxicology*, **05**(05).

Shaw, E.M., Beven, K.J., Chappell, N.A., Lamb, R. (2011). *Hydrology in Practice*. 4th edition. Oxon: Spon Press.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, pp. 117–126.

Shields, F.D., Alonso, C.V. (2012). Assessment of flow forces on large wood in rivers. *Water Resources Research*, **48**(4), pp. 1–16.

Shields, F.D.J., Gippel, C.J. (1995). Prediction of effects of woody debris removal on flow resistance. *Journal of Hydraulic Engineering*, **121**(4), pp. 341–354.

Sholtes, J.S., Doyle, M.W. (2011). Effect of Channel Restoration on Flood Wave Attenuation. *Journal of Hydraulic Engineering*, **137**(2), pp. 196–208.

Shuster, W., Rhea, L. (2013). Catchment-scale hydrologic implications of parcel-level stormwater management (Ohio USA). *Journal of Hydrology*, **485**, pp. 177–187.

Shuttleworth, E.L., Evans, M.G., Pilkington, M., Spencer, T., Walker, J., Milledge, D., Allott, T.E. (2019). Restoration of blanket peat moorland delays stormflow from hillslopes and reduces peak discharge. *Journal of Hydrology X*, **2**, pp. 1–11.

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny.
[Online] Available from: *https://plotly-r.com. https://plotly-r.com*

Simm, J., Gouldby, B., Sayers, P., Flikweert, J., Wersching, S., Bramley, M. (2009). Representing fragility of flood and coastal defences: Getting into the detail. In S. et al. (Ed.), *Flood Risk Management: Research and Practice*. London: Taylor & Francis Group, pp. 621–631.

Simm, J., Tarrant, O. (2018). Development of fragility curves to describe the performance of UK levee systems. *26th International Congress on Large Dams*, pp. 495–514.

Skinner, K.S., Bruce-Burgess, L. (2005). Strategic and Project Level River Restoration Protocols- Key Components for Meeting the Requirements of the Water Framework Directive (WFD). *Water and Environment Journal*, **19**(2), pp. 135–142.

Slow the Flow Calderdale (2017). A Natural Flood Management pilot project at Crimsworth Dean beck, Hardcastle Crags and Hebden Bridge, West Yorkshire. Technical report, Slow the Flow Calderdale.
[Online] Available from: *http://slowtheflow.net/wp-content/*

Smith, B., Clifford, N.J., Mant, J. (2014). The changing nature of river restoration. *WIREs Water*, **1**, pp. 249–261.

Smith, E.P. (2002). BACI Design. *Encyclopedia of Environmetrics*, **1**, pp. 141–148.

Som, N.A., Zégre, N.P., Ganio, L.M., Skaugset, A.E. (2012). Corrected prediction intervals for change detection in paired watershed studies. *Hydrological Sciences Journal*, **57**(1), pp. 134–143.

Soulsby, C., Dick, J., Scheliga, B., Tetzlaff, D. (2017). Taming the flood- How far can we go with trees? *Hydrological Processes*, **31**(17), pp. 3122–3126.

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American journal of psychology*, **15**(1), pp. 72–101.

Ssegane, H., Amatya, D.M., Chescheir, G.M., Skaggs, W.R., Tollner, E.W., Nettles, J.E. (2013). Consistency of Hydrologic Relationships of a Paired Watershed Approach. *American Journal of Climate Change*, **02**, pp. 147–164.

Stewart-Oaten, A., Murdoch, W.W., Parker, K.R. (1986). Environmental Impact Assessment: "Pseudoreplication" in Time? *Ecology*, **67**(4), pp. 929–940.

Stratford, C., Miller, J., House, A., Old, G., Acreman, M., Dueñas-Lopez, M.A., Nisbet, T., Newman, J., Burgess- Gamble, L., Chappell, N., Clarke, S., Leeson, L., Monbiot, G., Paterson, J., Robinson, M., Rogers, M., Tickner, D. (2017). Do Trees in UK- Relevant River Catchments Influence Fluvial Flood Peaks? A Systematic review. Technical Report 1, Centre for Ecology & Hydrology, Wallingford.
[Online] Available from: *http://nora.nerc.ac.uk/id/*

Swanson, F.J., Gregory, S.V., Iroumé, A., Ruiz-Villanueva, V., Wohl, E. (2021). Reflections on the history of research on large wood in rivers. *Earth Surface Processes and Landforms*, **46**(1), pp. 55–66.

Tashman, L.J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, **16**(4), pp. 437–450.

Taylor, J.W., Bunn, D.W. (1999). Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: A simulation study. *International Journal of Forecasting*, **15**(3), pp. 325–339.

Tencaliec, P., Favre, A.C., Prieur, C., Mathevet, T. (2015). Reconstruction of missing daily streamflow data using dynamic regression models. *Water Resources Research*, **51**, pp. 9447–9463.

The Rivers Trust (2021). The Rivers Trust NFM Project Monitoring & Evaluation Tool.
[Online] Available from: *https://www.arcgis.com/apps/* [Accessed: 24 August 2021]

Thirumalaiah, K., Deo, M.C. (1998). River Stage Forecasting Using Artificial Neural Networks. *Journal of Hydrologic Engineering*, **3**(1), pp. 26–32.

Thomas, H., Nisbet, T.R. (2007). An assessment of the impact of floodplain woodland on flood flows. *Water and Environment Journal*, **21**(2), pp. 114–126.

Thomas, H., Nisbet, T.R. (2012). Modelling the hydraulic impact of reintroducing large woody debris into watercourses. *Journal of Flood Risk Management*, **5**(2), pp. 164–174.

Thompson, V., Dunstone, N.J., Scaife, A.A., Smith, D.M., Slingo, J.M., Brown, S., Belcher, S.E. (2017). High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, **8**(1), pp. 1–6.

Trotter, E.H. (1990). Woody Debris, Forest-Stream Succession, and Catchment Geomorphology. *Journal of the North American Benthological Society*, **9**(2), pp. 141–156.

UK Centre for Ecology and Hydrology (2021). National River Flow Archive. [Online] Available from: *https://nrfa.ceh.ac.uk/* [Accessed: 25 August 2021]

Underwood, A.J. (1994). On Beyond BACI: Sampling Designs that Might Reliably Detect Environmental Disturbances. *Ecological Applications*, **4**(1), pp. 3–15.

University of Reading (2021). NERC Natural Flood Management Research Programme. [Online] Available from: *https://research.reading.ac.uk/nerc-nfm/* [Accessed: 6 September 2021]

USACE (2020). HEC-RAS 5.0.7. [Online] Available from: *http://www.hec.usace.army.mil/*

Uttley, C., Skinner, A. (2017). CIEEM Webinar: Natural Flood Management: policy and practical implementation. [Online] Available from: *http://www.cieem.net/events/*

van der Meer, J., ter Horst, W., van Velzen, E. (2008). Calculation of fragility curves for flood defence assets. In S. et al. (Ed.), *Flood Risk Management: Research and Practice*. London: Taylor & Francis Group, pp. 567–573.

Vira, B., Adams, W.M. (2009). Ecosystem services and conservation strategy: beware the silver bullet. *Conservation Letters*, **2**(4), pp. 158–162.

von Asmuth, J.R., Bierkens, M.F., Maas, K. (2002). Transfer function-noise modeling in continuous time using predefined impulse response functions. *Water Resources Research*, **38**(12), pp. 23.1–23.12.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, **5**(1), pp. 13–26.

Wallerstein, N., Thorne, C.R., Doyle, M.W. (1997). Spatial Distribution and impact of large woody debris in northern Mississippi. In S. Wang, E. Langendoen, J. Shields, F. D. (Eds.), *Management of Landscapes Disturbed by Channel Incision, Stabilization, Rehabilitation, and Restoration.* Mississippi: University of Mississippi, pp. 145–150.

Wallerstein, N.P., Thorne, C.R. (2004). Influence of large woody debris on morphological evolution of incised, sand-bed channels. *Geomorphology*, **57**(1-2), pp. 53–73.

Watson, F., Vertessy, R., McMahon, T., Rhodes, B., Watson, I. (2001). Improved methods to assess water yield changes from paired-catchment studies: Application to the Maroondah catchments. *Forest Ecology and Management*, **143**(1-3), pp. 189–204.

Waylen, K.A., Holstead, K.L., Colley, K., Hopkins, J. (2018). Challenges to enabling and implementing Natural Flood Management in Scotland. *Journal of Flood Risk Management*, **11**, pp. 1078–1089.

Wells, J., Labadz, J.C., Smith, A., Islam, M.M. (2020). Barriers to the uptake and implementation of natural flood management: A social-ecological analysis. *Journal of Flood Risk Management*, **13**(Suppl. 1), pp. 1–12.

Wensink, H.E., Broersen, P.M.T. (1994). Estimating the Kullbak-Leibler Information for autoregressive model order selection in finite samples. In *Signal Processing VII Proc. Eusipco Conf.* Edinburgh, pp. 1847–1850.

Wenzel, R., Reinhardt-Imjela, C., Schulte, A., Bölscher, J. (2014). The potential of in-channel large woody debris in transforming discharge hydrographs in headwater areas (Ore Mountains, Southeastern Germany). *Ecological Engineering*, **71**, pp. 1–9.

Wescoat, J.L., White, G.F. (2003). *Water for Life - Water Management and Environmental Policy.* Cambridge: Cambridge University Press.

Westerberg, I., Guerrero, J.L., Seibert, J., Beven, K.J., Halldin, S. (2011). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrological Processes*, **25**(4), pp. 603–613.

Wheater, H.S. (2006). Flood hazard and management: A UK perspective. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **364**(1845), pp. 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**(1), pp. 2511–2523.

Wilby, R.L., Clifford, N.J., De Luca, P., Harrigan, S., Hillier, J.K., Hodgkins, R., Johnson, M.F., Matthews, T.K., Murphy, C., Noone, S.J., Parry, S., Prudhomme, C., Rice, S.P., Slater, L.J., Smith, K.A., Wood, P.J. (2017). The 'dirty dozen' of freshwater science: detecting then reconciling hydrological data biases and errors. *Wiley Interdisciplinary Reviews: Water*, **4**(3), pp. 1–19.

Wilcox, A.C., Wohl, E.E. (2006). Flow resistance dynamics in step-pool stream channels: 1. Large woody debris and controls on total resistance. *Water Resources Research*, **42**(5).

Wild Trout Trust (2013). Advisory Visit for James Mawle and Stephen Mawle Coverhead Farm River Cover August 2013. Technical report, Wild Trout Trust.

Wilkinson, M.E., Addy, S., Quinn, P.F., Stutter, M. (2019). Natural flood management: small-scale progress and larger-scale challenges. *Scottish Geographical Journal*, **135**(1-2), pp. 23–32.

Wilkinson, M.E., Quinn, P.F., Barber, N.J., Jonczyk, J. (2014). A framework for managing runoff and pollution in the rural landscape using a Catchment Systems Engineering approach. *Science of the Total Environment*, **468-469**, pp. 1245–1254.

Wilkinson, M.E., Quinn, P.F., Welton, P. (2010). Runoff management during the September 2008 floods in the Belford catchment, Northumberland. *Journal of Flood Risk Management*, **3**(4), pp. 285–295.

Williams, W.H., Goodman, M.L. (1971). A simple method for the construction of empirical confidence limits for economic forecasts. *Journal of the American Statistical Association*, **66**(336), pp. 752–754.

Wingfield, T., Macdonald, N., Peters, K., Spees, J., Potter, K. (2019). Natural Flood Management: Beyond the evidence debate. *Area*, **51**(4), pp. 743–751.

Winsemius, H.C., Aerts, J.C., Van Beek, L.P., Bierkens, M.F., Bouwman, A., Jongman, B., Kwadijk, J.C., Ligtvoet, W., Lucas, P.L., Van Vuuren, D.P., Ward, P.J. (2016). Global drivers of future river flood risk. *Nature Climate Change*, **6**(4), pp. 381–385.

WMO (2009). Manual on Low-flow Estimation and Prediction. Technical report, WMO, Geneva.

Wohl, E. (2013). Floodplains and wood. *Earth-Science Reviews*, **123**, pp. 194–212.

Wohl, E., Bledsoe, B.P., Fausch, K.D., Kramer, N., Bestgen, K.R., Gooseff, M.N. (2016). Management of Large Wood in Streams: An Overview and Proposed Framework for Hazard Evaluation. *Journal of the American Water Resources Association*, **52**(2), pp. 315–335.

Wohl, E., Cadol, D. (2011). Neighborhood matters: Patterns and controls on wood distribution in old-growth forest streams of the Colorado Front Range, USA. *Geomorphology*, **125**(1), pp. 132–146.

Wohl, E., Lane, S.N., Wilcox, A.C. (2015). The science and practice of river restoration. *Water Resour. Res*, **51**, pp. 5974–5997.

Wood-Smith, R., Buffington, J. (1996). Multivariate geomorphic analysis of forest streams: implications for assessment of land use impacts on channel condition. *Earth Surface Processes and Landforms*, **21**(4), pp. 377–393.

Woodland Trust (2016). Natural Flood Management Guidance: Woody dams, deflectors and diverters. Technical Report 396, Woodland Trust, Grantham.

Wynne-Jones, S. (2016). Flooding and media storms – controversies over farming and upland land-use in the UK. *Land Use Policy*, **58**, pp. 533–536.

Wyżga, B., Mikuś, P., Zawiejska, J., Ruiz-Villanueva, V., Kaczka, R.J., Czech, W. (2017). Log transport and deposition in incised, channelized, and multithread reaches of a wide mountain river: Tracking experiment during a 20-year flood. *Geomorphology*, **279**, pp. 98–111.

Yamazaki, F., Motomura, H., Hamada, T. (1997). Damage assessment of expressway networks in Japan based on seismic monitoring. In *12th World Conference on Earthquake Engineering CD-ROM; Paper No. 0551.* pp. 1–8.
[Online] Available from: *http://www.iitk.ac.in/nicee/wcee/article/0551.pdf*

Yin, Z., Feng, Q., Wen, X., Deo, R.C., Yang, L., Si, J., He, Z. (2018). Design and evaluation of SVR, MARS and M5Tree models for 1, 2 and 3-day lead time forecasting of river flow data in a semiarid mountainous catchment. *Stochastic Environmental Research and Risk Assessment*, **32**(9), pp. 2457–2476.

Yorkshire Dales National Park Authority (2002). Landscape Character Assessment - Wensleydale.
[Online] Available from: *http://www.yorkshiredales.org.uk/about-the-dales/*
[Accessed: 18 October 2018]

Yorkshire Dales National Park Authority, Yorkshire Dales Rivers Trust, North Yorkshire City Council (2017). Natural Flood Management Measures – a practical guide for farmers. Technical report, Natural England, Environment Agency.
[Online] Available from: *https://thefloodhub.co.uk/wp-content/*

Yorkshire Dales Rivers Trust (2018). Naturally Resilient - Slowing the Movement of Water - Leaky Dams. Technical report, Interreg.
[Online] Available from: *https://www.ydrt.org.uk/wp-content/*

Young, P. (1986). Time series methods and recursive estimation in hydrological systems analysis. In D. Kraijenhoff, J. Moll (Eds.), *River Flow Modelling and Forecasting*. Dordrecht: D. Reidel, pp. 129–180.

Young, P. (2003). Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes*, **17**(11), pp. 2195–2217.

Young, P.C. (1978). A general theory of modeling for badly defined dynamic systems. In G.C. Vansteenkiste (Ed.), *Modeling, Identification and Control in Environmental Systems*. Amsterdam: North-Holland, pp. 103–135.

Young, P.C. (2002). Advances in real–time flood forecasting. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **360**, pp. 1433–1450.

Yuan, X., Xie, Z., Liang, M. (2009). Sensitivity of regionalized transfer-function noise models to the input and parameter transfer method. *Hydrological Sciences Journal*, **54**(3), pp. 639–651.

Appendix A

Stage-Discharge Rating Relationships

The smoothed and quality assured stage time series were converted to discharge through the means of rating relationships. The rating relationships were developed by calibrating 1D hydraulic models of the gauging sites to field measurements of discharge. All references for citations in the appendices are included in the references section in the main body of the thesis (p.189).

A.1 Method

Due to the small and turbulent nature of the study reaches, discharge measurements were measured using slug-injection dilution methods, following Moore (2005). Salt pulses were recorded using conductivity as a proxy at 1 second intervals using Campbell Scientific CR200 Data logger with a conductivity probe. A qualitative and a quantitative review of the suitability of the observed ratings for rating curve extension were carried out following the Environment Agency's best practice guidance (Ramsbottom and Whitlow, 2003)

To extrapolate the rating relationships beyond the gauged discharge the gauging sites were modelled using 1D hydraulic models. The channel cross-sections which formed the basis of the hydraulic models were surveyed in the Autumn-Winter of 2019/20 using a Leica TS15i total station (TPS) with a prism reflector.

Samuels' equation (Samuels, 1989) was used to determine the survey extent necessary to capture the length of backwater affecting the gauging location. The co-ordinates were recorded in a local co-ordinate system linked to the stage time series data by surveying the water level at the gauge.

The gauging sites were represented as 1D steady state models in HEC-RAS 5.0.7 (USACE, 2020), which solves the De Saint Venant equations for unsteady open channel flow. HEC-RAS is an industry standard hydraulic modelling software (USACE, 2020) which has been shown to reliably represent flood hydrographs in natural watercourses (Castellarin et al., 2009; Di Baldassarre and Montanari, 2009; Horritt and Bates, 2002) including steep watercourses (Reistad et al., 2007). The sites were represented as 1D features because in-channel and out of bank flows were approximately parallel and the steep gradient of the watercourse made hysteresis in the rating curves unlikely. HEC-RAS' mixed flow regime option was used to represent the subcritical and supercritical flows found in the modelled reaches.

The hydraulic loss coefficients were calibrated using the observed stage-discharge pairs. The friction coefficient, Manning's n was varied between 0.04 and 0.07 based on the values for mountain streams recommended by Chow (1959). In the step-pool reaches there were significant changes in elevation in bed level. In order to stabilise the models these were computationally represented as weir structures, as recommended in HEC-RAS documentation (Brunner, 2000). For each weir the shape of the surveyed cross-section was used as the weir geometry and the weir coefficient was calibrated within the range for natural weirs. The boundary condition at the furthest downstream cross-section were set as the hydraulic slope, assumed to be parallel to the surveyed bed-slope.

The fit of the modelled rating relationship to the data was assessed using the Pearson's correlation coefficient and root mean square error (RMSE). Bias in the relationships with respect to the gauged points was assessed visually. For each gauging site the best calibration run was chosen to provide a fit with the least bias at high flows with low RMSE and high Pearson's correlation coefficient.

A.1.1 Rating Relationship Uncertainty

Confidence intervals to represent the degree of uncertainty in the rating relationships were calculated using the methodology presented by Lamb et al. (2003) for the Environment Agency's river gauging station data quality (GSDQ) classification.

The standard error of the mean relationship (SMR) was calculated using Equations (A.1)-(A.2) in log space by comparing the gauged value of discharge, q_i^* (m^3/s), to the value of discharge estimated from the rating relationship, $\hat{q}^*$ (m^3/s), for each gauged value of stage, h_i^* (m). Finally, to obtain the 95% confidence interval the SMR at each value of gauged stage was multiplied by t_{95}, the critical value of the two-tailed Student t distribution with $N-2$ degrees of freedom, where N is the number of gauged points (equation (A.3)).

The SMR, and therefore the width of the confidence interval, is designed to vary with the distance of the stage of interest from the mean stage, h^*, of the gauged stage-discharge pairs to capture the increase in uncertainty further from the centre of the range of gauged stage. The 95% level of confidence interval was used because it is the recommended Hydrometric Uncertainty Guidance ISO standard (ISO/TS 25377:2007).

$$SEE = \frac{\sqrt{\sum_i (q_i^* - \hat{q}^*)^2}}{N-2} \tag{A.1}$$

$$SMR = SEE\sqrt{\frac{1}{N} + \frac{(h^* - E(h_i^*))^2}{\sum (h_i^* - E(h_i^*))^2}} \tag{A.2}$$

$$\hat{q}^* \pm t_{95}(SMR_{lnh}) \tag{A.3}$$

A.2 Results

A total of 146 stage - discharge pairs were collected across the study and main channel reaches over the monitoring period. The reaches LB6 (Downs Gill), LB7 (control) and LB8 (Fall Gill) were prioritised, resulting in an empirical rating curves with 10 to 26 points for each gauge. For the LB5 (Lock Gill) and main

channel gauges 7 points each were collected on average, covering less of the observed range (see Table A.1). Fitting either a power law or exponential curve to the data gives high values of R^2 for the gauged range; Table A.1 reports the highest R^2 of the two fits.

Table A.1: Summary of empirical rating data, R^2 refers to fit of power law or exponential relationship

Reach	gauging station	Range of stage gauged (%)	Number of stage-discharge pairs	R^2
LB5	Upstream	6	8	NA
(Lock Gill)	Middle	13	7	0.92
	Downstream	23	8	0.46
LB6	Upstream	70	20	0.95
(Downs Gill)	Downstream	76	26	0.98, 0.97
LB7	Upstream	46	21	0.92
(Control)	Downstream	38	10	0.98
LB8	Upstream	66	14	0.96
(Fall Gill)	Downstream	32	19	0.85

Rather than extrapolating relationships fitted to the stage-discharge pairs the rating curves for LB6, LB7, and LB8 were extended using 1D hydraulic models calibrated to the stage-discharge observations. The models were calibrated to the gauged data points using Manning's n of the channel and floodplain and weir coefficient of the hydraulic structures used to represent steps in bed level.

Goodness of fit measures were calculated for all of the model runs. Hence, the best fit could be determined based on the highest Pearson's correlation coefficient, or the lowest root mean square error (RMSE). However, as there are more data points at lower flows these measures give a better indication of the model fit at low flows than at high flows. As high flows are of particular interest in this study calibration runs which showed a better fit at high flows were chosen over calibration runs with the highest Pearson's correlation coefficient or lowest RMSE. Figure A.1 shows the rating relationships with the calibration data. The goodness of fit measures for the relationships are shown in Table A.2.

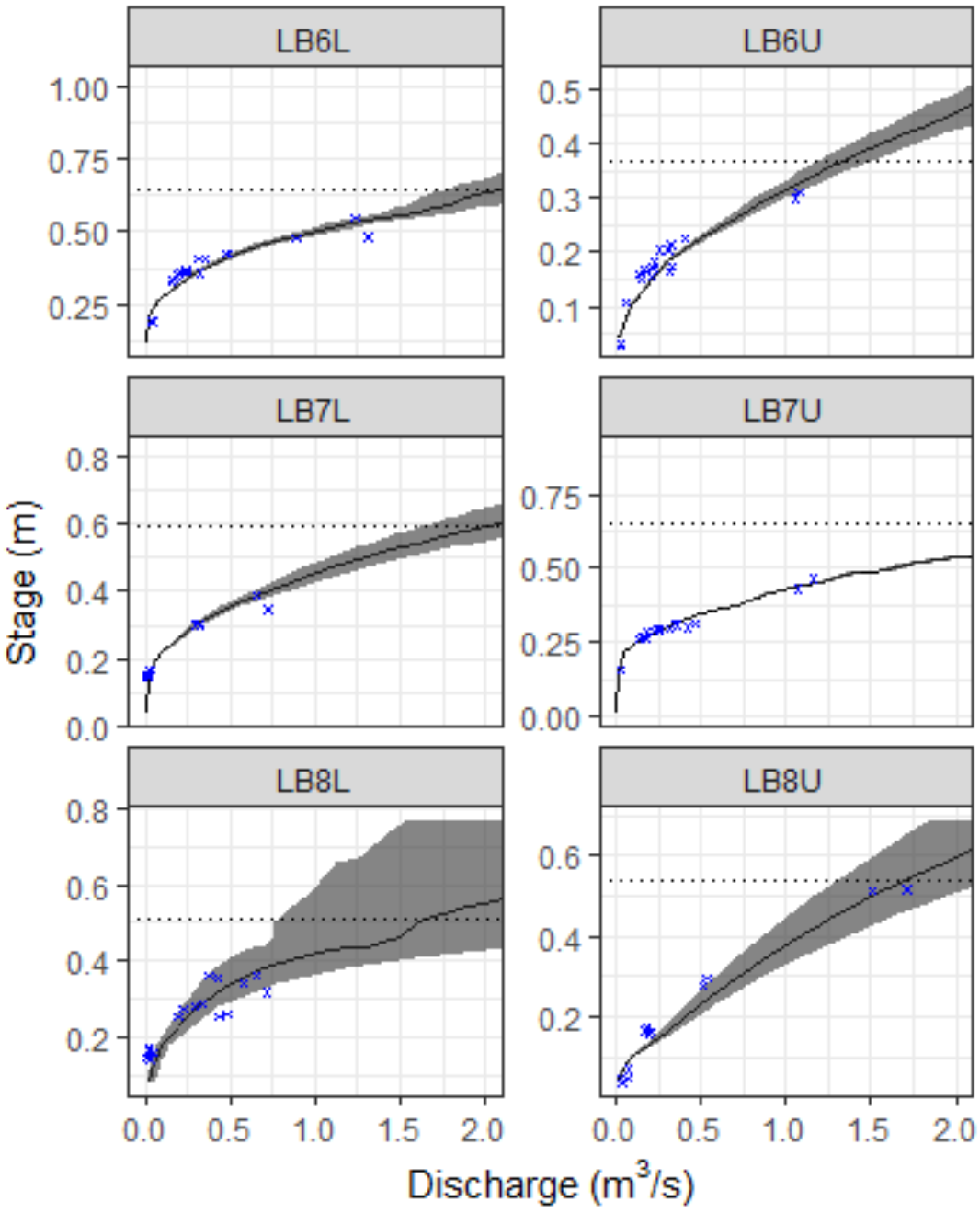


Figure A.1: Modelled and calibrated rating relationships with 95% Confidence Intervals (gauged points shown as blue 'x', dotted line indicates maximum observed stage)

Table A.2: Model calibration run with best fit at high flows

logger	channel n	weir coefficient	flood-plain n	Pearson's cor. coef.	d.o.f.	p value	RMSE
LB6U	0.06	1.4	0.03	0.97	16	5.88E-11	0.062
LB6L	0.06	1.4	0.03	0.93	15	4.03E-08	0.117
LB7U	0.07	1.4	0.03	0.98	19	1.41E-15	0.053
LB7L	0.07	1.4	0.03	0.97	9	1.16E-06	0.053
LB8U	0.06	1.4	0.03	0.99	12	1.49E-10	0.087
LB8L	0.06	1.4	0.03	0.85	17	3.15E-06	0.100

Appendix B

Data Quality

B.1 Introduction

To avoid data quality issues which can result in incorrect or misleading statistics (Rahm and Do, 2000) the data was subjected to a quality control process and potential sources of error were identified from the literature and field observations. Although the majority of uncertainty in discharge time series is usually attributed to the uncertainty in the rating curve the error in stage measurement can be significant (Horner et al., 2018). Sauer and Turnipseed (2010) identify datum errors due to vertical movements, stage reading errors due to water surge and instrument errors amongst the most common sources of error in the collection of stage data. All references for citations in the appendices are included in the references section in the main body of the thesis (p.189).

B.2 Quality Assurance Process

As recommended by Crochemore et al. (2020) quality assurance of the data was carried out by visual inspection of the data. After Pastorello et al. (2014), general trends and patterns were identified in the data to detect anomalous values through both single variable inspections/multi-variable inspection of correlated variables and detailed relationship examination. Single and multiple variable inspection was carried out using the free software environment R (R Core Team, 2020)

the data were interactively visualised at different scales and timeframes using the R package Plotly (Sievert, 2020). Detailed examination of the relationship between the upstream and downstream stage time series of each reach enabled the detection and inspection of anomalies in detail so that hypotheses could be developed about possible causes. Evidence was sought from field observations, notes, photography, spot discharge gaugings, aerial photography to support or reject the hypotheses. In the case of datum errors and rating relationship change, where sufficient evidence was available to support these hypotheses, a datum correction was made to the stage data.

First, all upstream stage data was plotted against all downstream stage data recorded on each stream during the baseline and post intervention monitoring periods (Figure B.1). Changes in the relationship between upstream and downstream stage can already be identified in Figure B.1. Assessing changes in the relationship between upstream and downstream stage as a quantification of the error in the data requires the assumption that the relationship between upstream and downstream stage is constant and does not vary for different types of events (e.g. short/long duration, single/multi-peaked, antecedent conditions) and the rising or falling limb of the hydrograph. These conditions relate to storage within the reach and are most likely satisfied due to the steep nature of the watercourses. In the post-intervention period, however, storage elements are introduced to the watercourse which may account for some of the spread in the data.

B.2.1 Downs Gill Example

To identify when these changes in the relationship occurred the upstream stage was plotted against the downstream stage for each month in the monitoring period for Downs Gill (Figure B.2). For example, Figure B.2 clearly shows that there was a shift in the relationship on Downs Gill during March 2019 which persisted to the end of the monitoring period.

The third step was to plot detailed plots of events during which changes in the relationship occurred at a range of temporal resolutions to identify the start and end point of the change if applicable. From here on the data QA process is

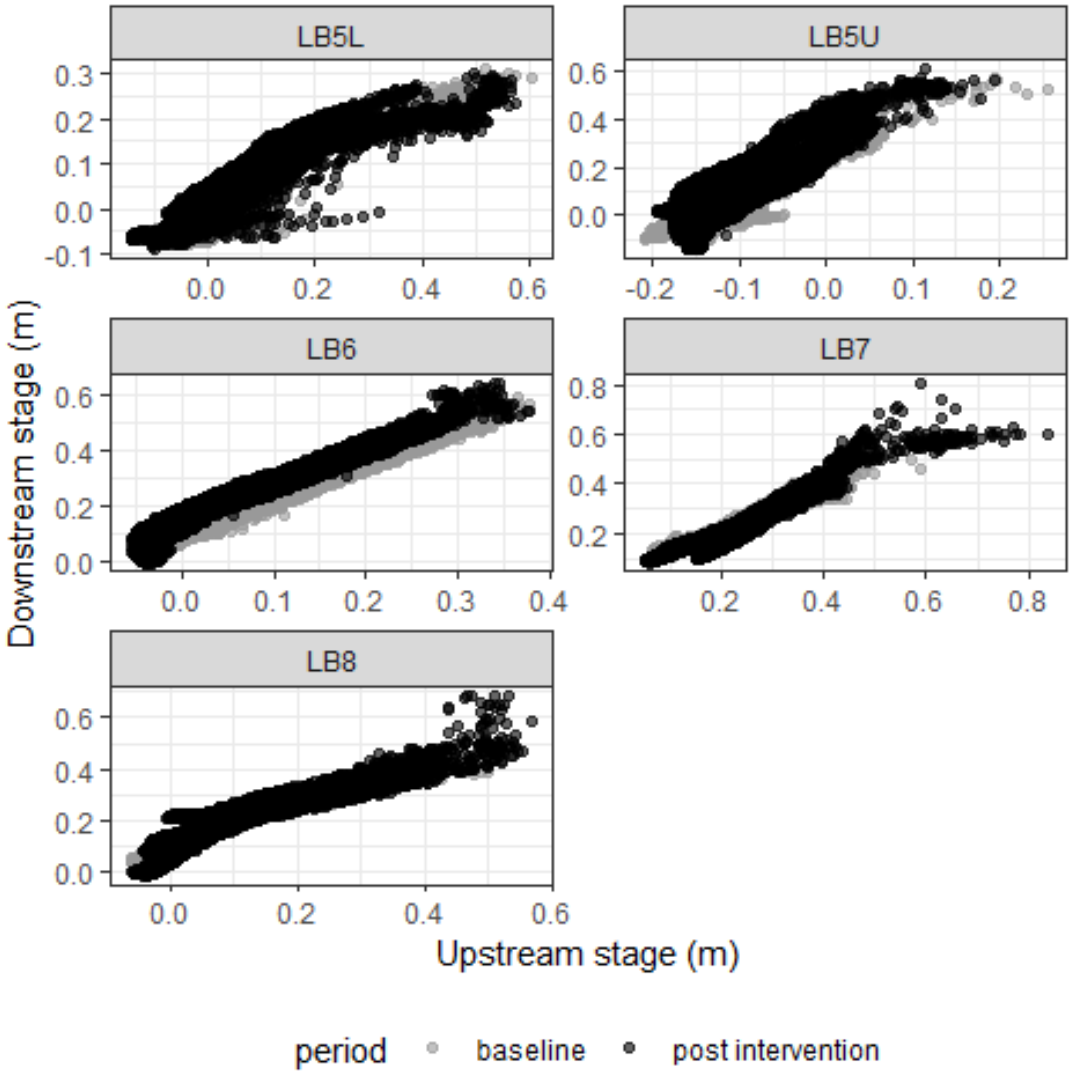


Figure B.1: Baseline and post-intervention monitoring period relationships between upstream and downstream stage on the study streams

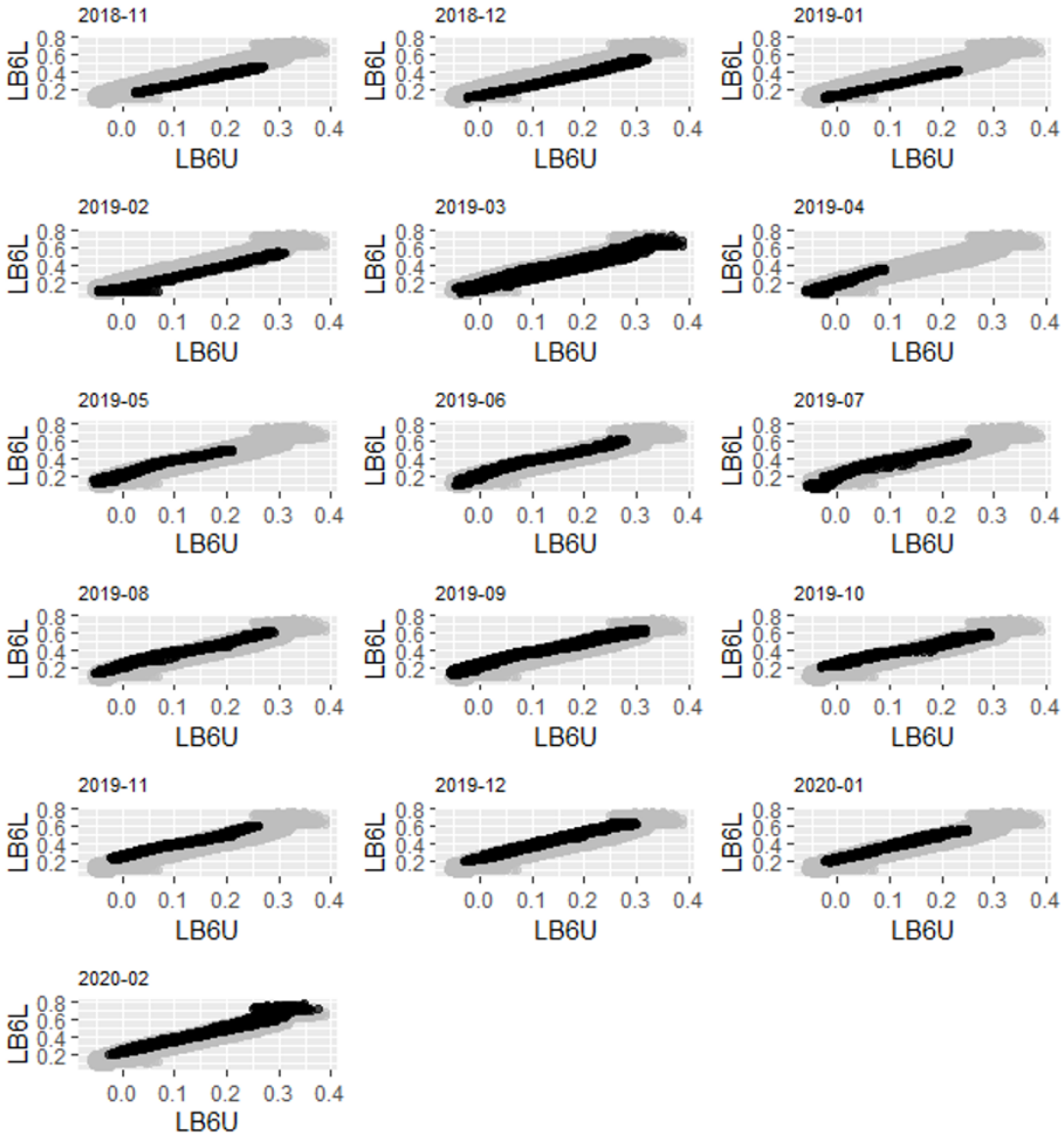


Figure B.2: Upstream stage plotted against downstream stage for every month in the post-intervention monitoring period on Downs Gill, the data for each month is given in black, overlying the data for the whole monitoring period in grey. LB6U refers to the upstream gauge, and LB6L refers to the downstream gauge on the stream

illustrated using a change which occurred on Downs Gill during Storm Gareth in March 2019 in the post-intervention monitoring period.

It can be seen in Figure B.3 that a datum change of approximately 0.1m occurred between the rising and falling limb of the hydrograph of Storm Gareth in March 2019. The time series plots shows that whilst the downstream logger returned to its original level by the 20th of March, the upstream logger dropped to a level approximately 0.1 m lower than its original stage. There can be three explanations for this change; either the upstream datum decreased by 0.1 m, the downstream datum increased by 0.1 m, or a combination of the two. The datum remained changed until the end of the monitoring period, so the change was permanent. To determine which of these changes had occurred physical explanations were sought.

An increase in the datum could be caused by damage to the stilling well, sedimentation or blockage of the stilling well, deposition of material in the gauging cross-section, or the backwater effect from a downstream blockage or cross section change. At the upstream logger a decrease in datum could be caused by scouring of the cross section, the removal of sedimentation or blockage at the gauge, or the removal of a downstream blockage causing a backwater effect. At the upstream logger scouring is unlikely as the gauging location is largely on bedrock, a backwater effect is also unlikely as there is a large drop in bed level after the gauge. Blockage at the gauge is unlikely as wrack was cleared from the loggers at every download and would have had to be present throughout the monitoring period before this point. Equally, sedimentation impacting the hanger length would have been noticeable at data download and has not been noted at this logger. Damage to the stilling well is the only remaining explanation for a permanent drop in datum at the logger. This has not been observed but could be too subtle to detect by eye. At the downstream logger an increase in stage can be explained by a backwater effect caused by the collapse of a footbridge approximately 5 m downstream of the logger (Figure B.4).

A change in datum is equivalent to a change in the rating relationship. The same discharge would have a different stage associated with it before and after the change in datum. The discharge-stage pairs spot-gauged before and after Storm Gareth are shown in Figure B.5 for both the upstream and downstream

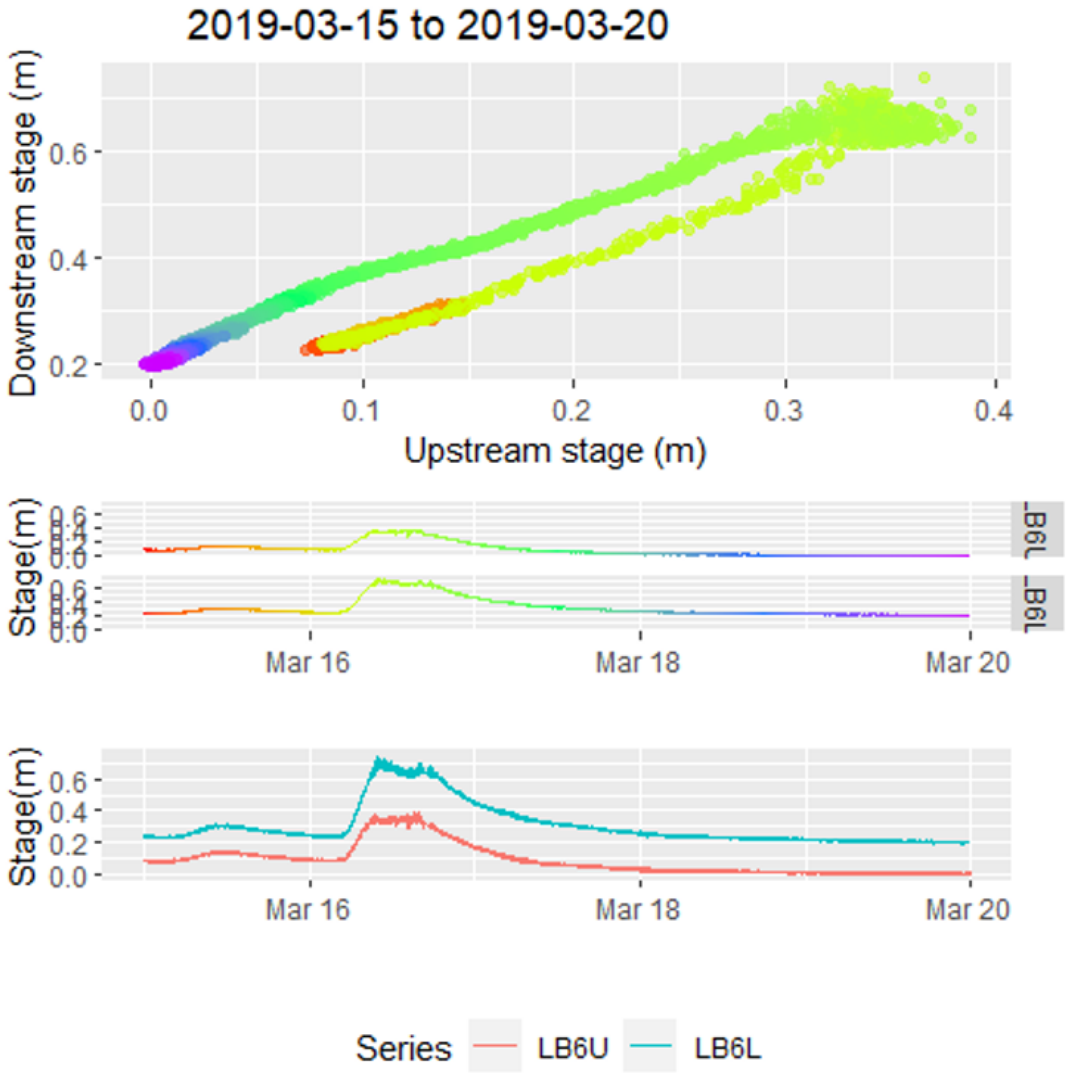


Figure B.3: Downs Gill Diagnostic plot for Storm Gareth (16 March 2019) The legend for the colours in the first plot are given as the time series shown in the second plot

Figure B.4: Collapsed footbridge causing a backwater effect at the downstream gauge on Downs Gill

gauge. Fitting curves to the points shows that there is indeed a change in the rating relationship at the downstream gauge, and not at the upstream gauge. The magnitude of the change varies with stage, as would be expected as the backwater effect increases as more of the collapsed bridge becomes submerged. At the top of the gauged range (1.3 m^3/s) the difference in stage before and after storm Gareth is 0.17 m, although the relationship is extrapolated above 0.9 m^3/s before Storm Gareth. At the top of the range gauged in both periods (0.9 m^3/s) the difference in stage is 0.15 m. Within the gauged range at 0.6 m^3/s the difference is 0.13 m, further reducing to 0.10 m and 0.04 m at 0.3 m^3/s and 0.1 m^3/s respectively. To improve confidence in the extrapolated relationship before Storm Gareth the rating curve is modelled in HEC-RAS 1D and a high flow point (2 m^3/s) is added and used to fit the logarithmic relationship. The rating curve after Storm Gareth is not added to for two reasons; higher flows (1.3 m^3/s) have been gauged after Storm Gareth so less extrapolation is required and modelling of the collapsed bridge in HEC-RAS 1D introduces uncertainties.

Fitting linear regression relationships to the data before and after Storm Gareth shows that whilst the data from the period before Storm Gareth is approximately linear ($R^2 = 0.99$), the data after Storm Gareth is less well explained by the linear relationship ($R^2 = 0.91$). This makes sense, as the impact of the collapsed bridge has not been observed to be linear in the rating relationships (Figure B.6).

Whether this difference in stage can be attributed to the difference in the rating relationship at the lower logger is assessed by comparing the differences in the two linear relationships shown in Figure B.6 with the difference between the logarithmic rating relationships at the downstream logger shown in Figure B.5. The differences are shown in Table B.1, it can be seen that at the bottom of the range of stage gauged on the lower gauge of Downs Gill the difference in the rating relationship and the difference in the upstream to downstream stage relationship before and after Storm Gareth is similar. At higher stages the rating relationship gives a smaller difference before and after Storm Gareth than the linear relationships between upstream and downstream logger suggest. However, as can be seen in Figure B.6 the linear relationships over-estimate the difference at high stage. This fit could be improved by fitting non-linear relationships to the

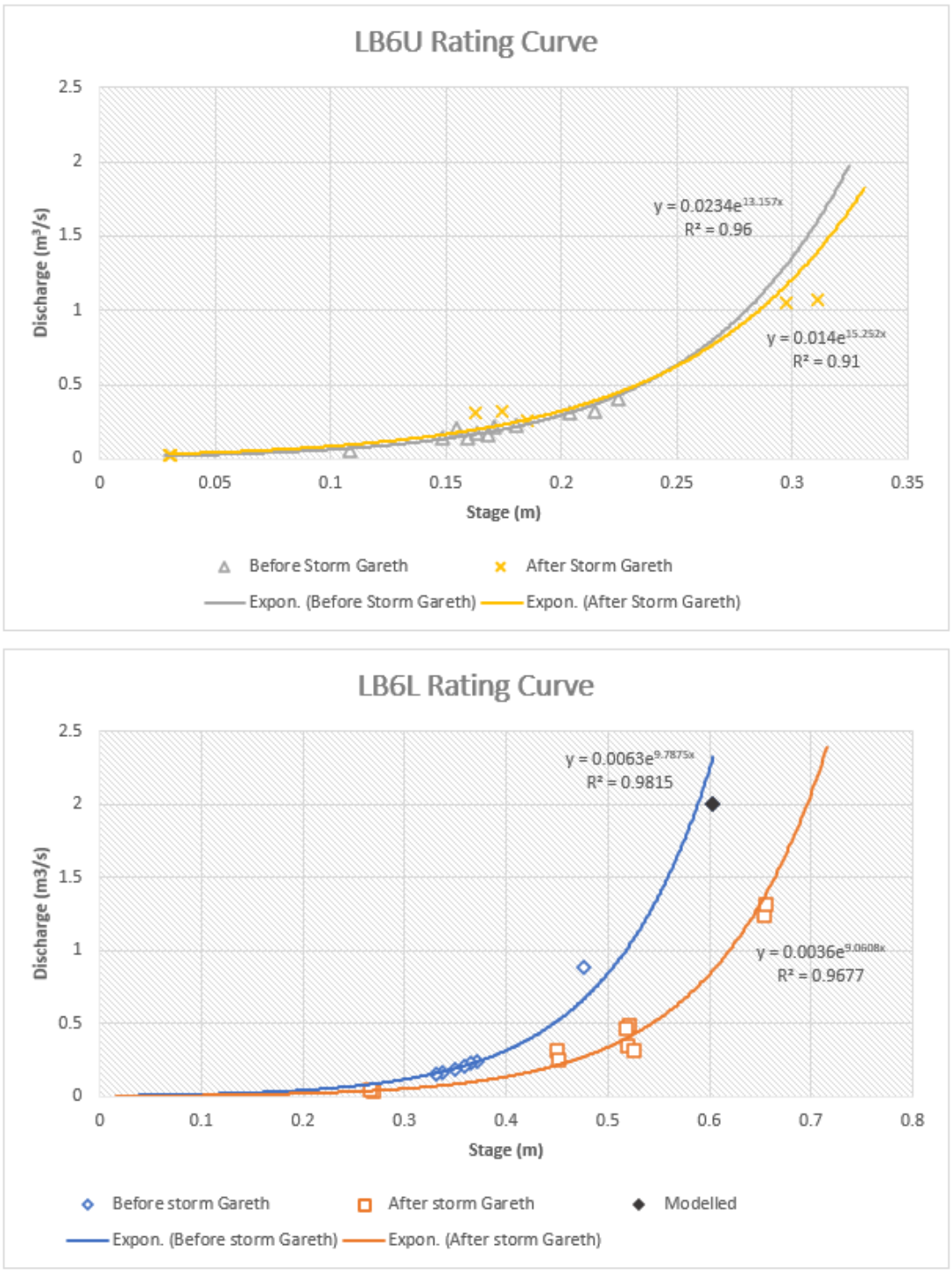


Figure B.5: Rating Curves showing spot gaugings and their exponential relationship before and after Storm Gareth

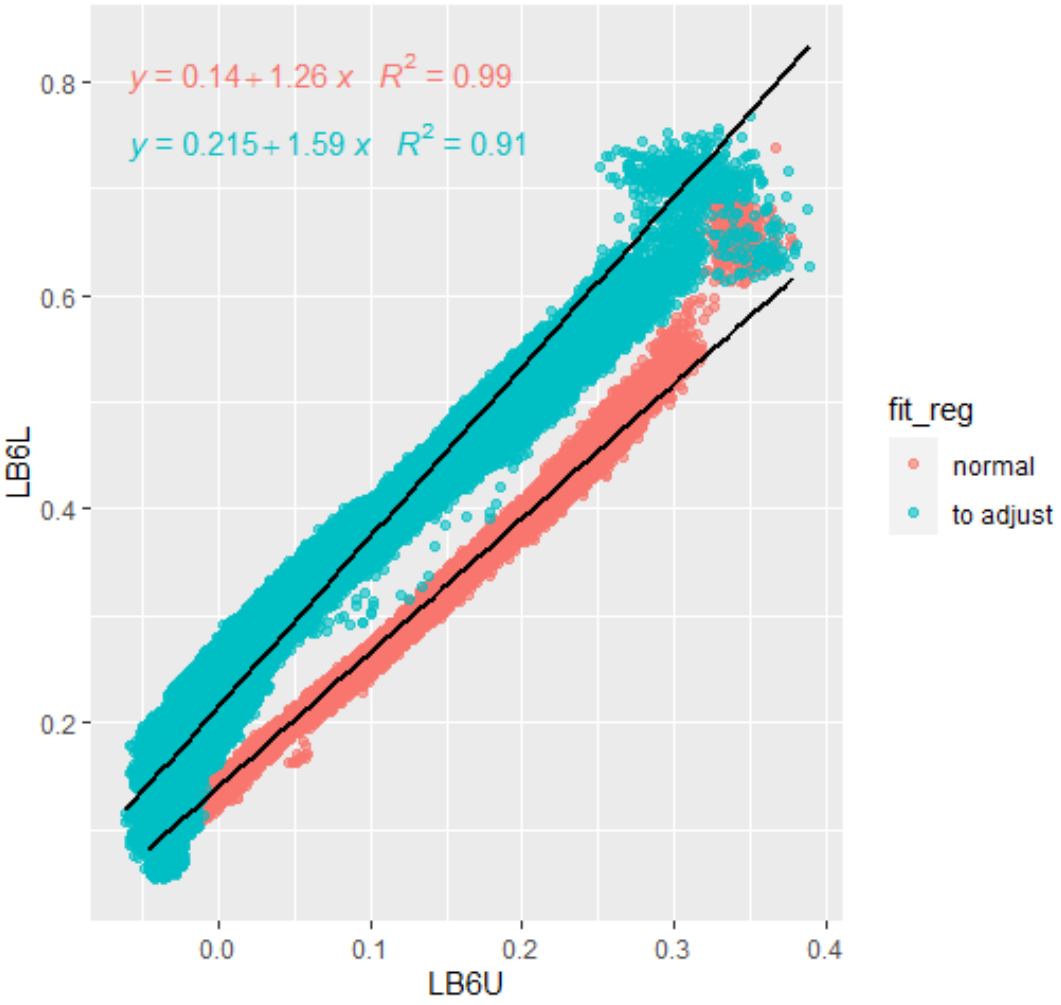


Figure B.6: Linear regression before and after Storm Gareth

data. However, the data is adjusted using the difference in the rating relationship, not the difference in upstream to downstream stage. This comparison merely serves as a sense check to ascertain that the magnitude of change is comparable so that the physical explanation can be deemed as viable.

Table B.1: Difference in lower gauge stage on Downs Gill before and after Storm Gareth

Lower Stage (m)	difference in rating relationship (m)	difference in upstream to downstream stage relationship (m)
0.65	0.11	0.17
0.6	0.10	0.15
0.5	0.09	0.13
0.4	0.09	0.11
0.3	0.08	0.09

By plotting the event peak in detail it becomes apparent that there is a drop in stage at the upstream logger during the event. The data drops by approximately 0.1 m on the upper logger, but then jumps back up, by about 0.05 m 20 minutes later so that the level goes from fluctuating around 0.36 m, to 0.30 m to 0.34 m whilst the lower logger goes from fluctuating around 0.63 m to 0.64 m. Whilst it is possible that this change persists the assumption is made here that the data drops and returns to normal during the event. This is a key assumption. If the datum was adjusted for this change it would make the upstream peak magnitude of this event and all subsequent events 0.1 m greater. However, as can be seen in Figure B.5 the rating relationship before and after Storm Gareth is reasonably constant, so it is unlikely that a permanent change occurred at the upstream logger.

Finally, the downstream logger data was adjusted from Storm Gareth (16 March 2019) to the end of the monitoring period. It is reasoned that the change in relationship between the upstream and downstream time series was brought about by the varying backwater effect of a footbridge which collapsed into the channel downstream of the downstream logger during Storm Gareth. Stage-discharge spot-gaugings taken before and after Storm Gareth, and extrapolated

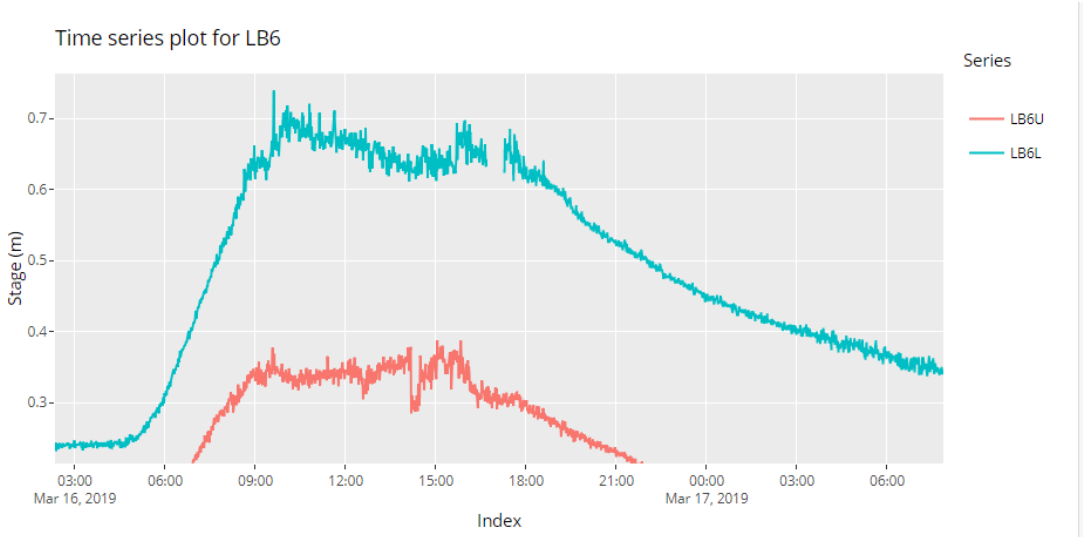


Figure B.7: Peak of Storm Gareth

using a HEC-RAS 1D hydraulic model are used to adjust the time series data after storm Gareth. As shown in Figure B.5 the relationship between stage (x) and discharge (Q) is described well by a logarithmic equation both before and after Storm Gareth ($R^2 \geq 0.97$). Equation (B.1) and (B.2) give the relationship between stage (x) and discharge (Q) before and after Storm Gareth, B and b are exponents of the logarithmic relationship of the data before Storm Gareth and A and a are exponents of the logarithmic relationship after Storm Gareth.

$$Q = Be^{bx_{before}} \tag{B.1}$$

$$Q = Ae^{ax_{after}} \tag{B.2}$$

As Q is equal in both equations it follows that the adjustment required to change x_{after} into x_{before} is given by equation B.4.

$$Be^{bx_{before}} = Ae^{ax_{after}} \tag{B.3}$$

$$x_before = \frac{Ae^{ax_{after}}}{B}\frac{1}{B} \tag{B.4}$$

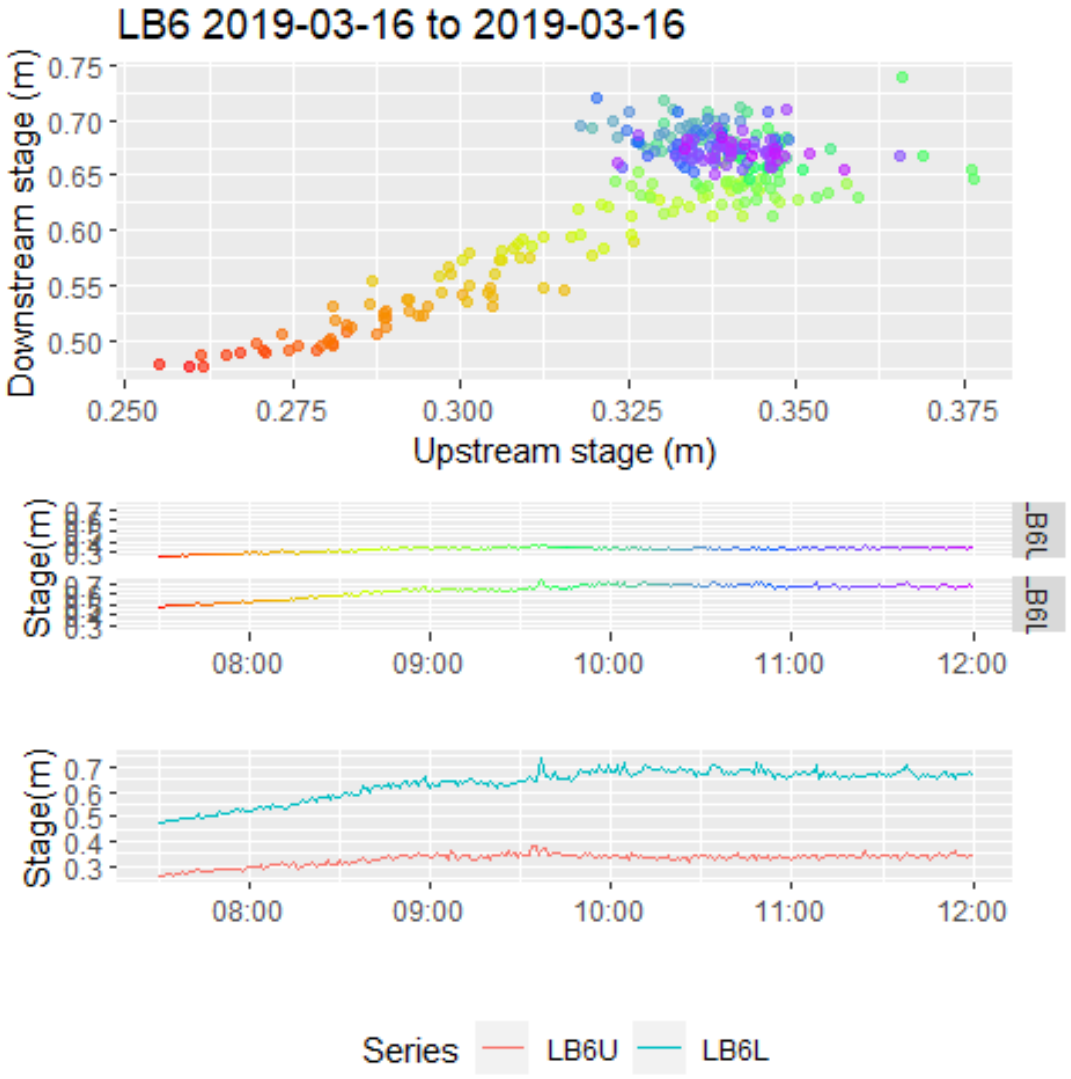


Figure B.8: Downs Gill Diagnostic plot of the rising limb and peak of Storm Gareth used to detect the point in time at which the change in the relationship occurred

It should be noted that as this adjustment is derived from gauged flows uncertainties are introduced, especially beyond the range of gauged flows where the relationship is extrapolated (stage >0.47 m before Storm Gareth, and Stage >0.65 m after Storm Gareth). The highest stage recorded during the monitoring period at the lower gauge on Downs Gill is 0.77 m in unadjusted terms. To improve the extrapolated relationship the rating curve at the lower gauge on Downs Gill was modelled in HEC-RAS 1D. A high flow point (2 m^3/s) was obtained from the model and added to the observed rating curve to increase confidence in the extrapolated relationship. This results in the adjustment shown in Figure B.9. Additionally, the erroneous data recorded in February 2020 was removed following a similar evidence gathering procedure. Figure B.9 shows the original and adjusted data for Downs Gill.

Table B.2: Data adjustments for Downs Gill

Gauge	from	to	adjustment
Upper	2019-02-03 00:00	2019-02-03 07:00	Data removed
Lower	2019-03-16 09:25	2020-02-17 (end of monitoring period)	Non-linear datum adjustment using equation (B.4) with B = 0.0005 b = 16.441 A = 0.0036 a = 9.0608

B.2.2 Fall Gill Example

Figure B.10 gives an example of the changing relationship between upstream and downstream stage during the baseline monitoring period. For the first 10 months the relationship appears relatively constant, but in January 2018 the downstream stage was 0.07 m lower for a given upstream stage than in the preceding months and this continues into February before changing back to its original level in March 2018. The change coincided with the dates on which the data on the downstream stage logger was downloaded so it was assumed that it was due to

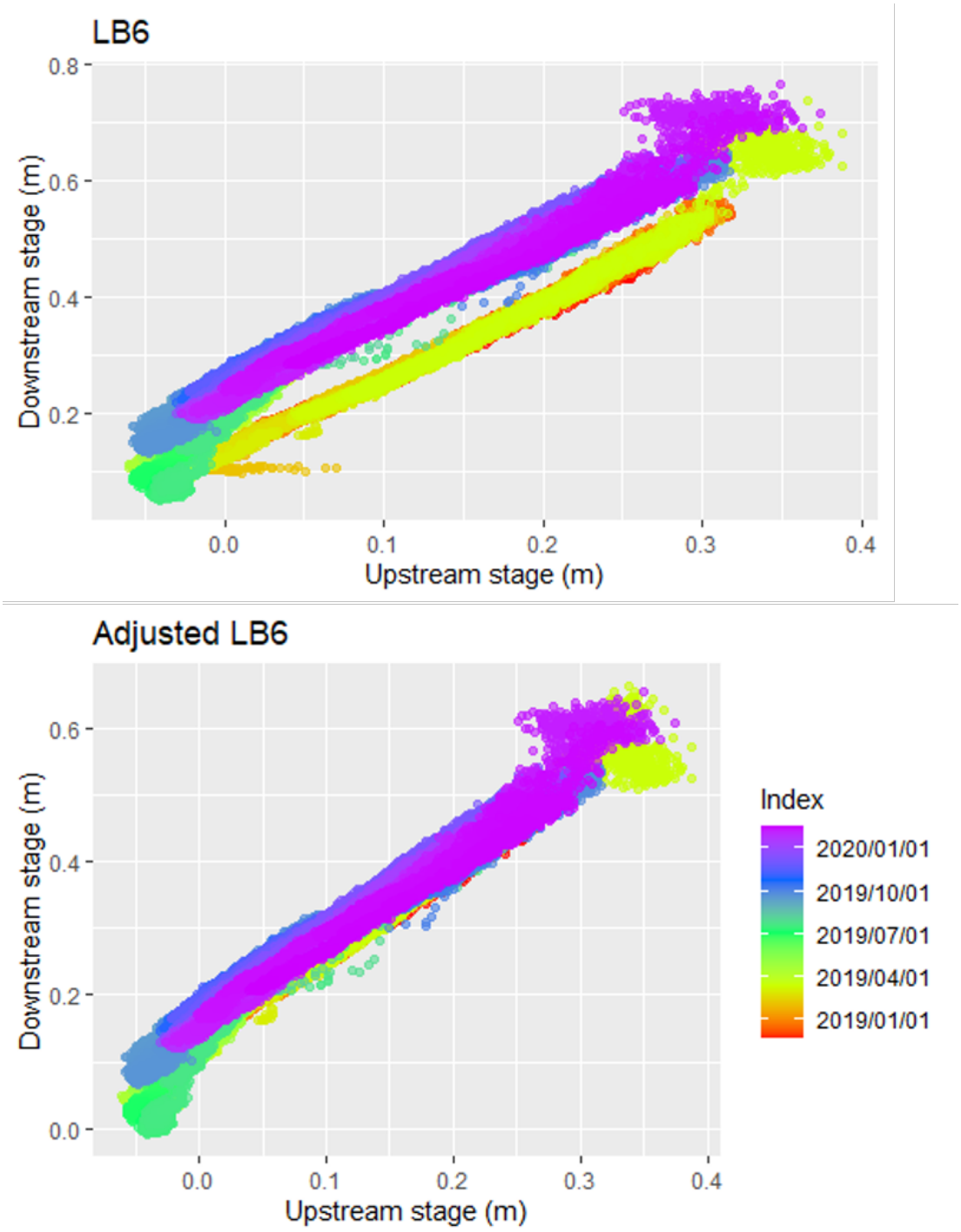


Figure B.9: Upstream stage plotted against downstream stage on Downs Gill for the post-intervention monitoring period before and after data adjustment

a change in the hanger length due to improper replacement of the logger and an appropriate correction was made to the data. In Figure B.11 a sudden drop in stage at the upstream logger on stream 3 during Storm Gareth (16th March 2019) in the post-intervention monitoring period can be seen. The start and end points of this change could be identified by plotting upstream against downstream stage for the event, and so the data was corrected. It was possible to correct the data in the above examples because it was clear at which logger the change had occurred and start and end points could readily be identified because the changes were sudden rather than gradual. For example, in Figure B.12 it can be seen that there is a small change (<0.05 m) in the relationship between upstream and downstream stage which occurred when the upstream stage was around 0.30 m, however, there was no step change in the time series plot for the event and so a correction to the data could not be justified.

B.2.3 Data Corrections

This process was repeated for all identified changes in the relationship between upstream and downstream stage for all streams and for both the baseline and post-intervention monitoring period. A summary of the data adjustments is given in Table B.3. Errors below 0.05 m were common but were too small to be detected using the methods described here because they often gradual rather than sudden which made their start and end points difficult to detect, particularly where they were a similar magnitude to the natural fluctuations of the water level.

Gauge	from	to	adjustment	period
LB5U	17/03/2018 17:00	18/03/2018 12:00	Data removed	before
LB5U	12/02/2018 00:00	15/02/2018 00:00	Data removed	before
LB5U	07/01/2018 18:00	09/01/2019 00:00	Data removed	before
LB5U	08/12/2017 08:00	11/12/2017 00:00	Data removed	before
LB5U	16/12/2017 04:00	16/12/2017 12:30	Data removed	before
LB5L	12/02/2018 00:00	15/02/2018 00:00	Data removed	before
LB5L	28/12/2017 20:00	29/12/2017 18:00	Data removed	before
LB7U	13/09/2017 02:28	13/09/2017 03:20	Data not to be used in magnitude analysis	before
LB8L	12/12/2017 16:00	02/02/2018 20:15	+0.066 m	before
LB8L	18/01/2018 00:00	18/01/2018 23:59	Data removed	before

LB8U	27/07/2018 16:00	27/09/2018 10:04	Data removed	before
LB8L	27/07/2018 16:00	27/09/2018 10:04	Data removed	before
LB5U	17/01/2019 19:00	18/01/2019 17:30	Data removed	after
LB5U	22/01/2019 22:00	24/01/2019 14:30	Data removed	after
LB5U	28/01/2019 07:00	01/02/2019 09:00	Data removed	after
LB5M_U	16/03/2019 07:46	17/02/2020 10:34	-0.06m	after
LB5M_L	16/03/2019 07:46	17/02/2020 10:34	-0.06m	after
LB5U	18/11/2019 17:30	19/11/2019 13:30	Data removed	after
LB5U	01/12/2019 18:30	02/12/2019 00:00	Data removed	after
LB5L	17/01/2019 22:00	18/01/2019 12:00	Data removed	after
LB5L	23/01/2019 03:00	23/01/2019 12:00	Data removed	after
LB5M_U	16/03/2019 07:46	17/02/2020 10:34	-0.06m	after
LB5M_L	16/03/2019 07:46	17/02/2020 10:34	-0.06m	after
LB5L	14/05/2019 00:00	05/07/2019 00:00	Data not to be used in timing analysis	after
LB6U	03/02/2019 00:00	03/02/2019 07:00	Data removed	after
LB6L	16/03/2019 09:25	17/02/2020 10:34	Non-linear datum adjustment	after
LB7U	09/02/2020 02:37	17/02/2020 10:34	Data not to be used in magnitude analysis	after
LB7U	16/03/2019 08:58	16/03/2019 17:35	Data not to be used in magnitude analysis	after
LB7L	09/02/2020 02:37	17/02/2020 10:34	Data not to be used in magnitude analysis	after
LB8U	12/09/2018 00:00	27/09/2018 09:00	+0.146 m	after
LB8L	12/09/2018 00:00	27/09/2018 09:00	+0.146 m	after
LB8U	16/03/2019 09:16	16/03/2019 18:55	+0.09 m	after
LB8L	09/02/2020 10:28	17/02/2020 10:34	+0.08 m	after
LB8L	09/02/2020 03:15	17/02/2020 10:34	+0.08 m	after

Table B.3: Summary of data corrections, U= upper gauge, M = middle gauge, L= lower gauge, LB5= Lock Gill, LB6 = Downs Gill, LB8= Fall Gill

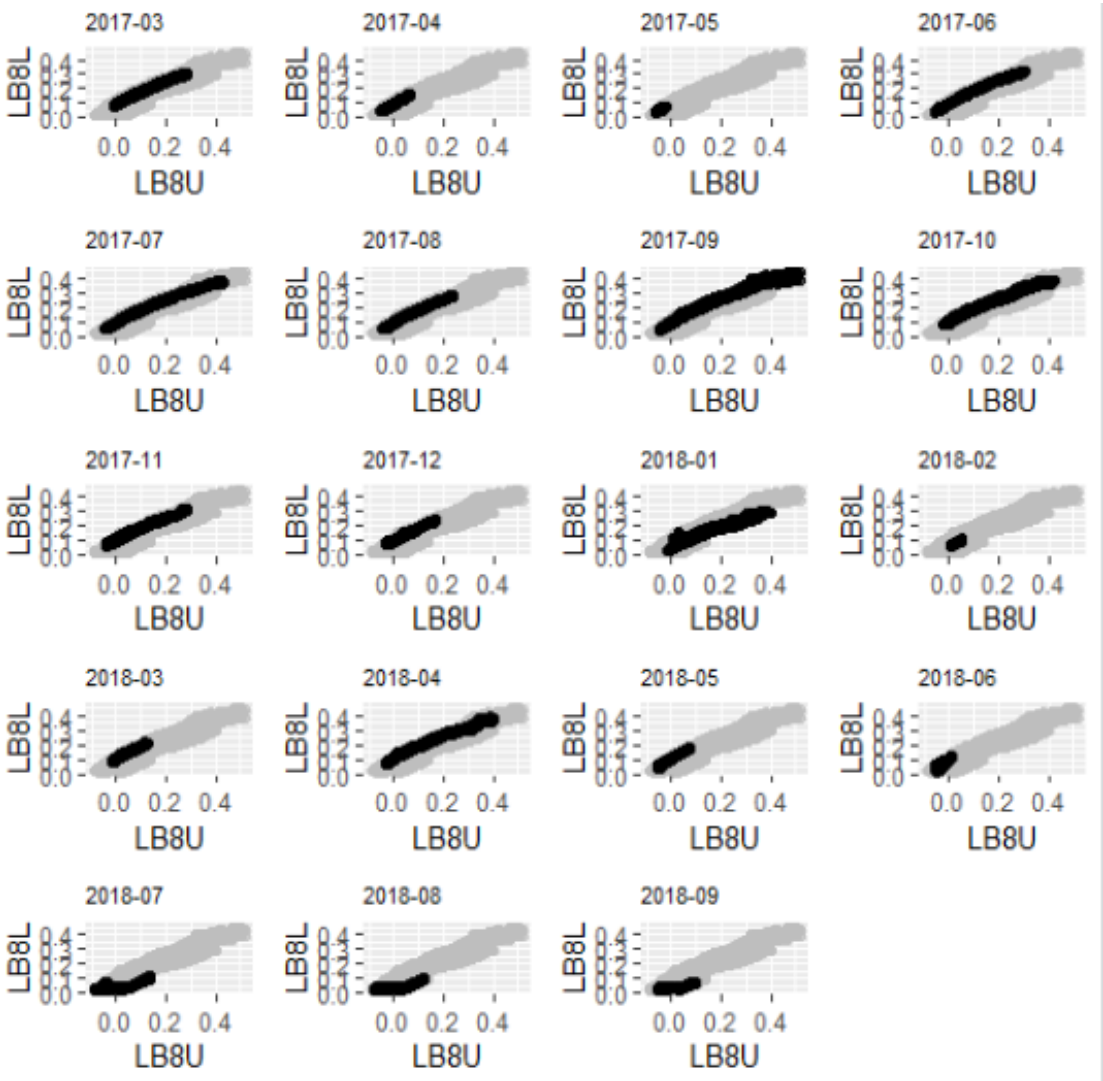


Figure B.10: Upstream stage (labelled LB8U) plotted against downstream stage (labeled LB8L) on Fall Gill for every month in the baseline monitoring period per month in black, overlying the data from the entire baseline monitoring period in grey

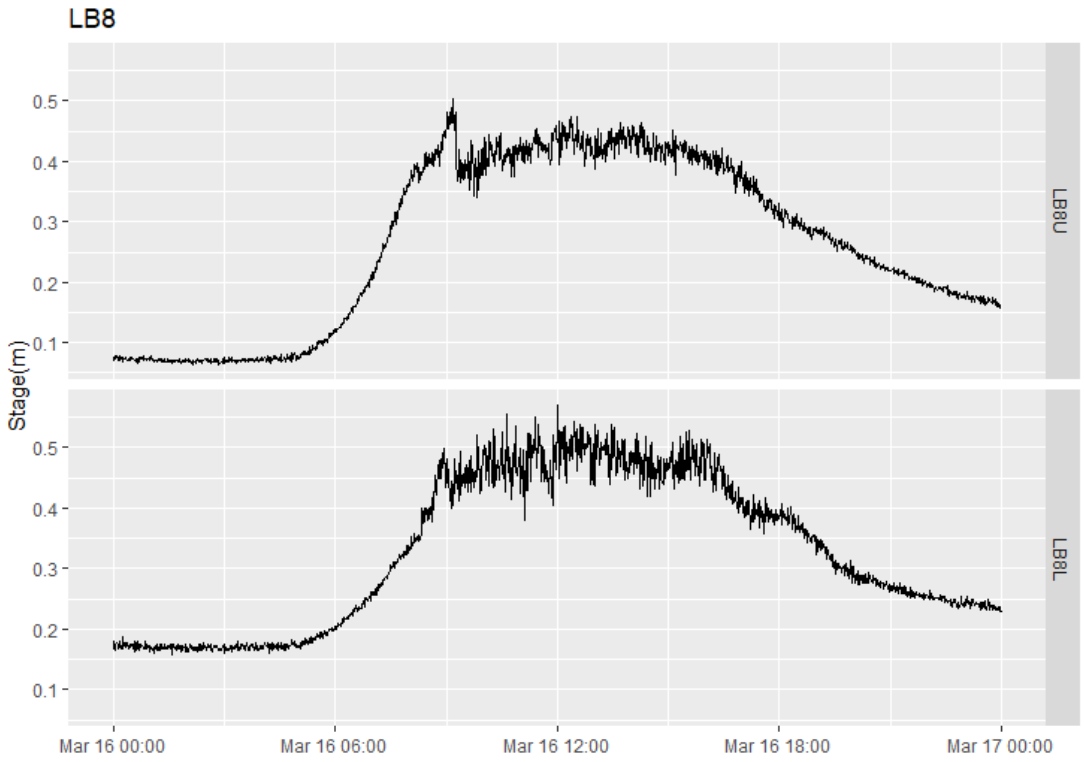


Figure B.11: Upstream (LB8U) and downstream (LB8L) stage measured on Fall Gill during Storm Gareth

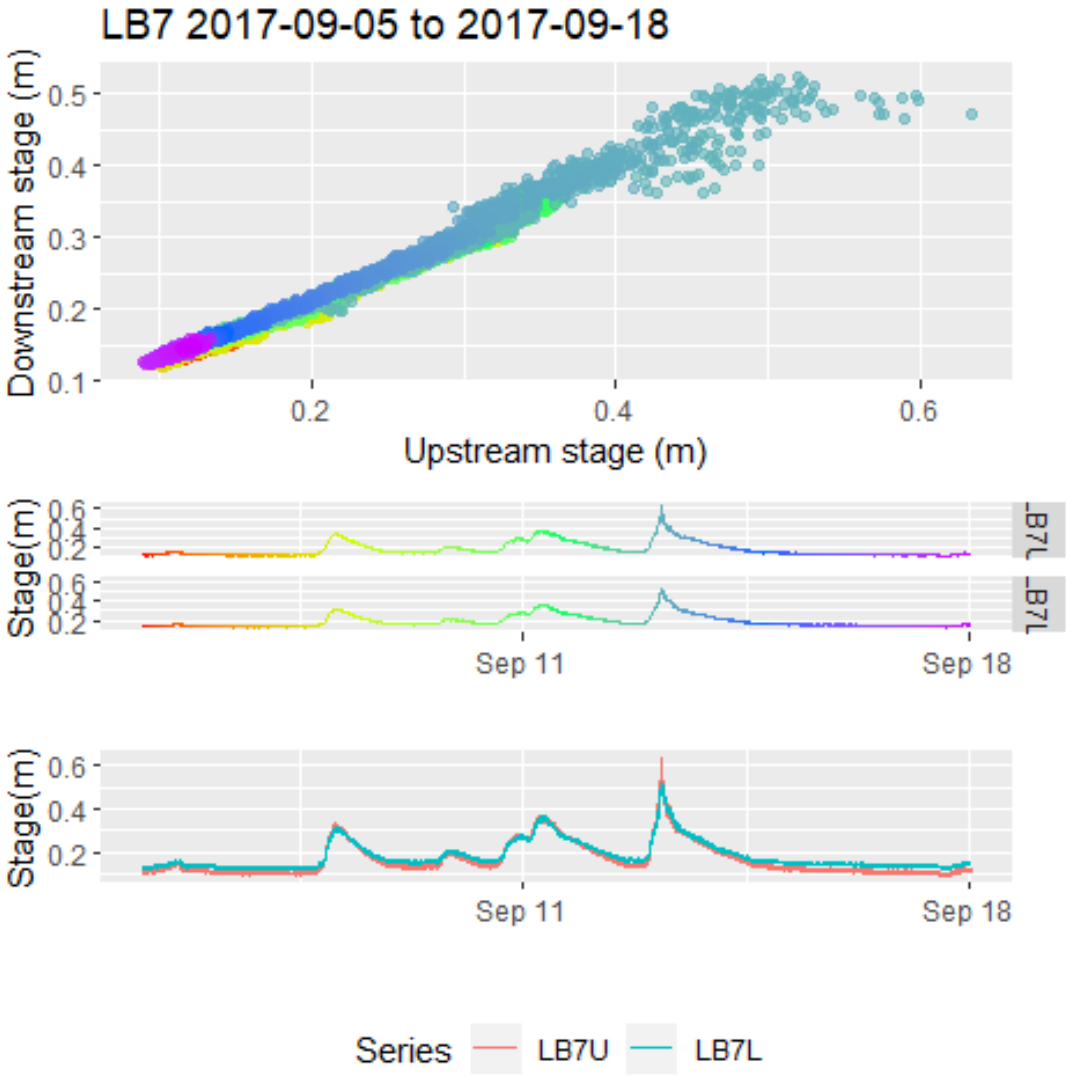


Figure B.12: Detailed relationship examination of upstream and downstream stage data collected on the control stream, West Gill, during a high flow event. The colouring of the time series plots provides the legend for the scatter plot.

B.3 Impact of uncertainty on quantifying change in peak stage magnitude and timing

Errors in the stage data identified during the QA process led to high levels of uncertainty in the peak magnitudes observed in the baseline and post-intervention data. Three types of common data quality issues were identified in the data. Namely, problems of shifting control (datum errors), changes in the rating relationship and missing data. Figure B.13 shows field observations of some of the sources of these errors. Figure B.13A illustrates the problem of monitoring high flows which cause afflux at the gauge and cause large fluctuations in the recorded stage levels. The fluctuations, or noise, increased with the magnitude of the stage up to a magnitude of ±0.1 m. However, the level fluctuated about a mean stage and so the data could be smoothed using the procedure described in Section 2.2. In Figure B.13B it can be seen that the gauges were prone to blockage by coarse woody material, which affected the readings of stage at the gauge by changing the local dynamics (e.g. afflux as seen in Figure B.13A). Additionally, silt was deposited in the bottom of stilling wells which could lead to the logger no longer being suspended at the correct level. Erosion of the river banks was observed at the upstream gauge on Fall Gill (Figure B.13D) but the impacts of this could not be discerned from the data. The collapse of a footbridge 5 m downstream of Downs Gill (Figure B.13C), on the other hand, caused a backwater effect which affected the logger during high flows. As described in the previous section spot discharge gaugings performed before and after the collapse of the footbridge were used to calculate the magnitude of the backwater effect at a range of flows. The data collected following the bridge collapse were adjusted accordingly.

The error which was most commonly identified in the stage data was sudden or gradual 'jumps' in the water level, or 'datum errors'. The distinction made here between datum errors and rating relationship errors is that datum errors were of constant magnitude across the range of flows, whereas for changes in the rating relationship the magnitude of the stage error varied with the stage magnitude. Datum errors occurred on all of the reaches, in both the pre and post intervention monitoring periods. The magnitude of the datum errors varied from <0.05 m to 0.15 m. Where start and end points could be readily identified by

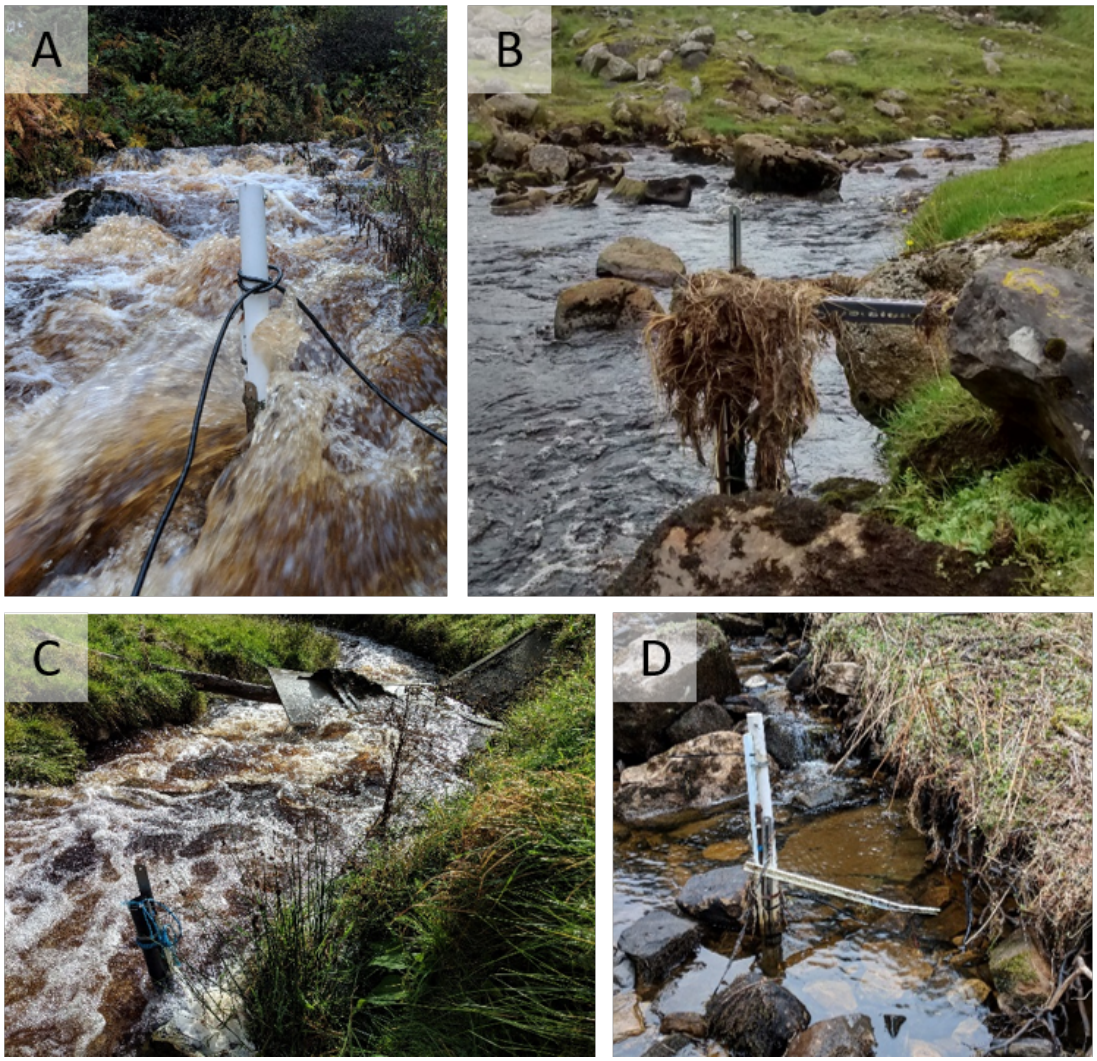


Figure B.13: Site observations of sources of data quality problems. A: Afflux at stilling well during high flows. B: Material deposited on stilling well. C: Gauging station affected by backwater effect due to collapsed footbridge. D: Bank collapse at gauging station

detailed examination of time series and the relationship between upstream and downstream stage these datum errors were corrected. However, this was often not possible as the changes occurred gradually rather than instantaneously making them difficult to detect, particularly if they were of a similar magnitude to natural fluctuations in the data. Consequently, datum errors <0.05 m could not be corrected and are common in the data. To take this uncertainty into account the stage measurements (h) are presented with an estimated error (E_{est}) of $h \pm 0.05m$. Errors in timing due to the internal clock on the upstream and downstream gauge manifested themselves as loops in plots of upstream against downstream stage. Timing errors were not common and were therefore not corrected.

B.3.1 Propagation of stage datum errors to stage-discharge rating relationship

The stage data uncertainty was reflected to some degree in the scatter and confidence intervals of the stage-discharge pair measurements, shown in Figure B.14. The uncertainty in the rating relationships varied between gauging stations due to the number and spread of the stage-discharge pairs. When only considering the uncertainty associated with the rating relationship at a discharge of 1 m^3/s the width of the 95% uncertainty bound varied from 0.05 m^3/s to 1.33 m^3/s, or 5% to 133% of the flow. Taking into account the uncertainty in the stage series measurements, $\pm E_{est}$, the error propagates to confidence interval widths >70% of the flow at 1m^3/s on all gauges.

B.3.2 Comparison of events before and after installation of leaky dams

Two single-peaked events with an upstream peak magnitude of approximately 1 m^3/s on Downs Gill were observed in each the baseline and post-intervention monitoring period (Figure B.15). The change in peak magnitude and timing from upstream to downstream for each event is given in Table B.4 and Table B.5 respectively. By comparing events of similar magnitude before and after the installation of leaky dams it can be seen that, when taking into account stage

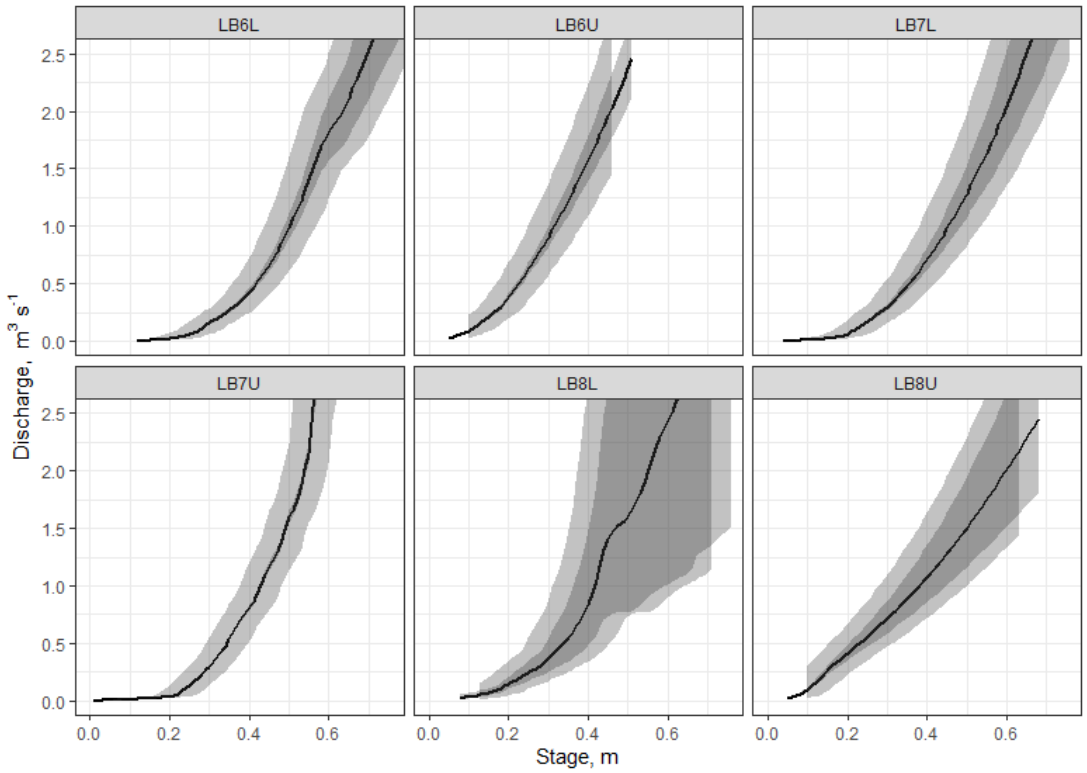


Figure B.14: Rating relationship uncertainty (dark grey is 95% CI based on rating relationship uncertainty alone, light grey is the uncertainty incorporating E_{est}, the error in the stage measurements

datum errors, the uncertainty in the discharge data is too great to discern any differences in the stream response. On stream 1 (Downs Gill) the data appears to show that installing leaky dams increased downstream event peak magnitude, but on the control stream the downstream peak magnitude response appears to have decreased in the post-intervention monitoring period despite no leaky dams being installed in the stream. On stream 3 (Fall Gill) downstream magnitude response appears decreased during storm Erik but increased during storm Atiyah. Finally, downstream discharge was smaller than upstream discharge during several of the events (e.g. for all four events on Stream 3 (Fall Gill) in Figure B.15), indicating further problems with the data or rating relationships.

Table B.4: Percentage change in peak magnitude from upstream to downstream for events in Figure B.15. The error when only the rating curve uncertainty is taken into account is given in brackets.

	Baseline ΔQ_p (%)		Post-intervention ΔQ_p (%)	
Event	Dec '17	Brian	Erik	Atiyah
Stream 1	7 ±95 (20)	10 ±95 (20)	-43 ±90 (23)	-37 ±94 (22)
Stream 2 (Control)	8 ±98 (15)	11 ±97 (15)	32 ±85 (17)	34 ±87 (16)
Stream 3	44 ±143 (63)	32 ±151 (67)	45 ±164 (76)	11 ±171 (74)

Table B.5: Change in event peak timing (mins) from upstream to downstream for events in Figure B.15

	Baseline ΔQ_p (%)		Post-intervention ΔQ_p (%)	
Event	Dec '17	Brian	Erik	Atiyah
Stream 1	-13	18	7	-2
Stream 2 (Control)	-4	-6	39	-9
Stream 3	-35	-17	14	28

By plotting the ΔQ_p against Q_p for all observed events (Figure B.17), along with the confidence intervals it can be seen that positive and negative values

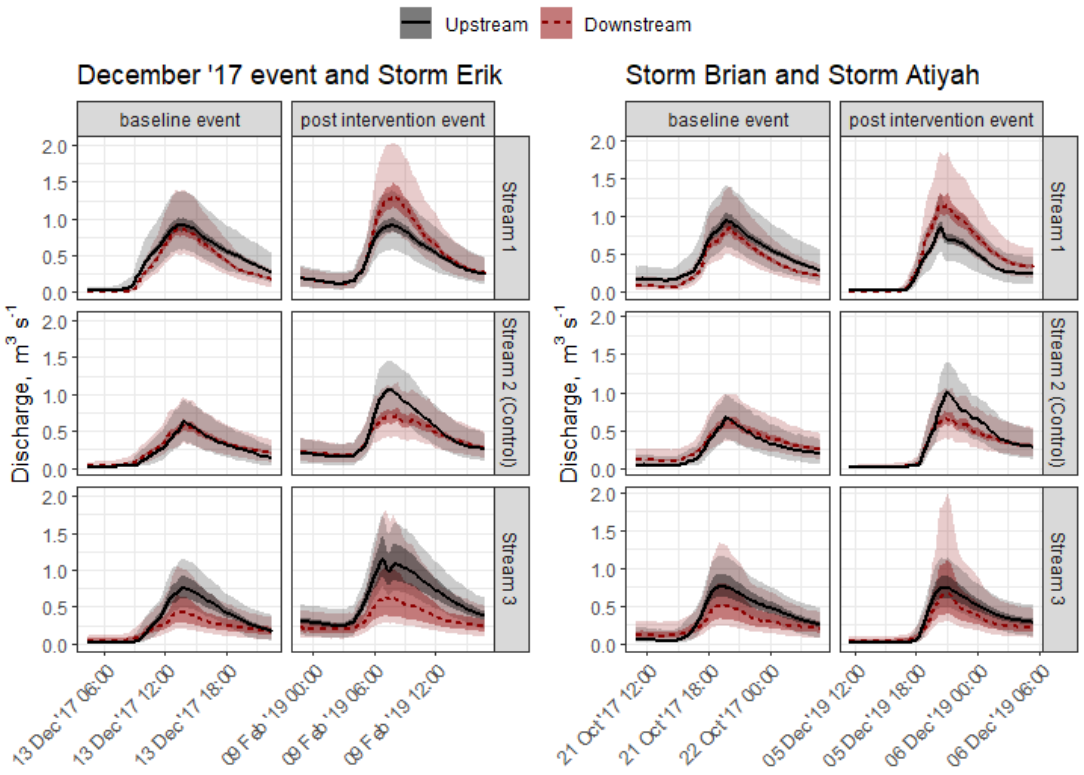


Figure B.15: High flow events during baseline and post-intervention period. In the first panel a high flow event on 13th of December 2017 and Storm Erik (February 2019) are compared. In the second panel Storm Brian (October 2017) and Storm Atiyah (December 2019) are compared. The shaded areas indicate the 95% confidence intervals, the darker shading indicates uncertainty due to discharge, and the lighter shading includes uncertainty due to stage datum errors

of ΔQ_p were observed on all three streams throughout the monitoring period, indicating that the data problems illustrated in Figure B.15 are common in the data. On the control stream and on stream 3 (Fall Gill) there appears to be a trend of increasing ΔQ_p with Q_p whilst on stream 1 (Downs Gill) the ΔQ_p appears to decrease with Q_p . However, in all cases the confidence intervals are too wide to assess whether there was a change in stream response in the post-intervention period. The results on Stream 1 (Downs Gill) illustrate the importance of considering datum change errors as the points lie outside of the rating relationship confidence interval.

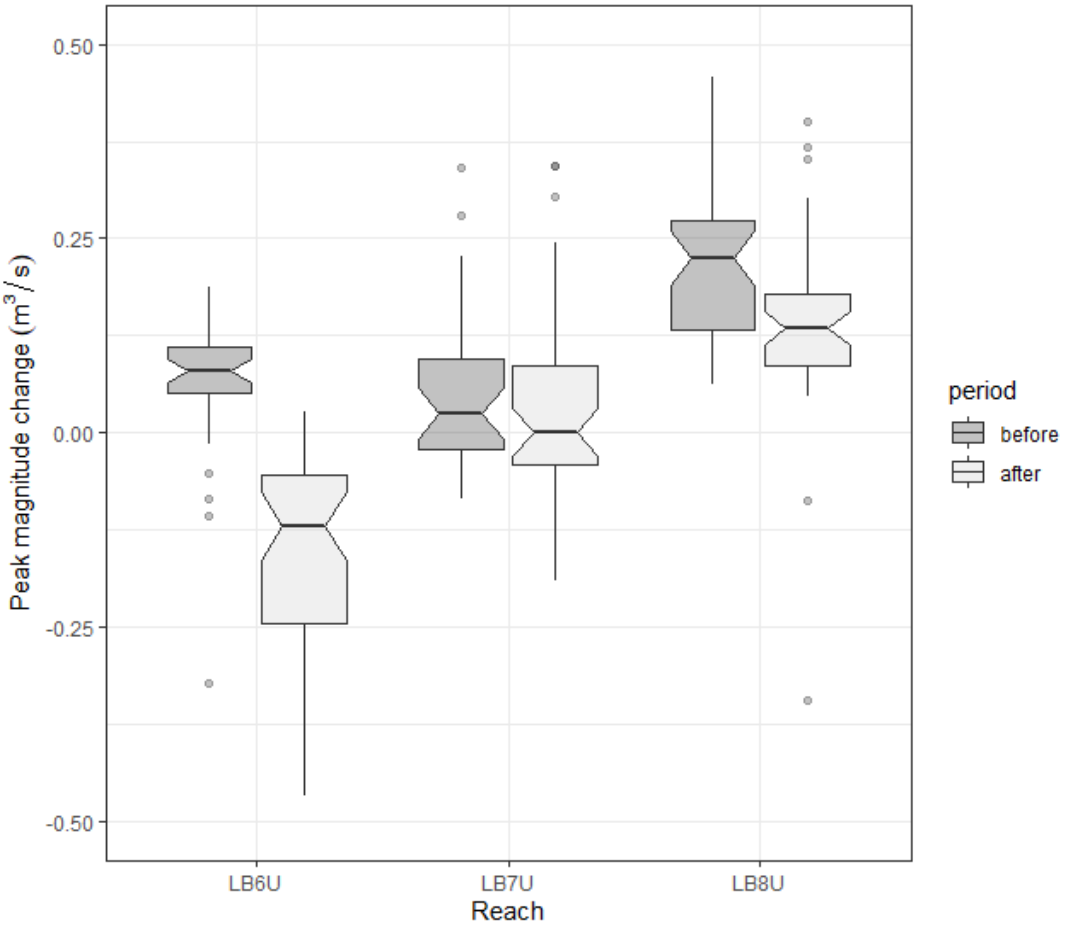


Figure B.16: Peak magnitude change (upstream magnitude – downstream magnitude) boxplots.

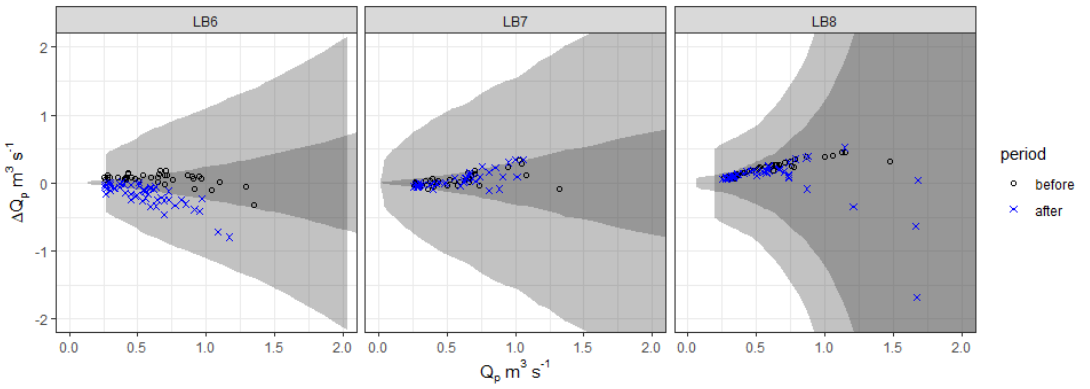


Figure B.17: Uncertainty in the measurement of event peak magnitude is greater than the change in peak magnitude from upstream to downstream of the study reaches. Points indicate the change in peak magnitude from upstream to downstream in the reach. The dark grey shaded area corresponds to the 95% uncertainty interval in the rating relationships whilst the light grey shaded area takes into account the uncertainty due to shifts up to ±0.05 m in the stage datum.

The median event peak timing increased by 5 minutes on every reach, including the control reach between the baseline and post-intervention monitoring period (Figure B.18). The change in timing of the centroid of the event, on the other hand, differed between streams. On the control reach the event centroid passed the downstream gauge 10 minutes earlier on average (median) during the post intervention monitoring period, whilst on the impact reaches the centroid travel time increased. On LB8 (Fall Gill) the centroid travel time increased by 4 minutes (not significant), but on LB6 (Downs Gill) the centroid travel time increased by 40 minutes, a significant difference ($p < 0.05$). However, in the baseline monitoring period the median peak centroid travel time was -33 minutes, suggesting that the event centroid passed the upstream gauge before the downstream gauge; indicative of further data problems. Although the median values of peak delay were similar (with the exception of stream 1 (Downs Gill)) in the baseline and post-intervention monitoring period the upper and lower quartiles were higher in the post intervention monitoring period for both impact streams using either method of defining peak delay (Figure B.18). The event peak timing of the first quartile of events increased by 16 and 10 minutes on the two impact

reaches, and the third quartile increased by 8 and 7 minutes whilst on the control reach the first and third quartile decreased by 5 and 4 minutes respectively. This means that whilst the median peak delay was similar in the baseline and post intervention monitoring periods, there were fewer short peak travel times and more long peak travel times following the installation of leaky dams on the impact reaches.

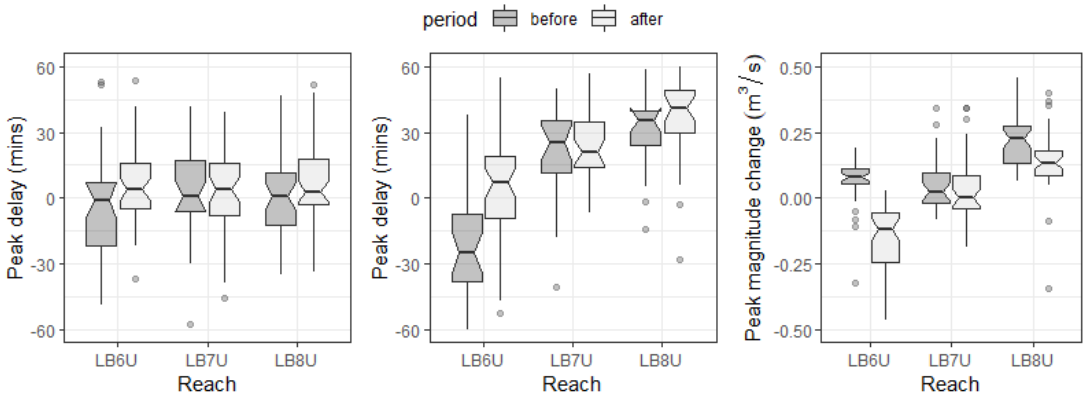


Figure B.18: Boxplots of (a) peak travel time; (b) event centroid travel time

Monotonic relationships between event peak delay and peak event magnitude were not evident in either the baseline or post intervention data (Figure B.19); the Spearman's rank correlation was between -0.45 and 0.28 for all groups of data. Inspection of the results is not complete without an appreciation of the uncertainties in the underlying stage data. During the quality assurance process inconsistencies in the stage datum and rating relationships were noted. Where these inconsistencies were large (>0.05 m) the data was corrected based on physical evidence and repeat discharge measurements. However, smaller datum errors (<0.05 m) could not be corrected as it was difficult to detect when they gradually started or ended. These smaller errors were found to be common in the data from all of the gauging stations and are thought to be due to blockage of the logger or river cross section with material carried by the stream during high flow events. The stage data (h) series were therefore given a range of $h \pm 0.05$ before being converted to discharge using the rating relationships with 95% confidence intervals developed in Appendix A.

B.3 Impact of uncertainty on quantifying change in peak stage magnitude and timing

The level of uncertainty associated with the stage time series and rating relationship, shaded grey and dark grey respectively in Figure B.17, precludes the assessment of leaky dam impacts on event peak magnitude. It can be seen in Figure B.17 that the difference between the upstream and downstream peak magnitude is smaller (generally $<\pm 0.5$ m^3/s) than the uncertainty in the data, which is >0.5m^3/s for all but the smallest events.

The majority of the points lie within the 95% uncertainty bounds of the rating relationship, indicating that even without taking the datum errors into account the results would be inconclusive. On Downs Gill the changes in peak magnitude recorded after the installation of leaky dams are larger than the uncertainty associated with the rating relationship. However, the change in peak magnitude is smaller than the uncertainty in the stage datum. Field evidence and inspection of the event hydrographs indicates that the apparent change in the relationship on Downs Gill (Figure B.17a) is likely to be due to a datum error following Storm Callum, which coincided with the installation of the dams.

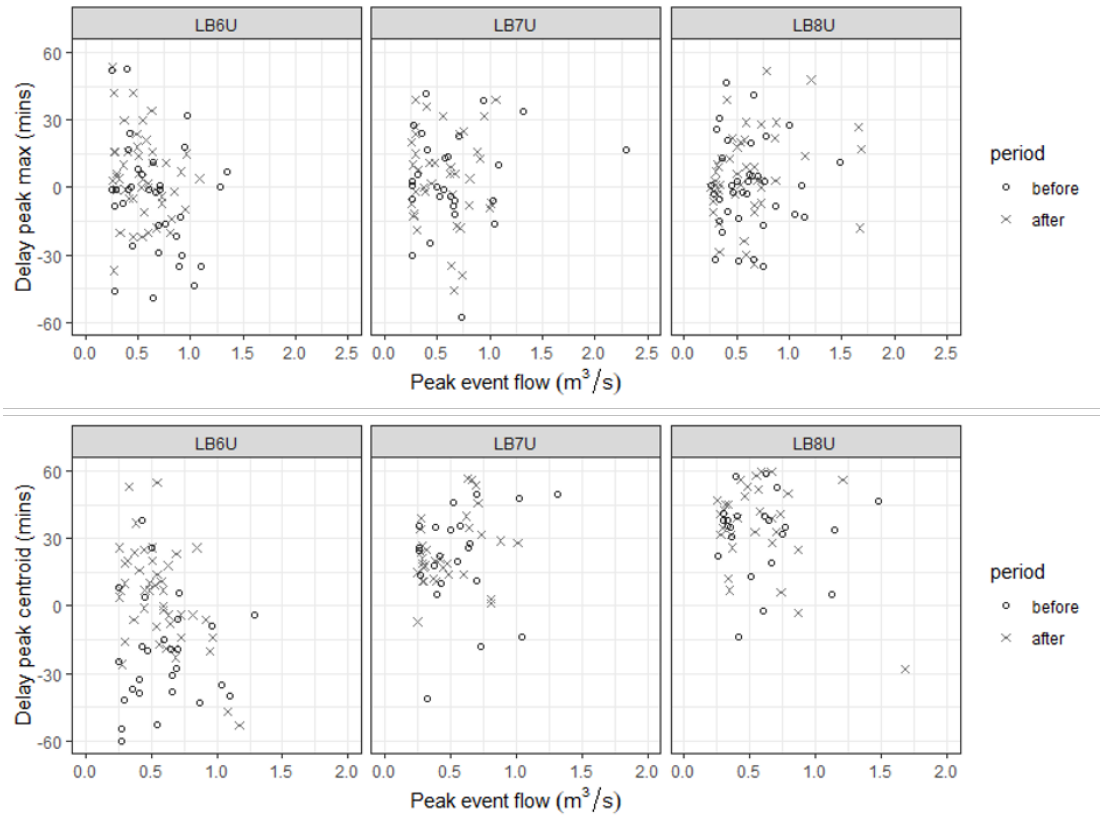


Figure B.19: Correlation of event characteristics a) peak travel time (b) event centroid travel time

Appendix C

Flood Frequency Estimation

Introduction

Flood frequency estimation was used to assess the extremeness of flood events (expressed as return period) based on long term records of annual maximum and peak over threshold data from Environment Agency operated hydrological monitoring stations.

The analysis followed the methodology presented in the standard hydrology text of Shaw et al. (2011). Generalised extreme value and Pareto distributions were fitted by maximum likelihood estimation to both the annual maximum and peak over threshold data to determine the best fit for estimating flood return period of the studied flood events. Goodness of fit plots, goodness of fit measures and standard errors estimates were used to evaluate the fit of the distributions to the data.

This appendix presents flood frequency estimation analysis for each Environment Agency operated hydrological monitoring station used in this study in the form of R markdown output. The R markdown displays R code alongside its outputs, including figures and narrative text.

Flood Frequency Estimation – Kilgram

Zora van Leeuwen

11/02/2020

C.1 Introduction

A GEV Distribution is estimated for the AMAX series downloaded from the National River Flow Archive https://nrfa.ceh.ac.uk/ using the MLE method.

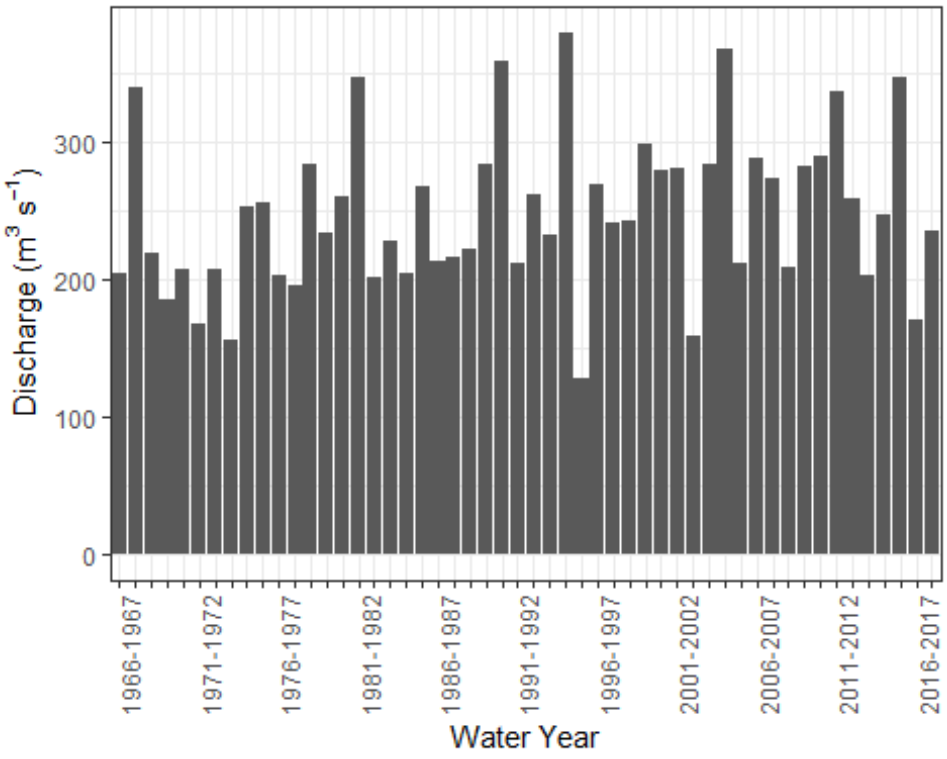
Goodness of fit plots are produced to evaluate the fit of the data along with goodness of fit measures such as the Negative log-likelihood, AIC and BIC. Standard Error Estimates are also provided. The fitted parameters are used to calculate the Annual Exceedance Probability (AEP) and Return Period (RP), the inverse of the AEP for a given level.

C.2 AMAX Data

Data file: `Input/Other catchments/27034 - Ure at Kilgram Bridge AMAX.csv`

```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
AMAX_data<-read.csv(params$filename)
head(AMAX_data)
```

```
##   Rank Water.Year       Date  Time Stage..m. Flow..m3.s.   Rating
## 1  N/A  1966-1967 18/08/1967 00:00     3.500     204.126  Extrap.
## 2    6  1967-1968 23/03/1968 00:00     5.170     340.544  Extrap.
## 3   32  1968-1969 31/10/1968 00:00     3.690     219.195  Extrap.
## 4   45  1969-1970 11/11/1969 00:00     3.260     185.287  Extrap.
## 5   38  1970-1971 12/02/1971 00:00     3.550     208.079 In Range
## 6   47  1971-1972 19/01/1972 09:45     3.038     168.069 In Range
##            Source Ref Comments
## 1           Chart  A1       NA
## 2           Chart  A1       NA
## 3           Chart  A1       NA
## 4           Chart  A1       NA
## 5           Chart  B3       NA
## 6 Digital Archive  B3       NA
```



Plot of the AMAX series which is used to fit the GEV distribution

The plot shows the AMAX series contained in the `Input/Other catchments/27034 - Ure at Kilgram Bridge AMAX.csv` dataset.

C.3 Fit GEV to AMAX series

The Package extRemes is used (Eric Gilleland, Richard W. Katz (2016). extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1-39. doi:10.18637/jss.v072.i08). The estimated shape parameter < 0, indicating a Weibull distribution.

```
library(extRemes)

fit<-fevd(AMAX_data$Flow..m3.s.,type="GEV")
fit
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Estimation Method used: MLE"
##
##  Negative Log-Likelihood Value:  282.7625
##
##  Estimated parameters:
##    location       scale       shape
## 225.8527167  52.0108047  -0.1676668
##
##  Standard Error Estimates:
```

```
##  location     scale     shape
## 8.0960274 5.7053478 0.1024526
##
##  Estimated parameter covariance matrix.
##            location      scale       shape
## location 65.5456589  8.8436137 -0.32775934
## scale     8.8436137 32.5509937 -0.27895267
## shape    -0.3277593 -0.2789527  0.01049654
##
##  AIC = 571.525
##  BIC = 577.3787
```

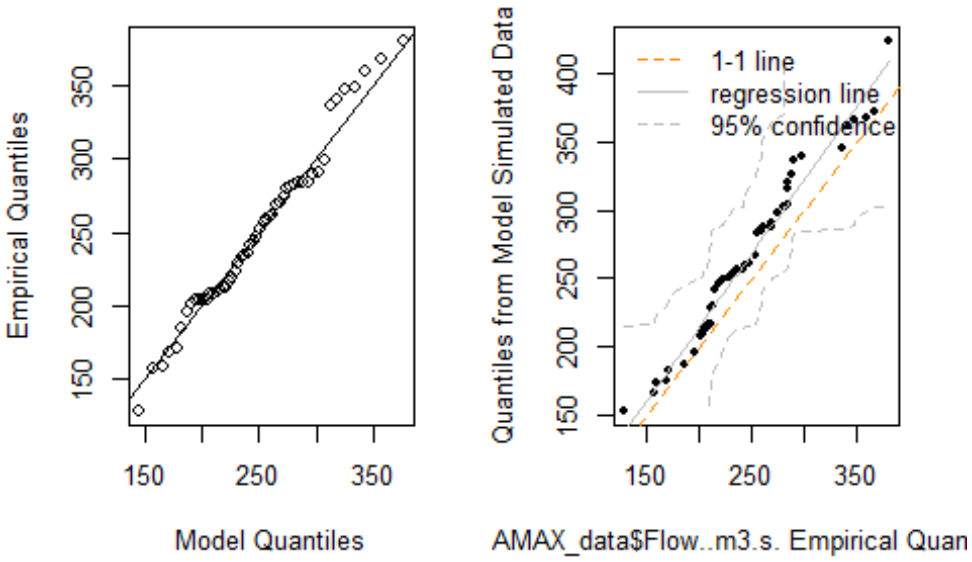
```
ci(fit,type="parameter")
```

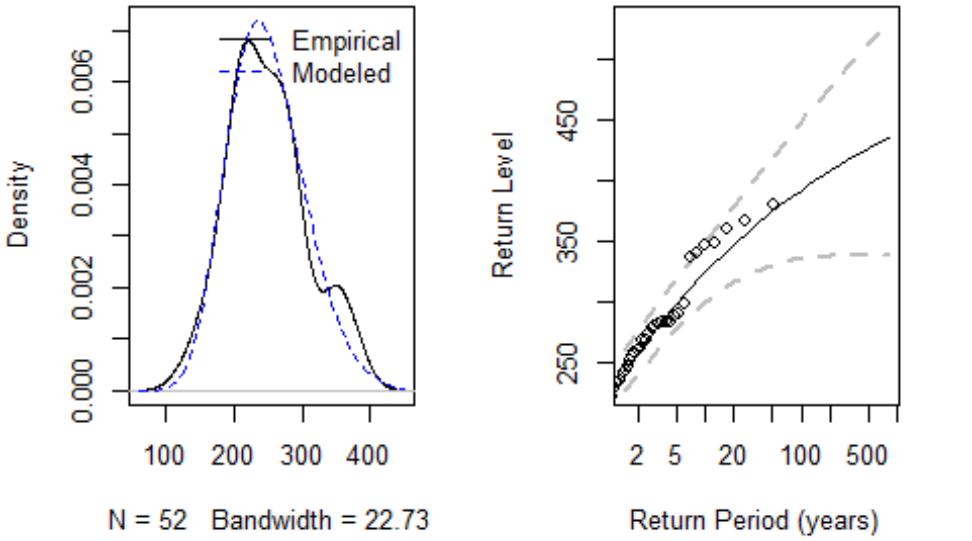
```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI    Estimate 95% upper CI
## location  209.9847947 225.8527167 241.72063873
## scale      40.8285284  52.0108047  63.19308092
## shape      -0.3684703  -0.1676668   0.03313672
```

Plotting the diagnostic plots shows a reasonable fit is achieved.

```
plot(fit)
```

```
fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
```

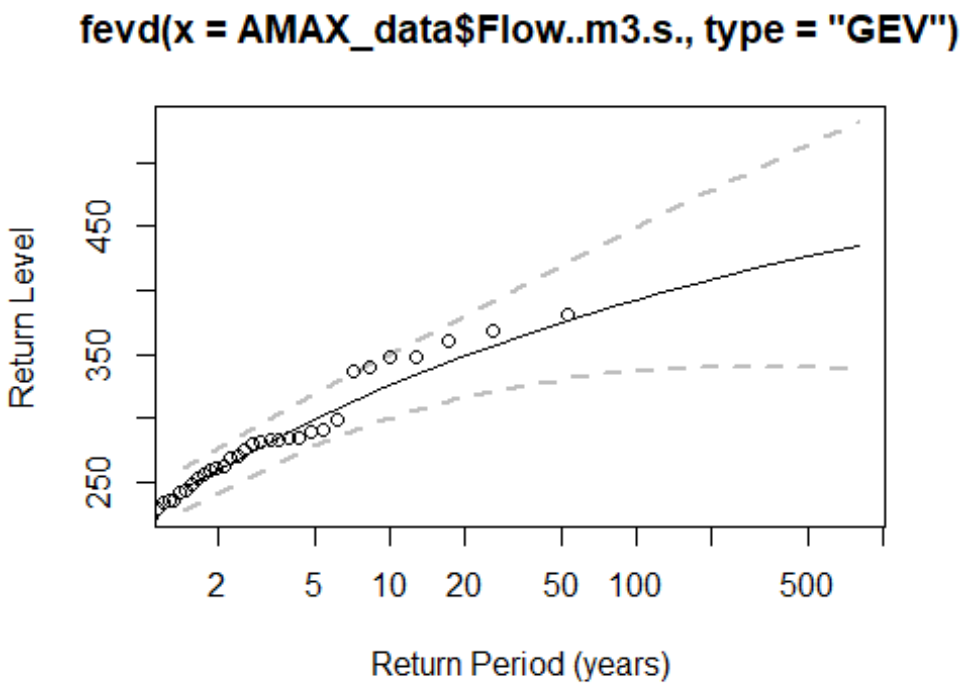




AMAX

Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit,type="rl")
```




```
return.level(fit,do.ci=TRUE)

## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##                         95% lower CI Estimate 95% upper CI
## 2-year return level         227.5725 244.3414     261.1104
## 20-year return level        316.6359 347.5322     378.4284
## 100-year return level       336.6664 392.6134     448.5604
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`157, 303`.

```
level_from_GEV_fun<-function(fit,level){
  #Get parameters
  u<-fit$results$`par`[["location"]]
  d<-fit$results$`par`[["scale"]]
  e<-fit$results$`par`[["shape"]]
  z<-level

  #calculate from Gumbel distribution equation
  a<-e*((z-u)/d)#intermediate
  b<-(1+a)^(-1/e)#intermediate
  P<-exp(-b)# Probability

  return(P) #Return the probability
```

```
}

#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit)
P

## [1] 0.03668397 0.83386077

AEP<-1-P
AEP

## [1] 0.9633160 0.1661392

RP<-1/AEP #return period (years)
RP

## [1] 1.038081 6.019048
```

For small return period events (<5 years) the distribution obtained using the AMAX and POT series are different (see http://evidence.environment-agency.gov.uk/FCERM/en/FluvialDesignGuide/Chapter2.aspx?pagenum=4). Hence, Fit a Generalised Pareto Distribution to the POT data instead.

C.4 Fit distribution to POT data

##POT Data Data file: `Input/Other catchments/27034 - Ure at Kilgram Bridge POT.csv`

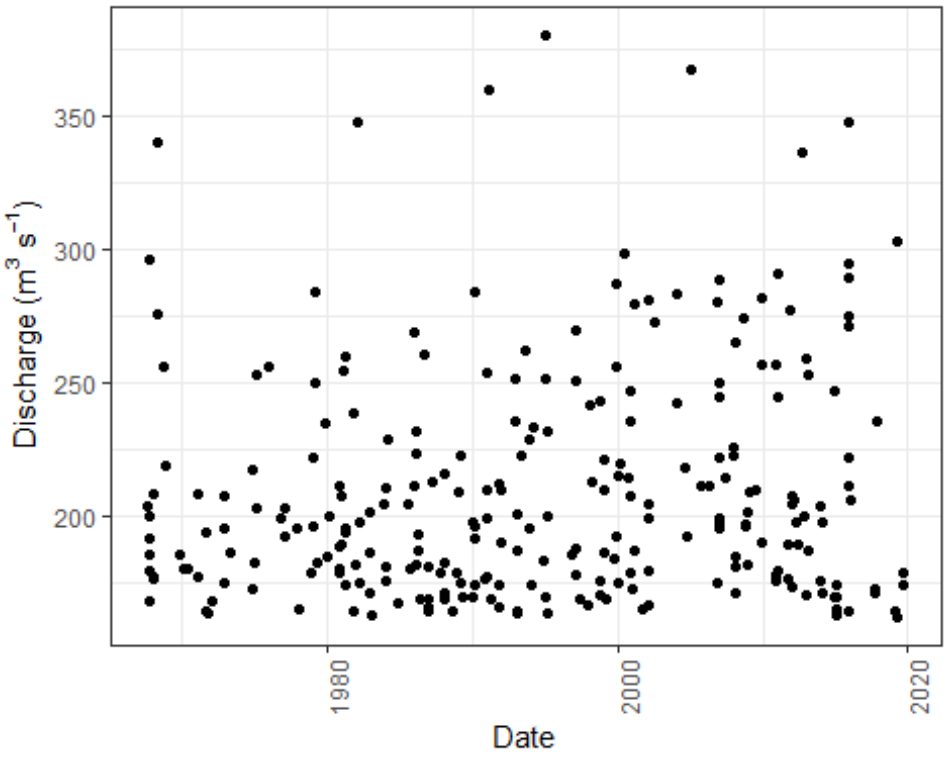
```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
POT_data<-read.csv(params$filename_POT)
POT_data$Date<-as.Date(POT_data$Date, format="%d/%m/%Y")

head(POT_data)

##   Rank       Date  Time Stage..m. Flow..m3.s.  Rating Source Ref Comments
## 1  111 1967-08-18 00:00      3.50     204.126 Extrap.  Chart  A1       NA
## 2  163 1967-10-01 00:00      3.26     185.287 Extrap.  Chart  A1       NA
## 3  182 1967-10-03 00:00      3.19     179.836 Extrap.  Chart  A1       NA
## 4  236 1967-10-07 00:00      3.04     168.223 Extrap.  Chart  A1       NA
## 5  147 1967-10-09 00:00      3.34     191.542 Extrap.  Chart  A1       NA
## 6  118 1967-10-14 00:00      3.45     200.183 Extrap.  Chart  A1       NA

summary(POT_data$Flow..m3.s.)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   162.5   177.5   197.4   209.6   228.9   380.4
```

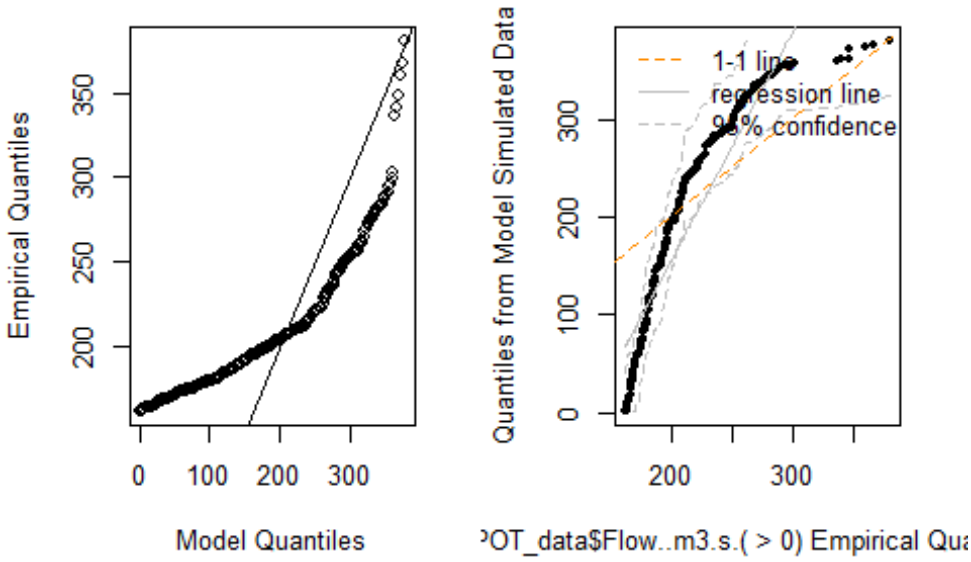


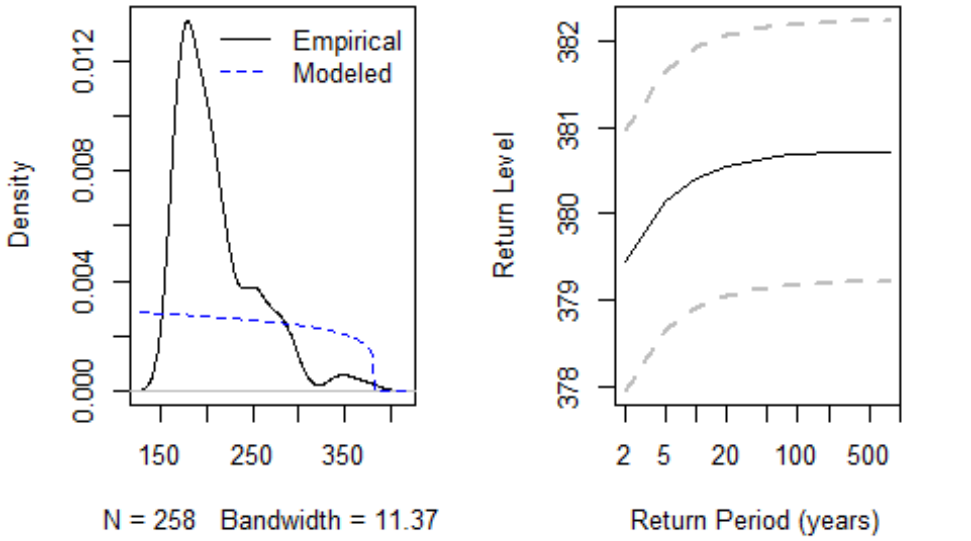
First, try to fit a Generalised-Pareto Distribution

```
#estimate number of years in POT series
library(zoo)
years<-as.yearmon(max(POT_data$Date))-as.yearmon(min(POT_data$Date))

#Terrible fit with the GP
fit_POT_GP<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GP",span=years)
plot(fit_POT_GP)
```

vd(x = POT_data$Flow..m3.s., threshold = 0, type = "GP", span = year





GP Distribution does not fit the data, hence the GEV distribution is used again.

```
#Try GEV instead
fit_POT_GEV<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GEV",span=years)
fit_POT_GEV
```

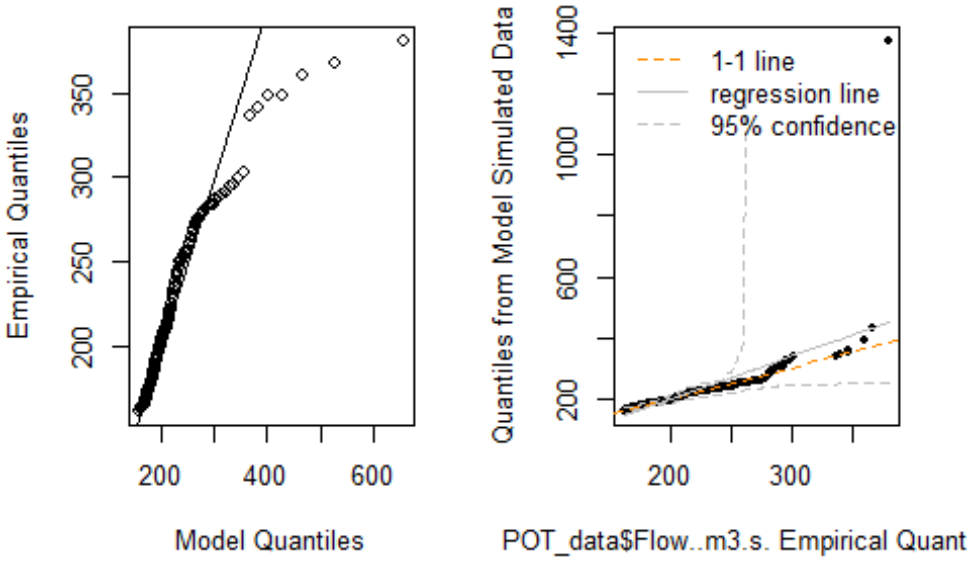
```
##
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  1268.453
##
##
##  Estimated parameters:
##    location       scale       shape
## 185.8352492  22.4679962   0.4056559
##
##  Standard Error Estimates:
##  location     scale     shape
## 1.7240597 1.5540468 0.0790239
##
##  Estimated parameter covariance matrix.
##             location       scale        shape
## location  2.97238181  1.97830455 -0.058661191
## scale     1.97830455  2.41506147 -0.023340759
## shape    -0.05866119 -0.02334076  0.006244777
##
##  AIC = 2542.906
##
##  BIC = 2553.565
```

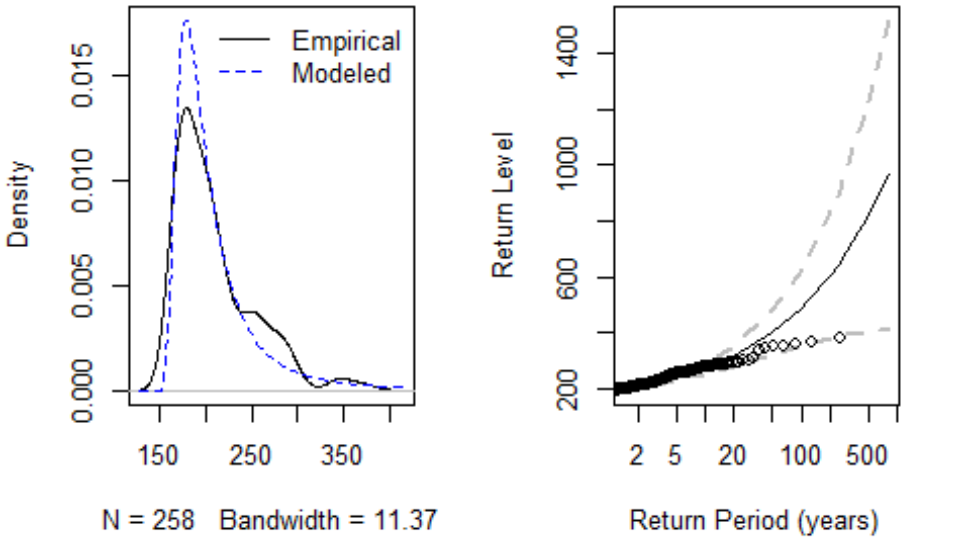
```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI    Estimate 95% upper CI
## location  209.9847947 225.8527167 241.72063873
## scale      40.8285284  52.0108047  63.19308092
## shape      -0.3684703  -0.1676668   0.03313672
```

```
plot(fit_POT_GEV)
```

d(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = yea

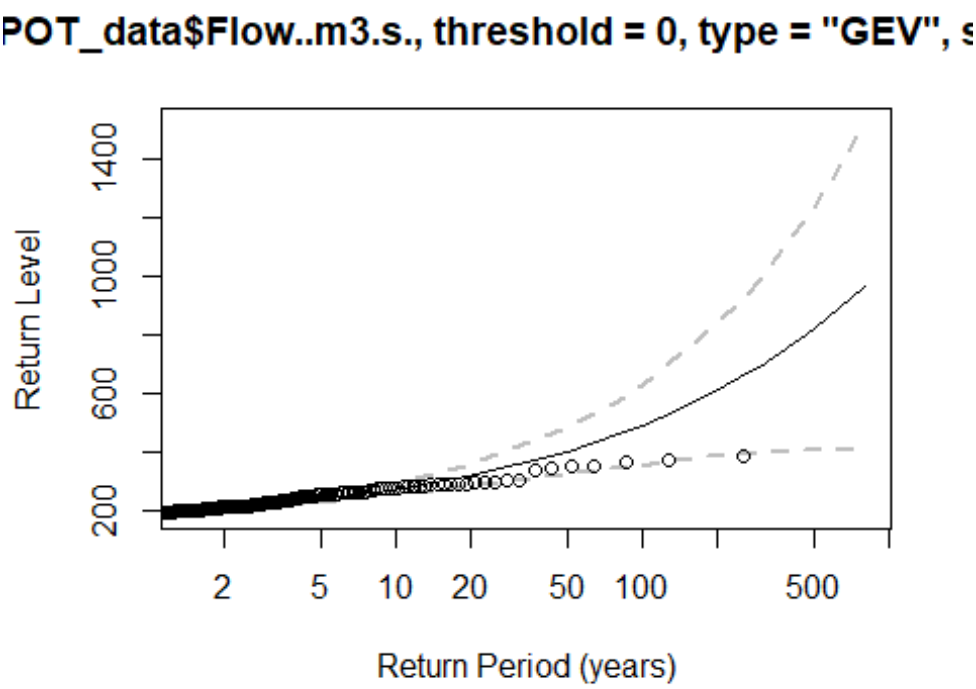




The GEV Distribution produces a reasonable fit at lower return periods.The estimated shape parameter is >0, indicating a Frechet distribution is likely.

##POT Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit_POT_GEV,type="rl")
```




```
return.level(fit_POT_GEV,do.ci=TRUE)
```

```
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Normal Approx."
##                       95% lower CI Estimate 95% upper CI
## 2-year return level       190.4628 194.7137     198.9646
## 20-year return level      278.7249 315.2383     351.7516
## 100-year return level     351.3596 488.4078     625.4560
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`157, 303`.

```
#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit_POT_GEV)
P
```

```
## [1] 0.002185829 0.941074948
```

```
AEP<-1-P
AEP
```

```
## [1] 0.99781417 0.05892505
```

```
RP<-1/AEP #return period (years)
RP
```

```
## [1]  1.002191 16.970711
```

Flood Frequency Estimation - Leck

Zora van Leeuwen

11/02/2020

C.5 Introduction

A GEV Distribution is estimated for the AMAX series downloaded from the National River Flow Archive https://nrfa.ceh.ac.uk/ using the MLE method.

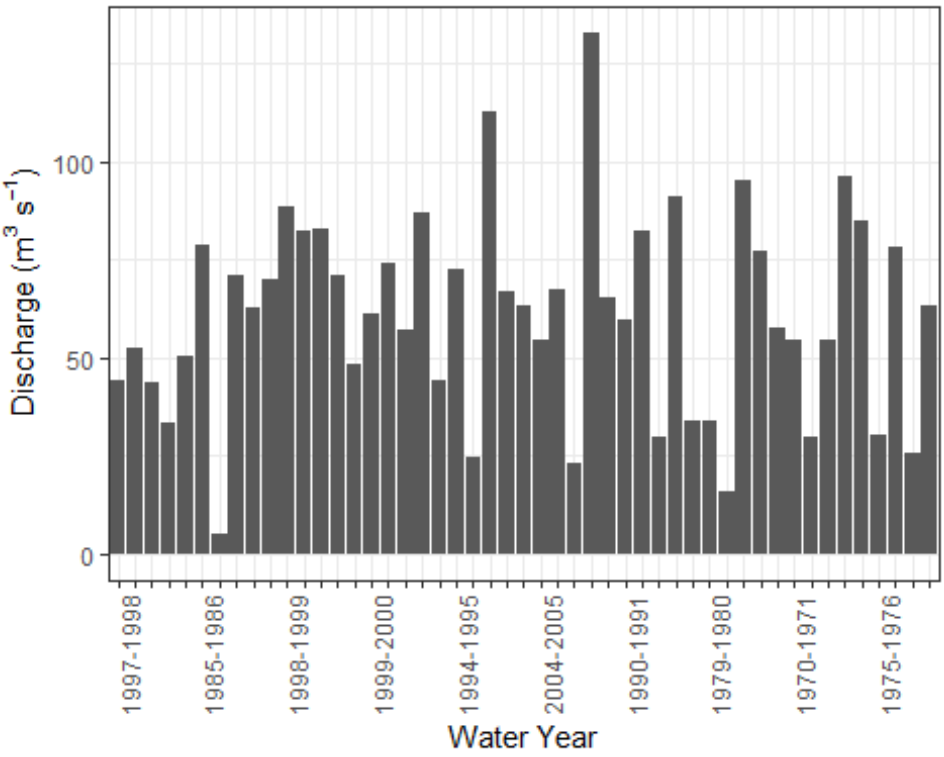
Goodness of fit plots are produced to evaluate the fit of the data along with goodness of fit measures such as the Negative log-likelihood, AIC and BIC. Standard Error Estimates are also provided. The fitted parameters are used to calculate the Annual Exceedance Probability (AEP) and Return Period (RP), the inverse of the AEP for a given level.

C.6 AMAX Data

Data file: `Input/Other catchments/Leck/33037 - Bedford Ouse at Newport Pagnell AMAX.csv`

```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
AMAX_data<-read.csv(params$filename)
head(AMAX_data)
```

```
##   Rank Water.Year       Date  Time Stage..m. Flow..m3.s. Rating          S
ource
## 1    1  1997-1998 10/04/1998 10:30       N/A       133.0    N/A Digital Ar
chive
## 2   10  2000-2001 13/02/2001 17:45       N/A        82.5    N/A Digital Ar
chive
## 3   11  1980-1981 28/04/1981 07:30       N/A        82.3    N/A Digital Ar
chive
## 4   12  2015-2016 10/03/2016 12:15       N/A        78.4    N/A Digital Ar
chive
## 5   13  2007-2008 17/01/2008 01:00       N/A        77.5    N/A Digital Ar
chive
## 6   14  1985-1986 11/01/1986 14:45       N/A        74.0    N/A Digital Ar
chive
##   Ref Comments
## 1 N/A
## 2 N/A
## 3 N/A
## 4 N/A
## 5 N/A
## 6 N/A
```



Plot of the AMAX series which is used to fit the GEV distribution

The plot shows the AMAX series contained in the `Input/Other catchments/Leck/33037 - Bedford Ouse at Newport Pagnell AMAX.csv` dataset.

C.7 Fit GEV to AMAX series

The Package extRemes is used (Eric Gilleland, Richard W. Katz (2016). extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1-39. doi:10.18637/jss.v072.i08). The estimated shape parameter < 0, indicating a Weibull distribution.

```
library(extRemes)

fit<-fevd(AMAX_data$Flow..m3.s.,type="GEV")
fit

##
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Estimation Method used: MLE"
##
##  Negative Log-Likelihood Value:  228.1423
##
##  Estimated parameters:
##   location      scale      shape
```

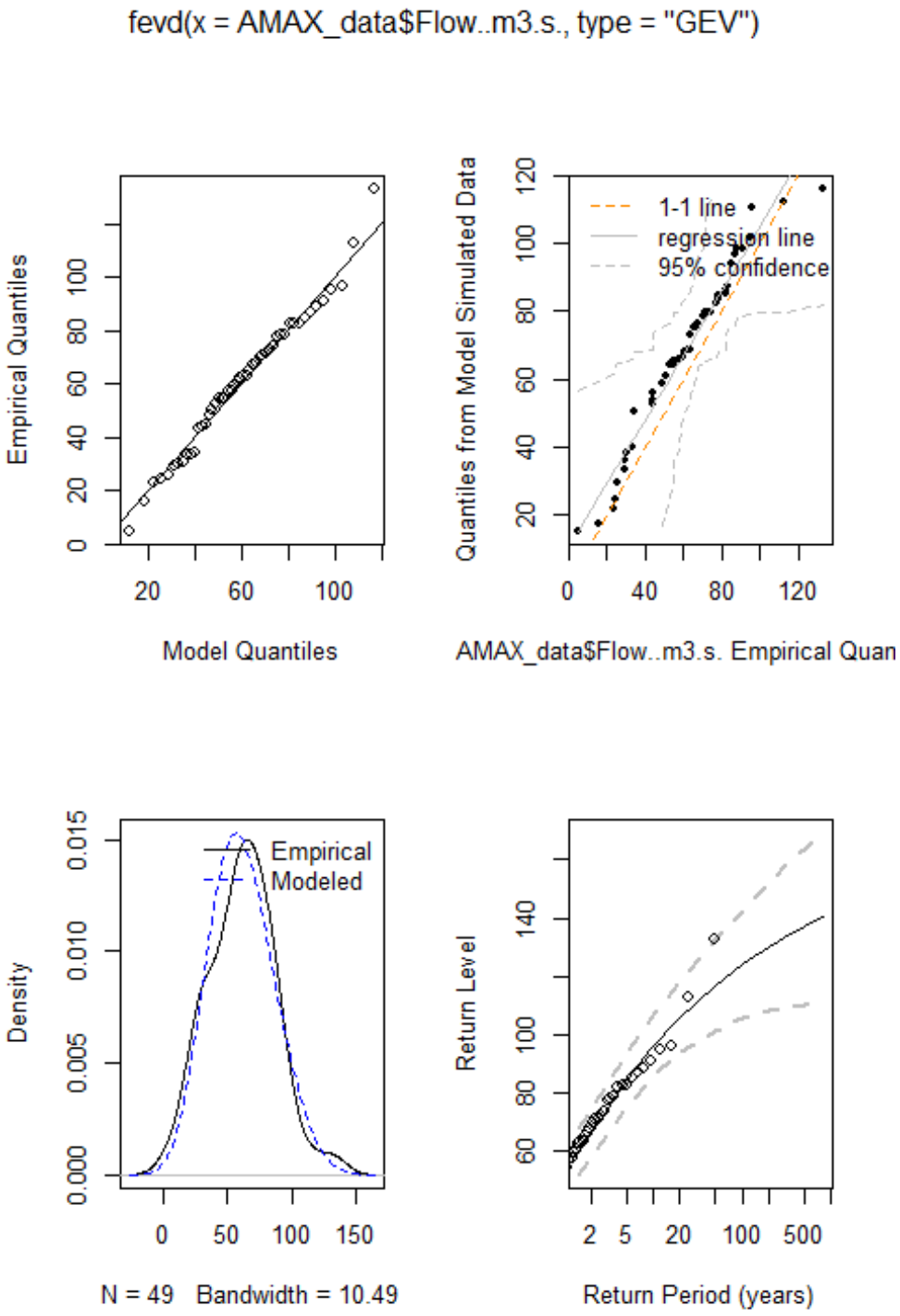
```
## 51.1756567 24.6094379 -0.2076387
##
##  Standard Error Estimates:
##   location      scale      shape
## 3.84791293 2.65009226 0.07807907
##
##  Estimated parameter covariance matrix.
##            location      scale        shape
## location 14.8064339  0.9574219 -0.104519878
## scale     0.9574219  7.0229890 -0.101435204
## shape    -0.1045199 -0.1014352  0.006096341
##
##  AIC = 462.2847
##  BIC = 467.9601
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI   Estimate 95% upper CI
## location   43.6338859 51.1756567  58.71742742
## scale      19.4153526 24.6094379  29.80352334
## shape      -0.3606709 -0.2076387  -0.05460654
```

Plotting the diagnostic plots shows a reasonable fit is achieved.

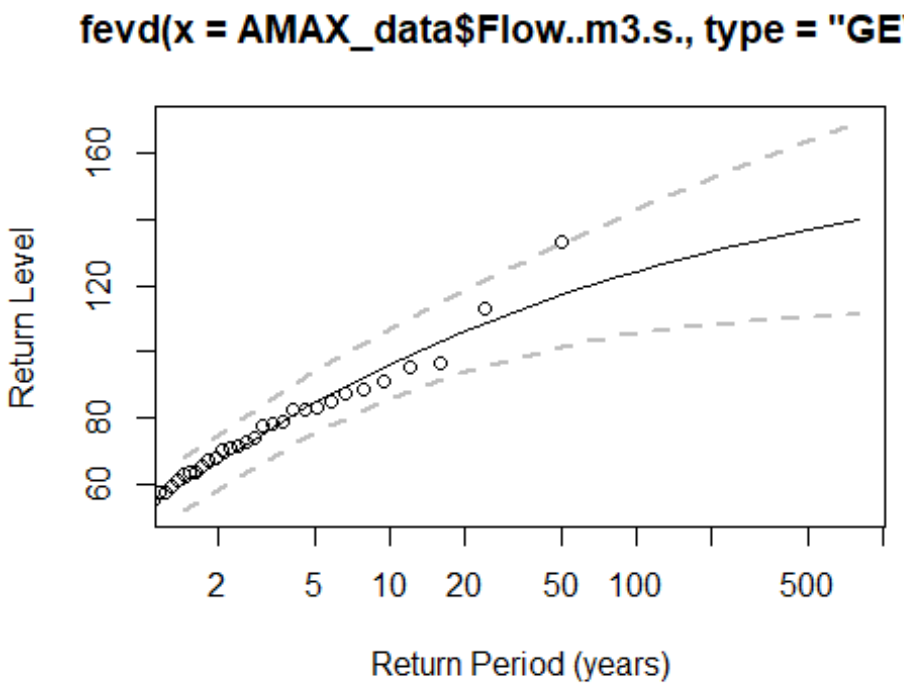
```
plot(fit)
```



AMAX

Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit,type="rl")
```




```
return.level(fit,do.ci=TRUE)

## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##                         95% lower CI  Estimate 95% upper CI
## 2-year return level        52.03774  59.86067     67.68359
## 20-year return level       93.66325 105.72971    117.79618
## 100-year return level     105.55448 124.09572    142.63696
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`51.23, 63.7`.

```
level_from_GEV_fun<-function(fit,level){
  #Get parameters
  u<-fit$results$`par`[["location"]]
  d<-fit$results$`par`[["scale"]]
  e<-fit$results$`par`[["shape"]]
  z<-level

  #calculate from Gumbel distribution equation
  a<-e*((z-u)/d)#intermediate
  b<-(1+a)^(-1/e)#intermediate
  P<-exp(-b)# Probability

  return(P) #Return the probability
```

```
}

#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit)
P

## [1] 0.3686920 0.5576691

AEP<-1-P
AEP

## [1] 0.6313080 0.4423309

RP<-1/AEP #return period (years)
RP

## [1] 1.584013 2.260751
```

For small return period events (<5 years) the distribution obtained using the AMAX and POT series are different (see http://evidence.environment-agency.gov.uk/FCERM/en/FluvialDesignGuide/Chapter2.aspx?pagenum=4). Hence, Fit a Generalised Pareto Distribution to the POT data instead.

C.8 Fit distribution to POT data

##POT Data Data file: `Input/Other catchments/Leck/33037 - Bedford Ouse at Newport Pagnell POT.csv`

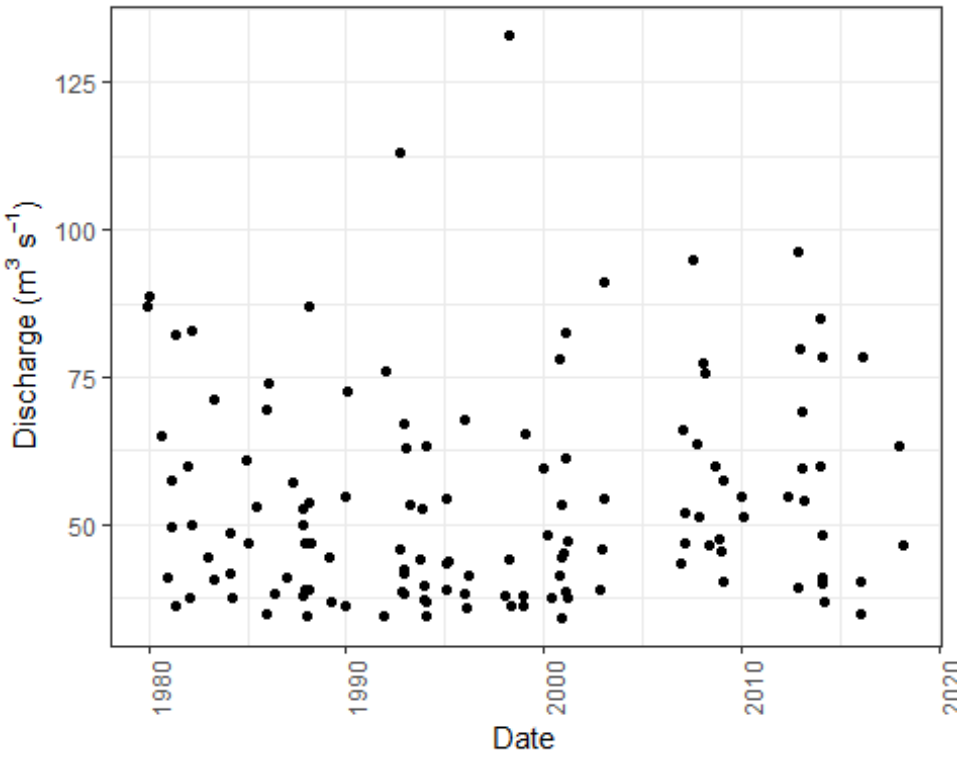
```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
POT_data<-read.csv(params$filename_POT)
POT_data$Date<-as.Date(POT_data$Date, format="%d/%m/%Y")

head(POT_data)

##   Rank       Date  Time Stage..m. Flow..m3.s. Rating          Source Ref
## 1    7 1979-12-15 08:45       N/A        87.1    N/A Digital Archive N/A
## 2    6 1979-12-29 02:30       N/A        88.6    N/A Digital Archive N/A
## 3   34 1980-08-16 20:15       N/A        65.2    N/A Digital Archive N/A
## 4  116 1980-12-21 03:30       N/A        41.2    N/A Digital Archive N/A
## 5   75 1981-03-03 15:00       N/A        49.7    N/A Digital Archive N/A
## 6   50 1981-03-12 20:45       N/A        57.5    N/A Digital Archive N/A
##   Comments
## 1       NA
## 2       NA
## 3       NA
## 4       NA
## 5       NA
## 6       NA

summary(POT_data$Flow..m3.s.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.30   39.33   47.30   53.58   62.67  133.00
```

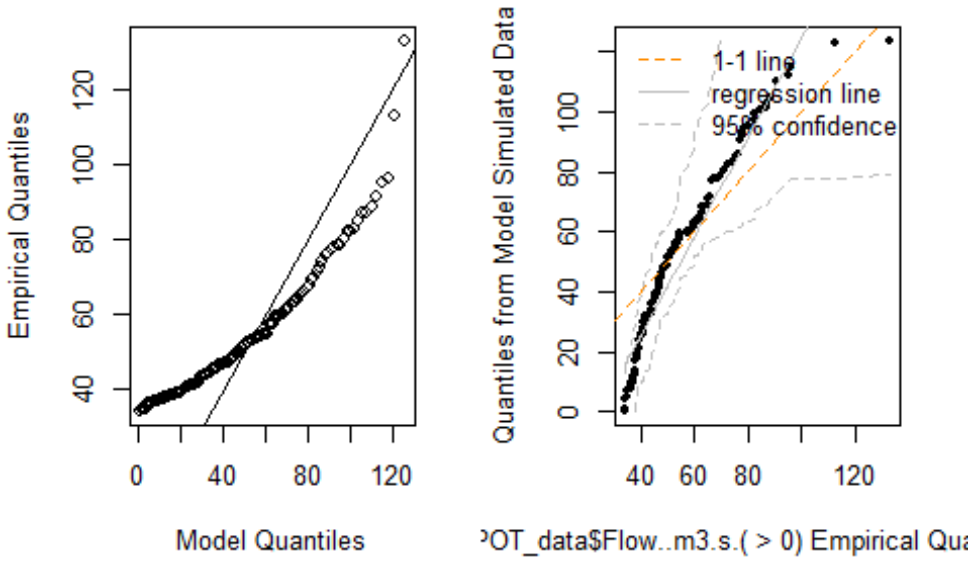


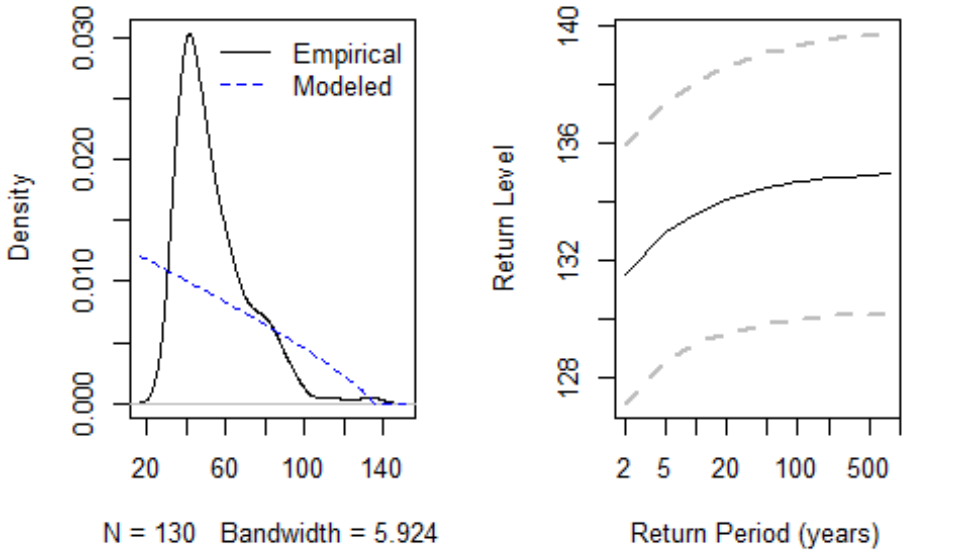
First, try to fit a Generalised-Pareto Distribution

```
#estimate number of years in POT series
library(zoo)
years<-as.yearmon(max(POT_data$Date))-as.yearmon(min(POT_data$Date))

#Terrible fit with the GP
fit_POT_GP<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GP",span=years)
plot(fit_POT_GP)
```

vd(x = POT_data$Flow..m3.s., threshold = 0, type = "GP", span = year





GP Distribution does not fit the data, hence the GEV distribution is used again.

```
#Try GEV instead
fit_POT_GEV<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GEV",span=years)
fit_POT_GEV
```

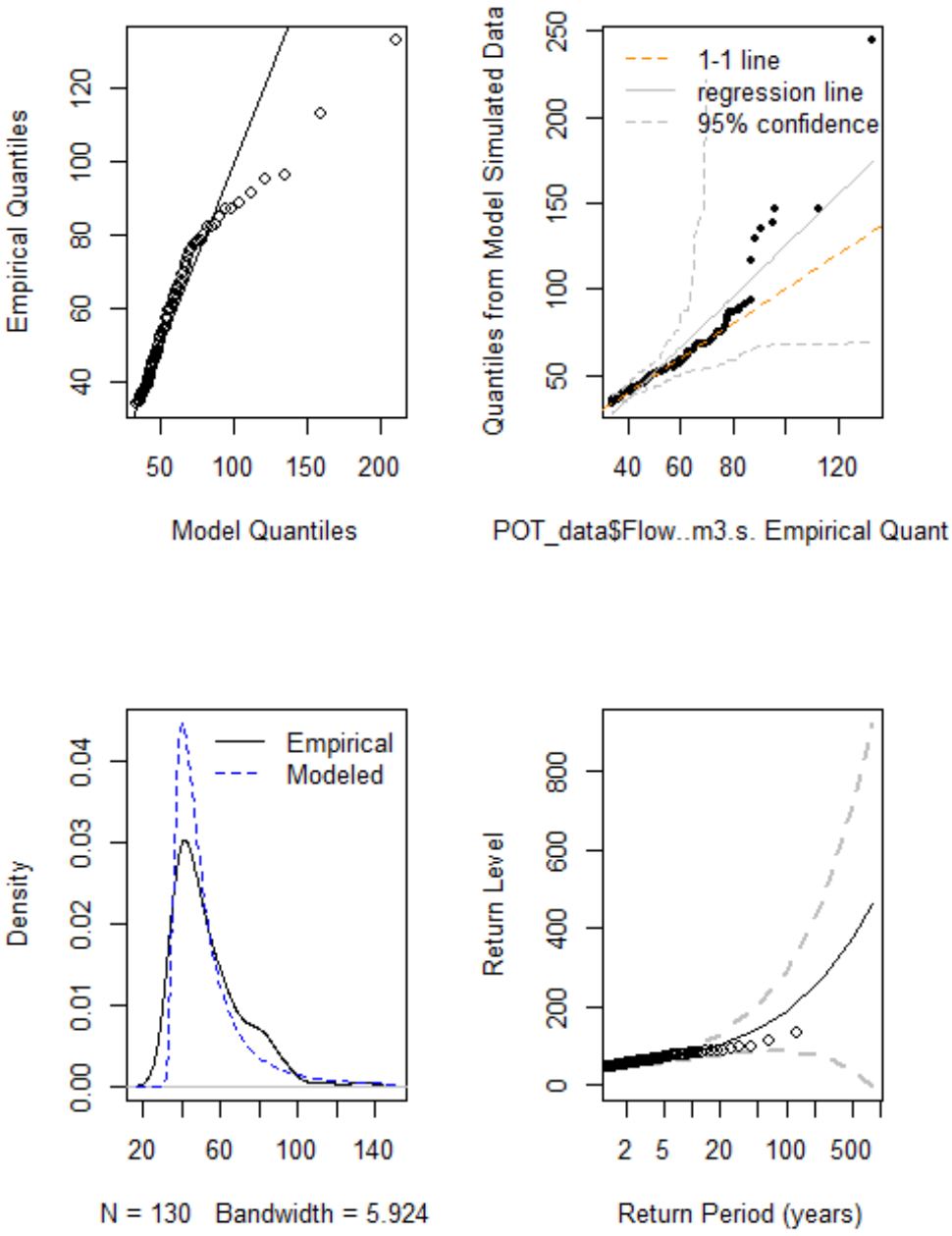
```
##
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  524.5709
##
##
##  Estimated parameters:
##   location      scale      shape
## 43.3700558  9.0314601  0.4652176
##
##  Standard Error Estimates:
##  location     scale     shape
## 0.9943562 0.9259065 0.1198595
##
##  Estimated parameter covariance matrix.
##             location       scale       shape
## location  0.98874429  0.70784591 -0.05338889
## scale     0.70784591  0.85730292 -0.02181819
## shape    -0.05338889 -0.02181819  0.01436631
##
##  AIC = 1055.142
##
##  BIC = 1063.744
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI   Estimate 95% upper CI
## location   43.6338859 51.1756567  58.71742742
## scale      19.4153526 24.6094379  29.80352334
## shape      -0.3606709 -0.2076387  -0.05460654
```

```
plot(fit_POT_GEV)
```

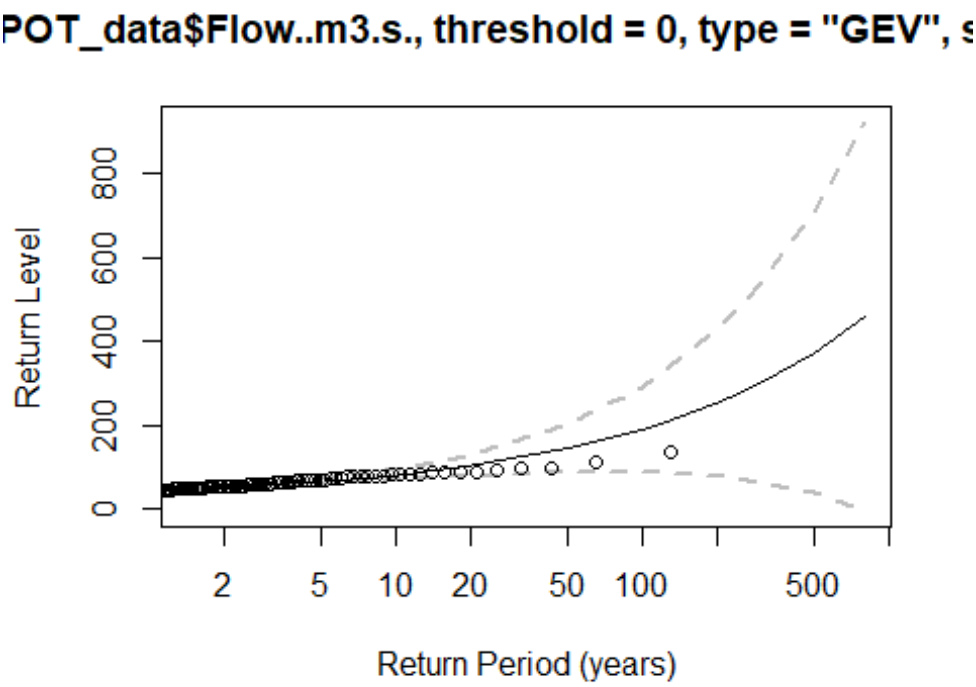
d(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = yea



The GEV Distribution produces a reasonable fit at lower return periods.The estimated shape parameter is >0, indicating a Frechet distribution is likely.

##POT Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit_POT_GEV,type="rl")
```




```
return.level(fit_POT_GEV,do.ci=TRUE)
```

```
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Normal Approx."
##
##                         95% lower CI  Estimate 95% upper CI
## 2-year return level        44.48951  46.97915      49.4688
## 20-year return level       76.19617 101.26107     126.3260
## 100-year return level      86.29556 188.97132     291.6471
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`51.23, 63.7`.

```
#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit_POT_GEV)
P
```

```
## [1] 0.6178190 0.8070569
```

```
AEP<-1-P
AEP
```

```
## [1] 0.3821810 0.1929431
```

```
RP<-1/AEP #return period (years)
RP ## [1] 2.616561 5.182876
```

Flood Frequency Estimation - Pickering

Zora van Leeuwen

11/02/2020

C.9 Introduction

A GEV Distribution is estimated for the AMAX series downloaded from the National River Flow Archive https://nrfa.ceh.ac.uk/ using the MLE method.

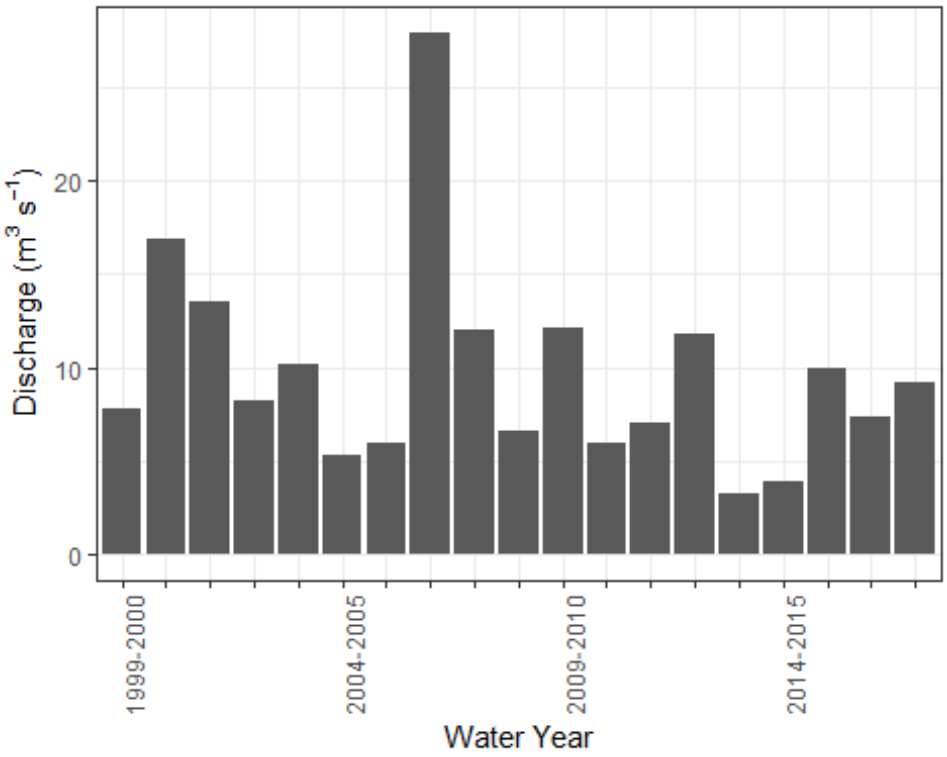
Goodness of fit plots are produced to evaluate the fit of the data along with goodness of fit measures such as the Negative log-likelihood, AIC and BIC. Standard Error Estimates are also provided. The fitted parameters are used to calculate the Annual Exceedance Probability (AEP) and Return Period (RP), the inverse of the AEP for a given level.

C.10 AMAX Data

Data file: `Input/Other catchments/27095 - Pickering Beck at Pickering_AMAX.csv`

```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
AMAX_data<-read.csv(params$filename)
head(AMAX_data)
```

```
##   Rank Water.Year       Date  Time Stage..m. Flow..m3.s.   Rating
## 1  N/A  1999-2000 20/09/2000 13:30     0.845       7.766 In Range
## 2    2  2000-2001 08/11/2000 10:30     1.532      16.897 In Range
## 3    3  2001-2002 02/08/2002 14:45     1.340      13.491 In Range
## 4   10  2002-2003 03/11/2002 10:45     0.890       8.261 In Range
## 5    7  2003-2004 01/02/2004 01:45     1.060      10.180 In Range
## 6   16  2004-2005 16/04/2005 08:00     0.610       5.284 In Range
##            Source Ref Comments
## 1 Digital Archive  B2       NA
## 2 Digital Archive  B3       NA
## 3 Digital Archive  B2       NA
## 4 Digital Archive  B2       NA
## 5 Digital Archive  B2       NA
## 6 Digital Archive  B2       NA
```



Plot of the AMAX series which is used to fit the GEV distribution

The plot shows the AMAX series contained in the `Input/Other catchments/27095 - Pickering Beck at Pickering_AMAX.csv` dataset.

C.11 Fit GEV to AMAX series

The Package extRemes is used (Eric Gilleland, Richard W. Katz (2016). extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1-39. doi:10.18637/jss.v072.i08). The estimated shape parameter < 0, indicating a Weibull distribution.

```
library(extRemes)

fit<-fevd(AMAX_data$Flow..m3.s.,type="GEV")
fit
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Estimation Method used: MLE"
##
##  Negative Log-Likelihood Value:  54.35152
##
##  Estimated parameters:
##  location     scale     shape
## 7.1707517 3.2576790 0.1780202
##
##  Standard Error Estimates:
```

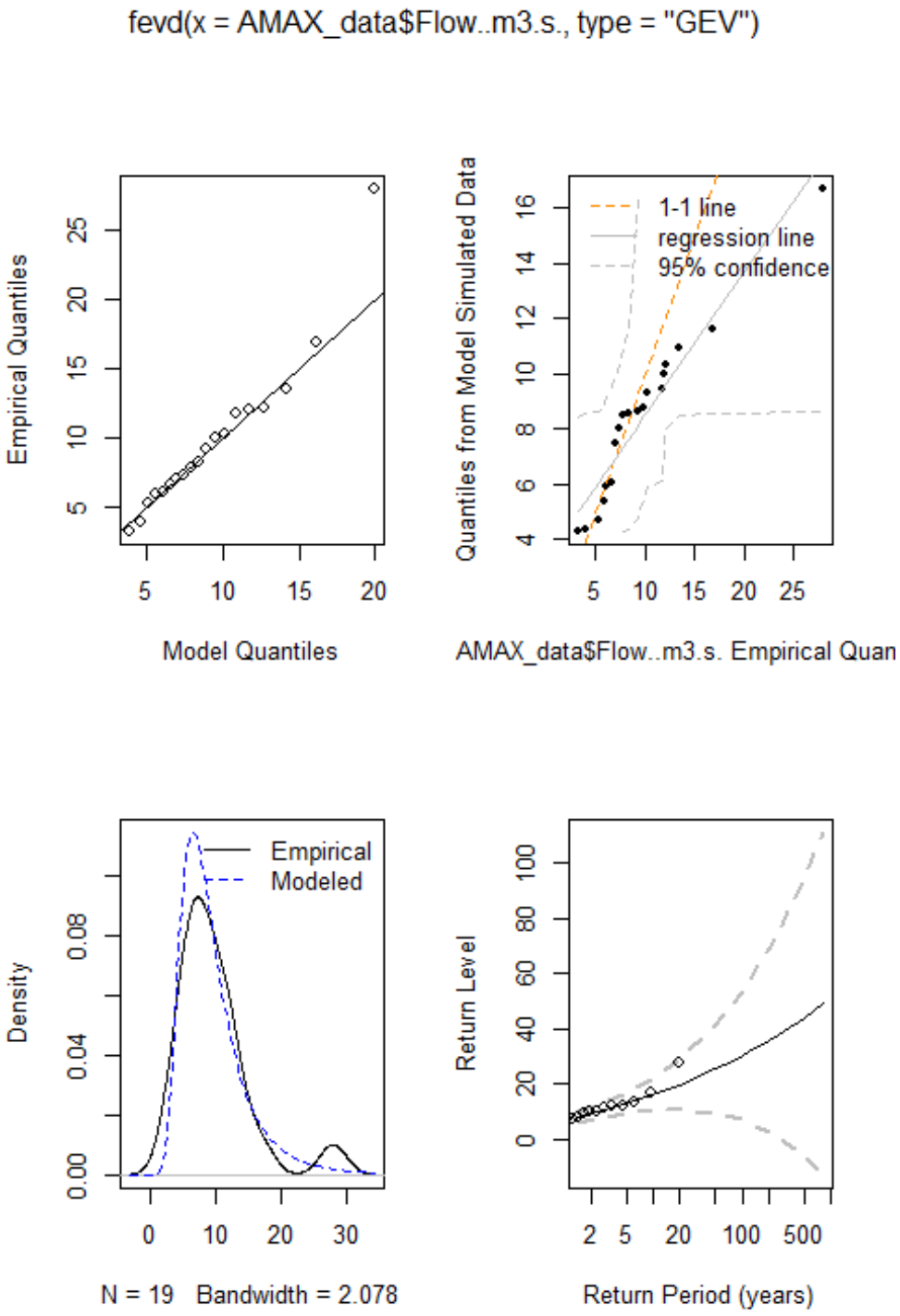
```
##  location     scale     shape
## 0.8498189 0.6657188 0.1881403
##
##  Estimated parameter covariance matrix.
##              location       scale       shape
## location  0.72219223  0.31210645 -0.05162366
## scale     0.31210645  0.44318152 -0.01371584
## shape    -0.05162366 -0.01371584  0.03539677
##
##  AIC = 114.703
##  BIC = 117.5363
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI  Estimate 95% upper CI
## location     5.505137 7.1707517    8.8363662
## scale        1.952894 3.2576790    4.5624639
## shape       -0.190728 0.1780202    0.5467684
```

Plotting the diagnostic plots shows a reasonable fit is achieved.

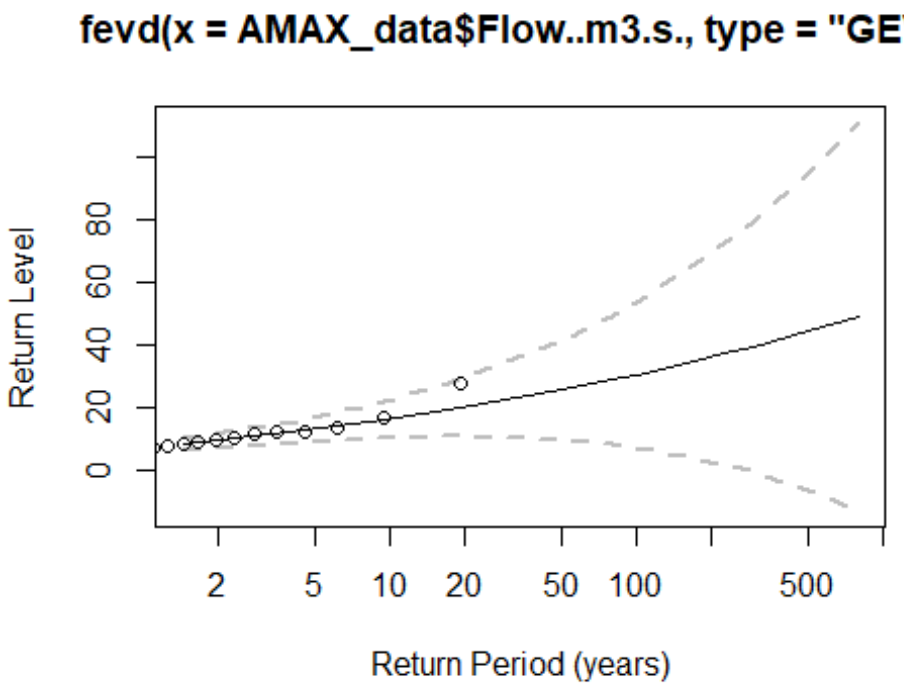
```
plot(fit)
```



AMAX

Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit,type="rl")
```




```
return.level(fit,do.ci=TRUE)
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##                         95% lower CI  Estimate 95% upper CI
## 2-year return level         6.446495  8.404546     10.36260
## 20-year return level       11.018741 19.922168     28.82559
## 100-year return level       7.230037 30.375384     53.52073
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`9.916, 11.778`.

```
level_from_GEV_fun<-function(fit,level){
  #Get parameters
  u<-fit$results$`par`[["location"]]
  d<-fit$results$`par`[["scale"]]
  e<-fit$results$`par`[["shape"]]
  z<-level

  #calculate from Gumbel distribution equation
  a<-e*((z-u)/d)#intermediate
  b<-(1+a)^(-1/e)#intermediate
  P<-exp(-b)# Probability

  return(P) #Return the probability
```

```
}

#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit)
P

## [1] 0.6337891 0.7533299

AEP<-1-P
AEP

## [1] 0.3662109 0.2466701

RP<-1/AEP #return period (years)
RP

## [1] 2.730667 4.053997
```

For small return period events (<5 years) the distribution obtained using the AMAX and POT series are different (see http://evidence.environment-agency.gov.uk/FCERM/en/FluvialDesignGuide/Chapter2.aspx?pagenum=4). Hence, Fit a Generalised Pareto Distribution to the POT data instead.

C.12 Fit distribution to POT data

##POT Data Data file: `Input/Other catchments/27095 - Pickering Beck at Pickering_POT.csv`

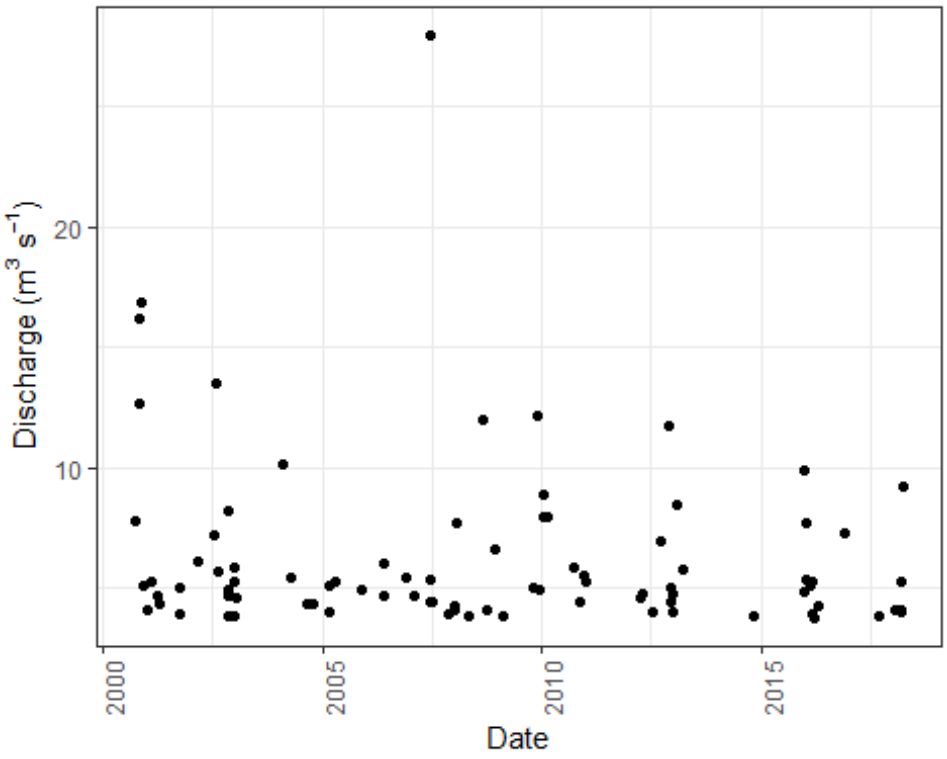
```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
POT_data<-read.csv(params$filename_POT)
POT_data$Date<-as.Date(POT_data$Date, format="%d/%m/%Y")

head(POT_data)

##   Rank       Date  Time Stage..m. Flow..m3.s.   Rating          Source Ref
## 1   17 2000-09-20 13:30     0.845       7.766 In Range Digital Archive  B2
## 2    3 2000-10-31 08:30     1.501      16.230 In Range Digital Archive  B3
## 3    5 2000-11-03 11:00     1.273      12.683 In Range Digital Archive  B2
## 4    2 2000-11-08 10:30     1.532      16.897 In Range Digital Archive  B3
## 5   41 2000-12-08 21:45     0.592       5.101 In Range Digital Archive  B2
## 6   71 2001-01-02 23:45     0.498       4.086 In Range Digital Archive  B1
##   Comments
## 1       NA
## 2       NA
## 3       NA
## 4       NA
## 5       NA
## 6       NA

summary(POT_data$Flow..m3.s.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.806   4.338   5.061   6.305   6.896  27.942
```

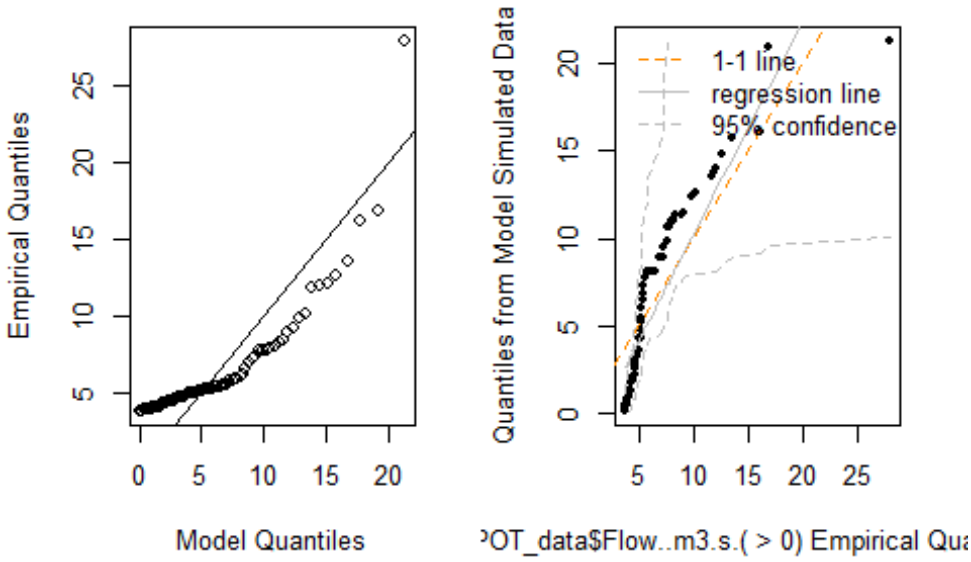


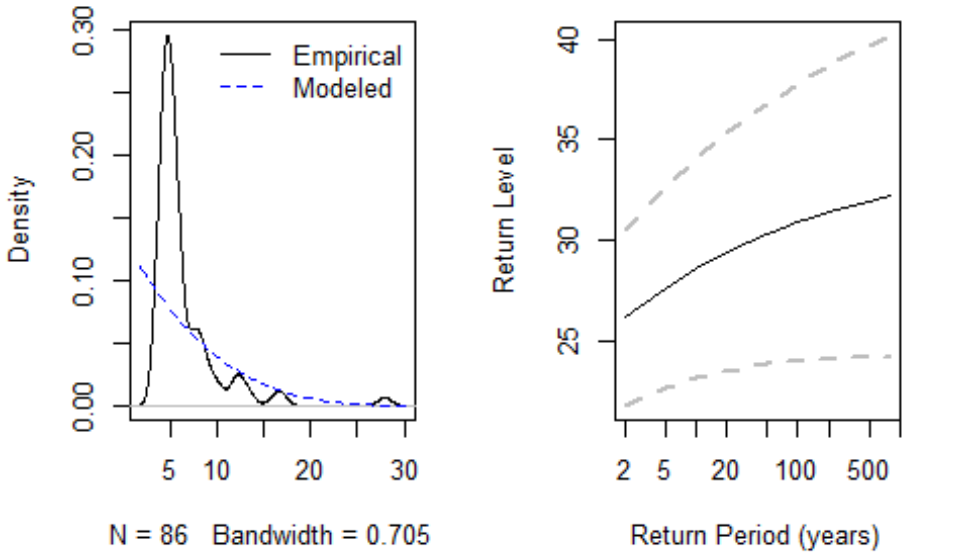
First, try to fit a Generalised-Pareto Distribution

```
#estimate number of years in POT series
library(zoo)
years<-as.yearmon(max(POT_data$Date))-as.yearmon(min(POT_data$Date))

#Terrible fit with the GP
fit_POT_GP<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GP",span=years)
plot(fit_POT_GP)
```

vd(x = POT_data$Flow..m3.s., threshold = 0, type = "GP", span = year





GP Distribution does not fit the data, hence the GEV distribution is used again.

```
#Try GEV instead
fit_POT_GEV<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GEV",span=years)
fit_POT_GEV
```

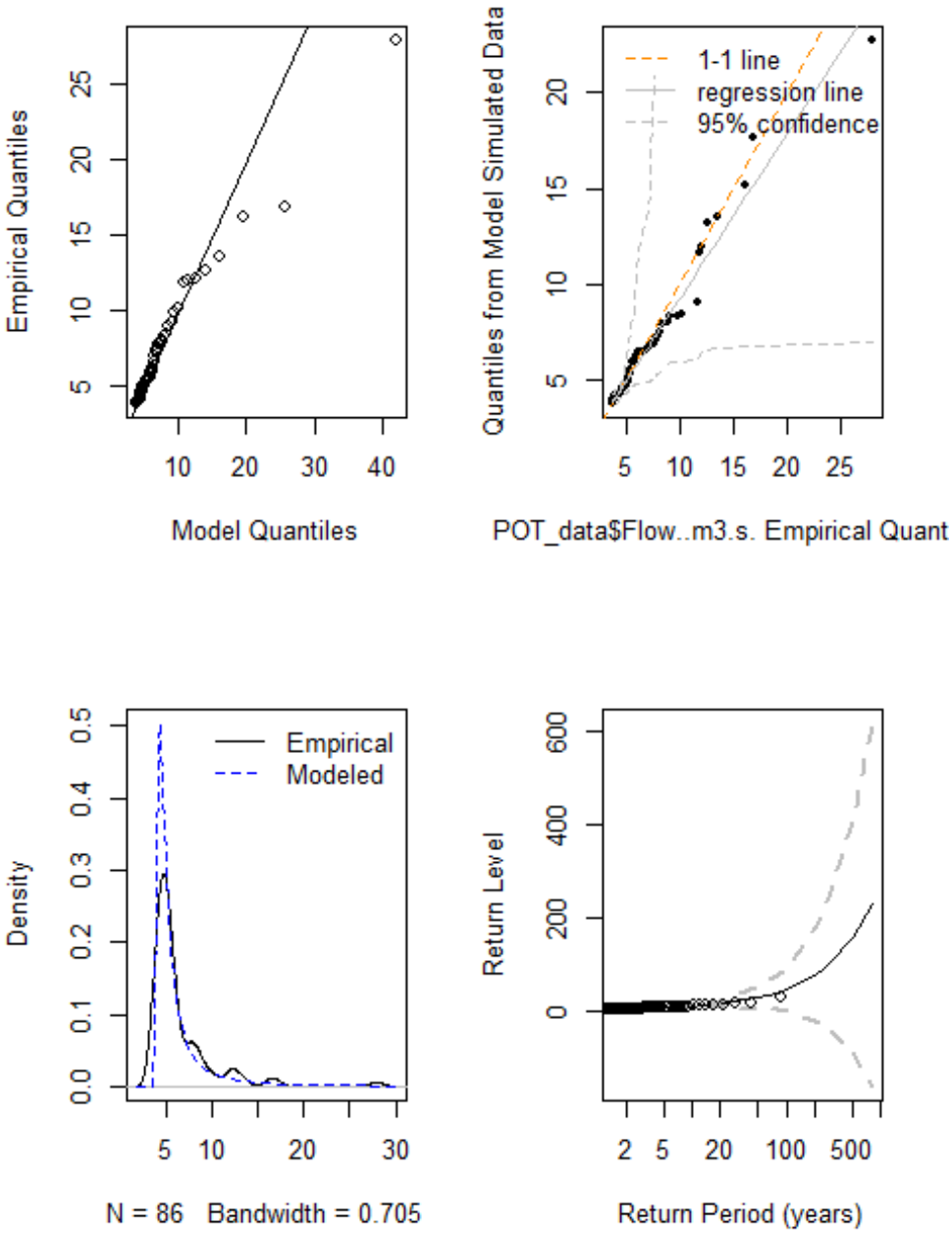
```
##
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  164.1223
##
##
##  Estimated parameters:
##  location     scale     shape
## 4.5561686 0.9037904 0.7890557
##
##  Standard Error Estimates:
##  location     scale     shape
## 0.1204020 0.1352730 0.1608015
##
##  Estimated parameter covariance matrix.
##              location        scale         shape
## location  0.014496641 0.0136412978 -0.0069591258
## scale     0.013641298 0.0182987895  0.0002755753
## shape    -0.006959126 0.0002755753  0.0258571081
##
##  AIC = 334.2446
##
##  BIC = 341.6076
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI  Estimate 95% upper CI
## location     5.505137 7.1707517    8.8363662
## scale        1.952894 3.2576790    4.5624639
## shape       -0.190728 0.1780202    0.5467684
```

```
plot(fit_POT_GEV)
```

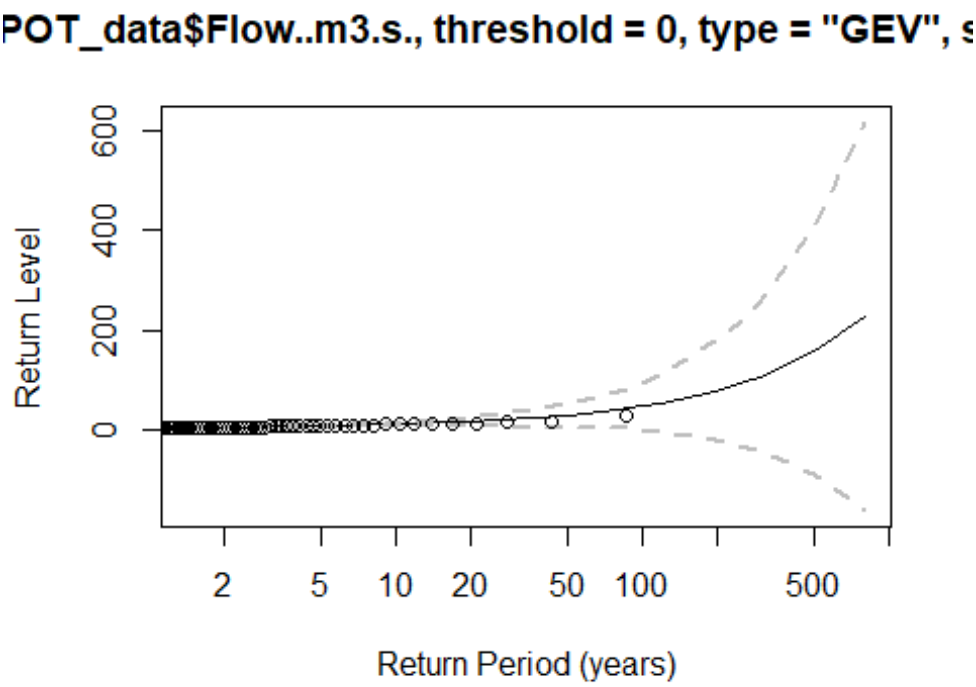
d(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = yea



The GEV Distribution produces a reasonable fit at lower return periods.The estimated shape parameter is >0, indicating a Frechet distribution is likely.

##POT Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit_POT_GEV,type="rl")
```




```
return.level(fit_POT_GEV,do.ci=TRUE)
```

```
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Normal Approx."
##
##                         95% lower CI Estimate 95% upper CI
## 2-year return level        4.6092438  4.94029     5.271336
## 20-year return level       7.7422265 15.34498    22.947738
## 100-year return level     -0.9742641 46.59754    94.169347
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`9.916, 11.778`.

```
#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit_POT_GEV)
P
```

```
## [1] 0.8952311 0.9227056
```

```
AEP<-1-P
AEP
```

```
## [1] 0.10476889 0.07729437
```

```
RP<-1/AEP #return period (years)
RP ## [1]  9.544818 12.937554
```

Flood Frequency Estimation - Flasby

Zora van Leeuwen

11/02/2020

C.13 Introduction

A GEV Distribution is estimated for the AMAX series downloaded from the National River Flow Archive https://nrfa.ceh.ac.uk/ using the MLE method.

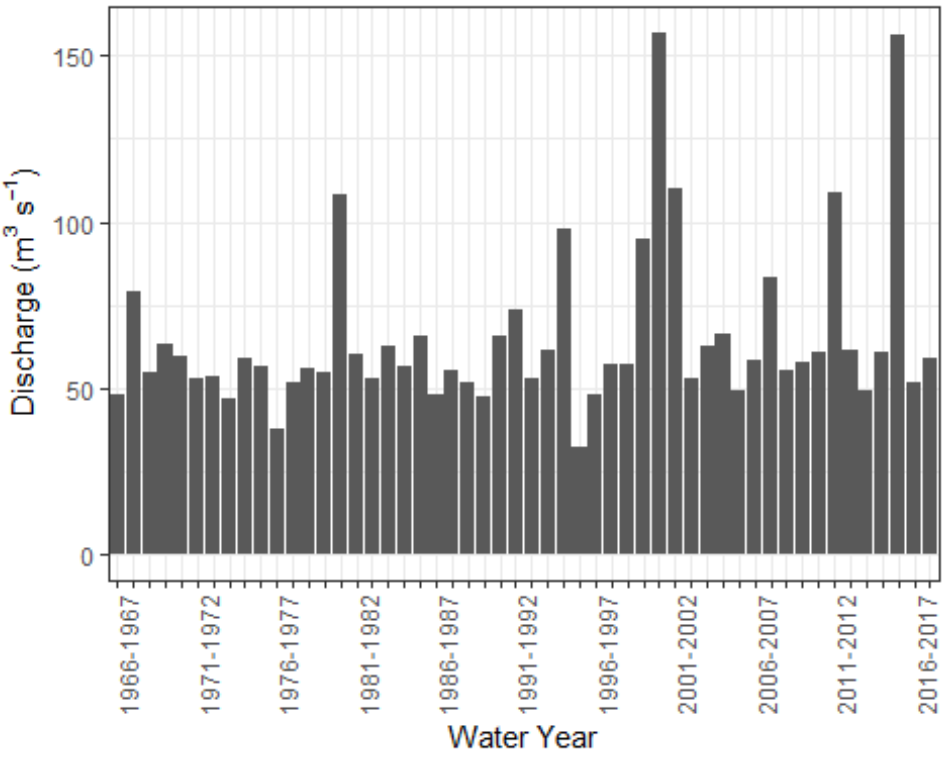
Goodness of fit plots are produced to evaluate the fit of the data along with goodness of fit measures such as the Negative log-likelihood, AIC and BIC. Standard Error Estimates are also provided. The fitted parameters are used to calculate the Annual Exceedance Probability (AEP) and Return Period (RP), the inverse of the AEP for a given level.

C.14 AMAX Data

Data file: `Input/Other catchments/Aire/27035 - Aire at Kildwick Bridge_AMAX.csv`

```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
AMAX_data<-read.csv(params$filename)
head(AMAX_data)
```

```
##   Rank Water.Year       Date  Time Stage..m. Flow..m3.s.   Rating
## 1   48  1966-1967 01/10/1967 00:00     1.780      47.889 In Range
## 2    9  1967-1968 17/10/1967 00:00     2.990      78.757  Extrap.
## 3   34  1968-1969 31/03/1969 00:00     2.020      54.567 In Range
## 4   14  1969-1970 11/11/1969 00:00     2.360      63.481  Extrap.
## 5   22  1970-1971 30/10/1970 00:00     2.220      59.879  Extrap.
## 6   37  1971-1972 19/10/1971 11:15     2.239      53.000 In Range
##            Source Ref Comments
## 1       CEH Files  1b       NA
## 2       CEH Files  1b       NA
## 3       CEH Files  1b       NA
## 4       CEH Files  1b       NA
## 5       CEH Files  1b       NA
## 6 Digital Archive  C3       NA
```



Plot of the AMAX series which is used to fit the GEV distribution

The plot shows the AMAX series contained in the `Input/Other catchments/Aire/27035 - Aire at Kildwick Bridge_AMAX.csv` dataset.

C.15 Fit GEV to AMAX series

The Package extRemes is used (Eric Gilleland, Richard W. Katz (2016). extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1-39. doi:10.18637/jss.v072.i08). The estimated shape parameter < 0, indicating a Weibull distribution.

```
library(extRemes)

fit<-fevd(AMAX_data$Flow..m3.s.,type="GEV")
fit
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Estimation Method used: MLE"
##
##  Negative Log-Likelihood Value:  220.3542
##
##  Estimated parameters:
##   location      scale      shape
## 54.6819808 12.9910227  0.1642523
##
##  Standard Error Estimates:
```

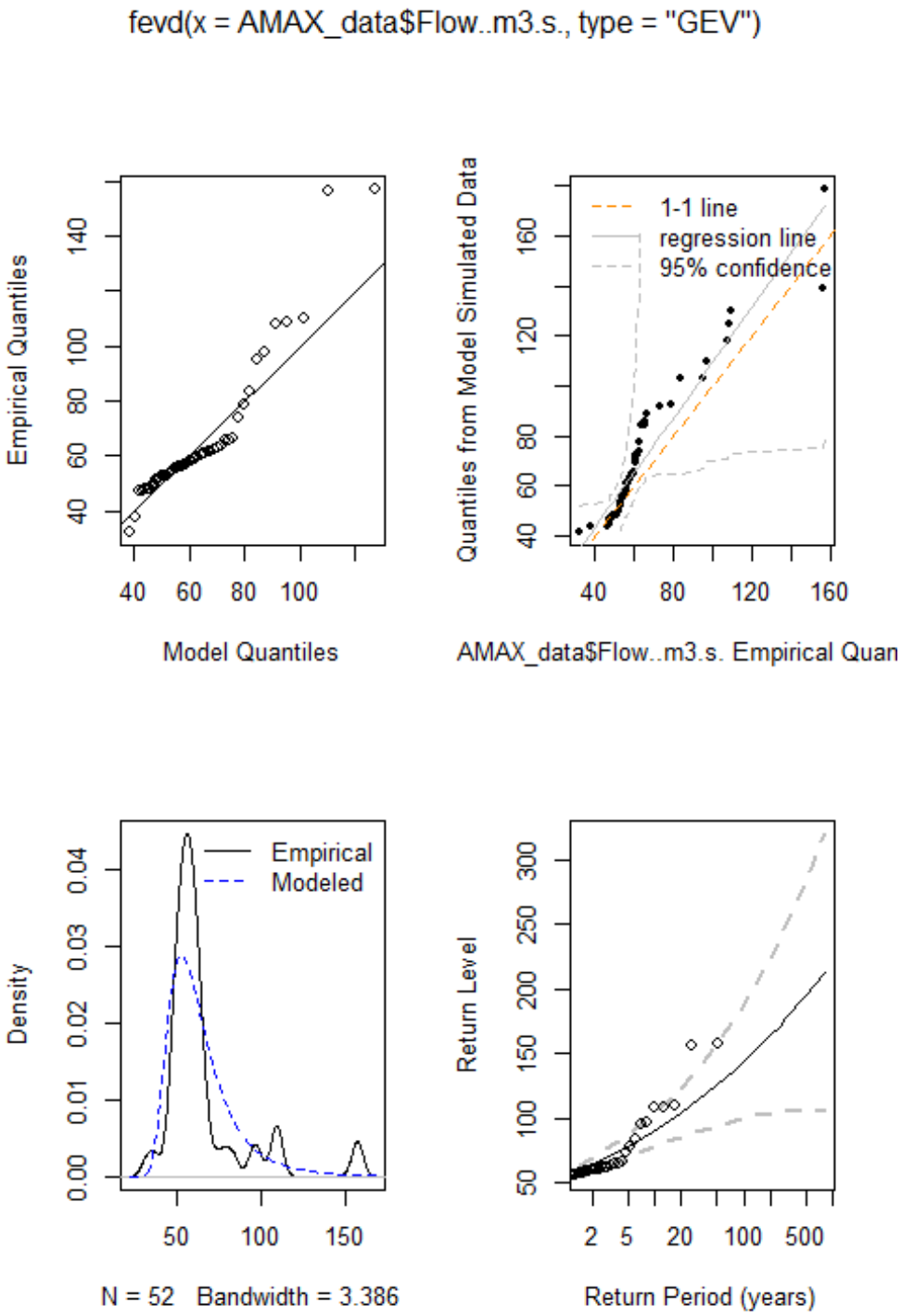
```
##   location      scale      shape
## 1.95108379 1.47648299 0.07734388
##
##  Estimated parameter covariance matrix.
##              location      scale        shape
## location  3.80672796 1.37979688 -0.026915191
## scale     1.37979688 2.18000201  0.021267510
## shape    -0.02691519 0.02126751  0.005982075
##
##  AIC = 446.7085
##  BIC = 452.5622
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI   Estimate 95% upper CI
## location  50.85792689 54.6819808   58.5060348
## scale     10.09716924 12.9910227   15.8848762
## shape      0.01266112  0.1642523    0.3158435
```

Plotting the diagnostic plots shows a reasonable fit is achieved.

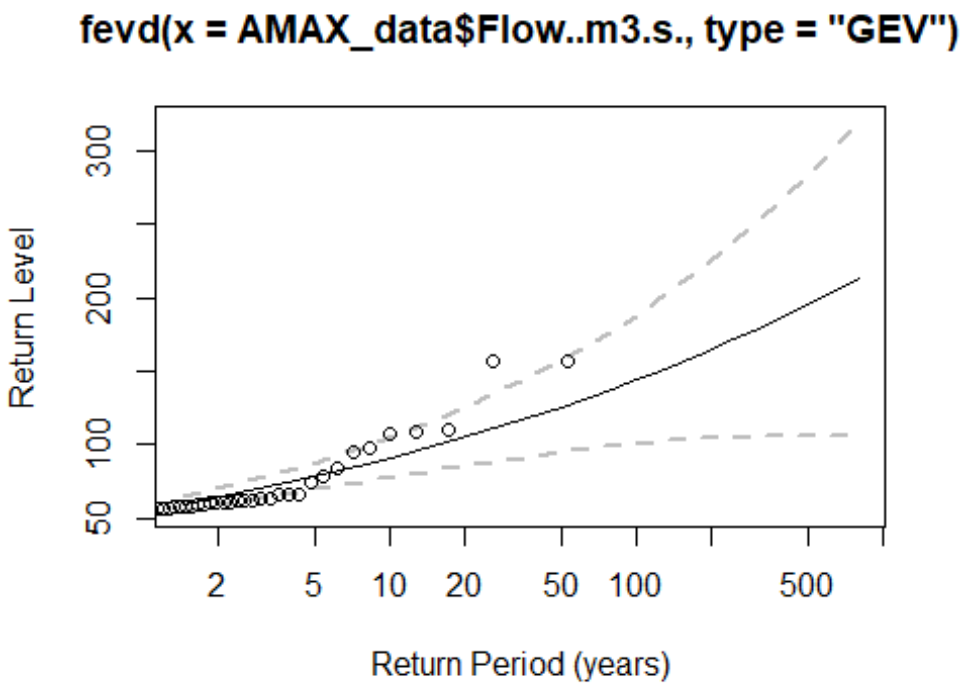
```
plot(fit)
```



AMAX

Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit,type="rl")
```




```
return.level(fit,do.ci=TRUE)
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##                         95% lower CI Estimate 95% upper CI
## 2-year return level        55.14996  59.5896     64.02923
## 20-year return level       85.49943 104.4173    123.33508
## 100-year return level     100.14032 143.9654    187.79046
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`75.662, 107`.

```
level_from_GEV_fun<-function(fit,level){
  #Get parameters
  u<-fit$results$`par`[["location"]]
  d<-fit$results$`par`[["scale"]]
  e<-fit$results$`par`[["shape"]]
  z<-level

  #calculate from Gumbel distribution equation
  a<-e*((z-u)/d)#intermediate
  b<-(1+a)^(-1/e)#intermediate
  P<-exp(-b)# Probability

  return(P) #Return the probability
```

```
}

#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit)
P

## [1] 0.7876276 0.9555631

AEP<-1-P
AEP

## [1] 0.21237239 0.04443686

RP<-1/AEP #return period (years)
RP

## [1]  4.70871 22.50384
```

For small return period events (<5 years) the distribution obtained using the AMAX and POT series are different (see http://evidence.environment-agency.gov.uk/FCERM/en/FluvialDesignGuide/Chapter2.aspx?pagenum=4). Hence, Fit a Generalised Pareto Distribution to the POT data instead.

C.16 Fit distribution to POT data

##POT Data Data file: `Input/Other catchments/Aire/27035 - Aire at Kildwick Bridge_POT.csv`

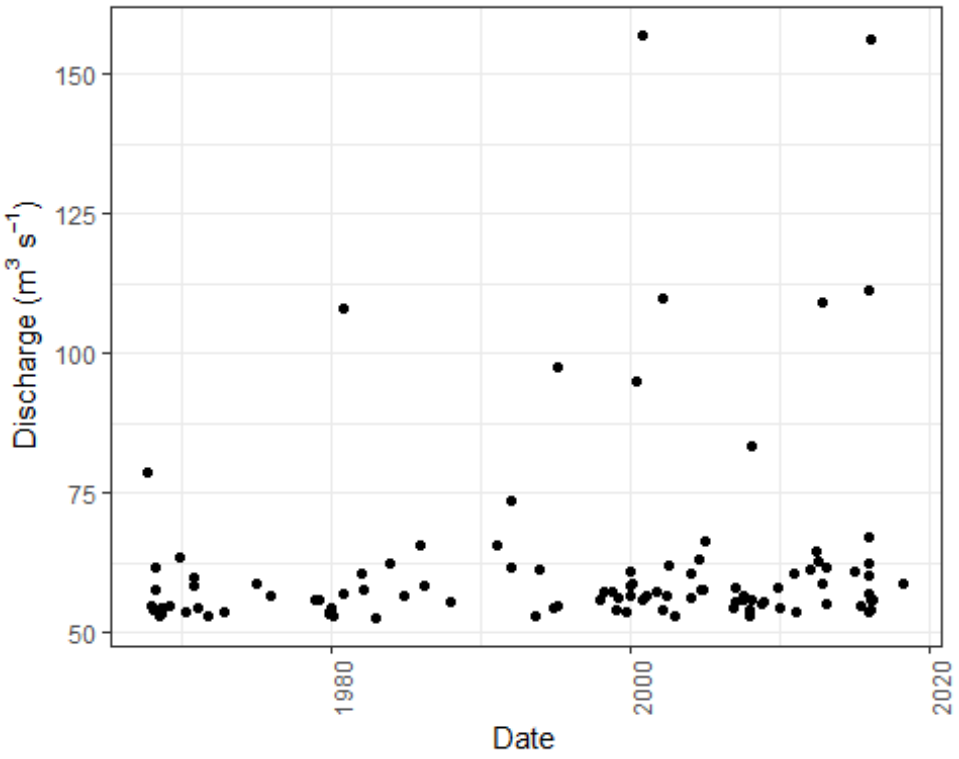
```
setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
POT_data<-read.csv(params$filename_POT)
POT_data$Date<-as.Date(POT_data$Date, format="%d/%m/%Y")

head(POT_data)

##   Rank       Date  Time Stage..m. Flow..m3.s.   Rating    Source Ref Comme
nts
## 1   10 1967-10-17 00:00      2.99      78.757  Extrap. CEH Files  1b
NA
## 2   76 1968-01-16 00:00      2.02      54.567 In Range CEH Files  1b
NA
## 3   87 1968-03-19 00:00      2.00      54.024 In Range CEH Files  1b
NA
## 4   24 1968-03-23 00:00      2.29      61.691  Extrap. CEH Files  1b
NA
## 5   45 1968-03-24 00:00      2.14      57.779 In Range CEH Files  1b
NA
## 6  100 1968-07-01 00:00      1.96      52.932 In Range CEH Files  1b
NA

summary(POT_data$Flow..m3.s.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   52.70   54.53   56.88   62.60   61.25  157.00
```

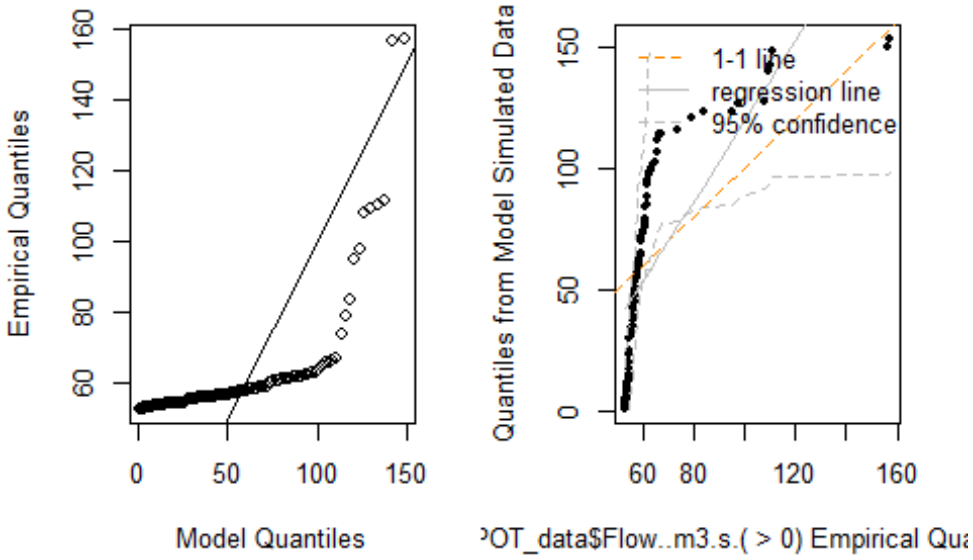


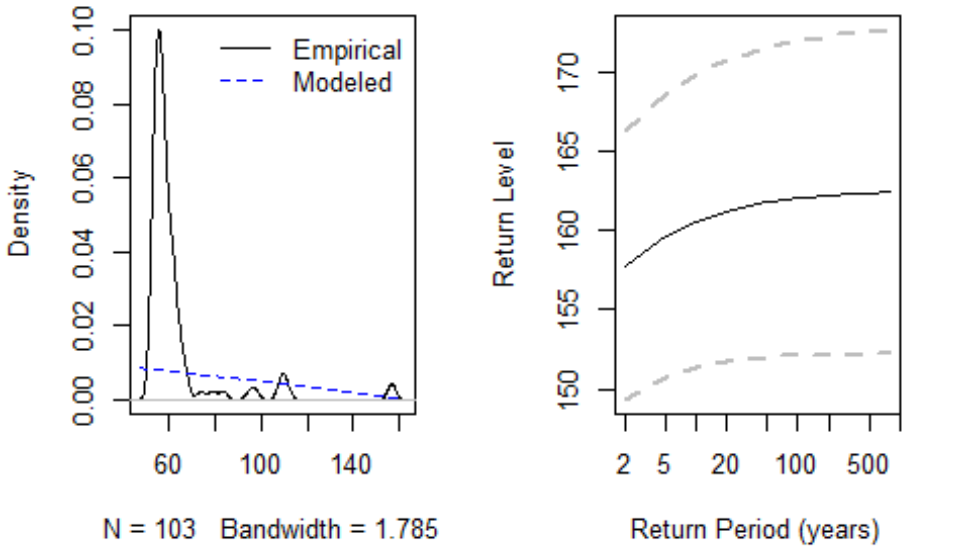
First, try to fit a Generalised-Pareto Distribution

```
#estimate number of years in POT series
library(zoo)
years<-as.yearmon(max(POT_data$Date))-as.yearmon(min(POT_data$Date))

#Terrible fit with the GP
fit_POT_GP<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GP",span=years)
plot(fit_POT_GP)
```

vd(x = POT_data$Flow..m3.s., threshold = 0, type = "GP", span = year





GP Distribution does not fit the data, hence the GEV distribution is used again.

```
#Try GEV instead
fit_POT_GEV<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GEV",span=years)
fit_POT_GEV
```

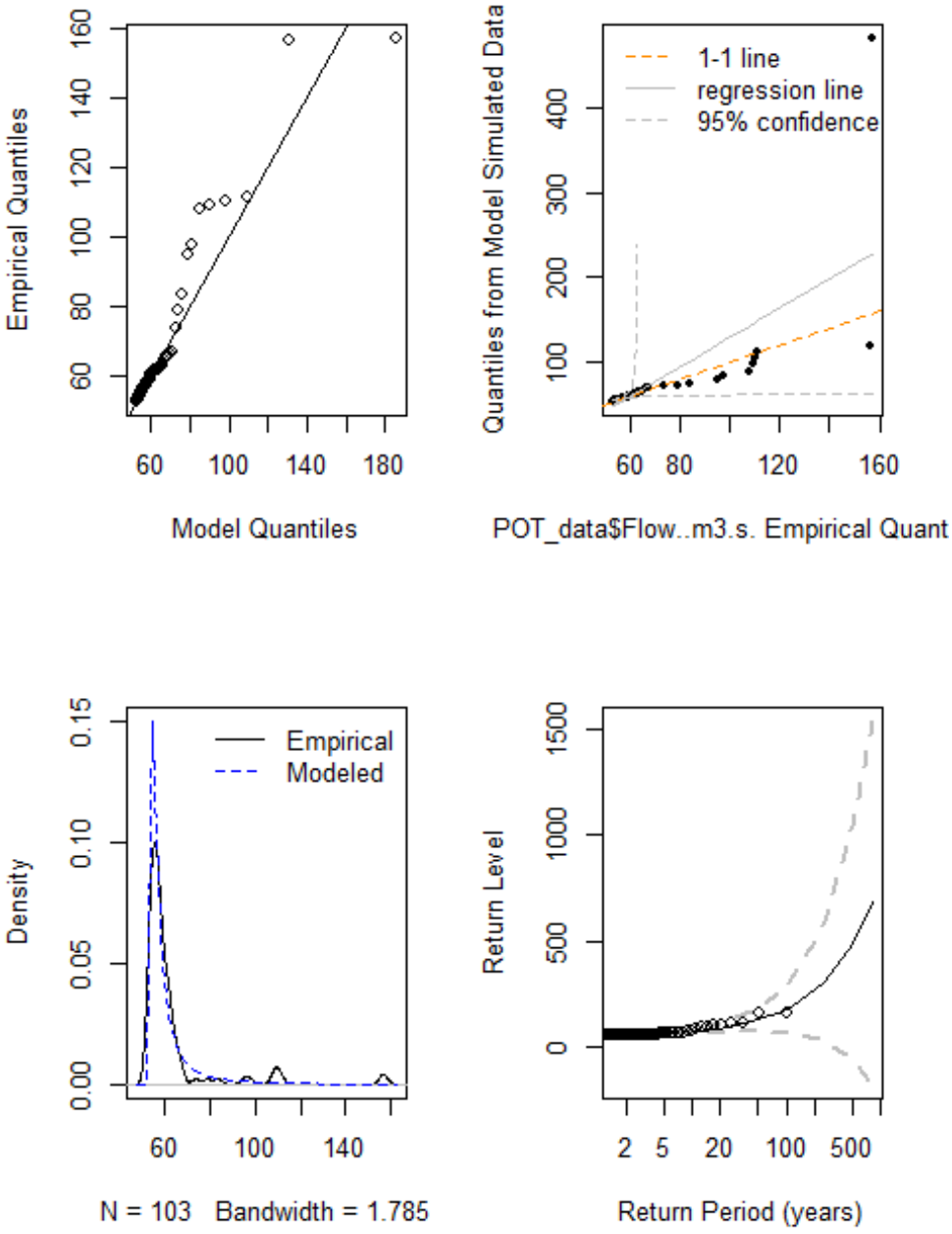
```
##
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  321.582
##
##
##  Estimated parameters:
##   location      scale      shape
## 55.4269156  3.0474042  0.7555805
##
##  Standard Error Estimates:
##  location     scale     shape
## 0.3508209 0.3942199 0.1232556
##
##  Estimated parameter covariance matrix.
##             location       scale        shape
## location  0.12307533 0.112826703 -0.010724024
## scale     0.11282670 0.155409349  0.007920718
## shape    -0.01072402 0.007920718  0.015191937
##
##  AIC = 649.164
##
##  BIC = 657.0682
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI   Estimate 95% upper CI
## location  50.85792689 54.6819808   58.5060348
## scale     10.09716924 12.9910227   15.8848762
## shape      0.01266112  0.1642523    0.3158435
```

```
plot(fit_POT_GEV)
```

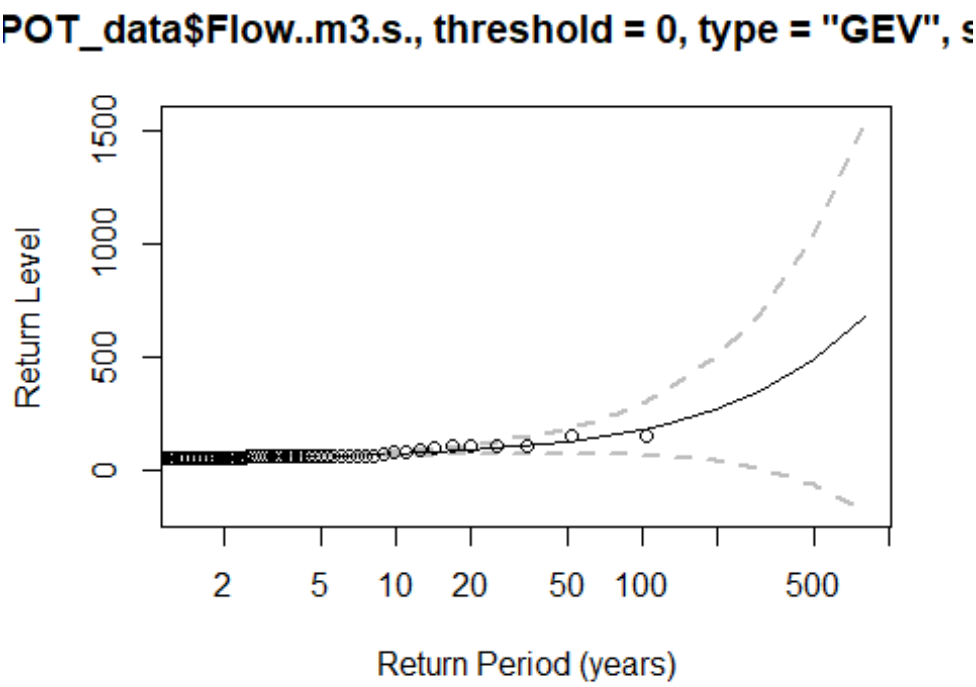
d(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = yea



The GEV Distribution produces a reasonable fit at lower return periods.The estimated shape parameter is >0, indicating a Frechet distribution is likely.

##POT Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit_POT_GEV,type="rl")
```




```
return.level(fit_POT_GEV,do.ci=TRUE)
```

```
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Normal Approx."
##
##                         95% lower CI  Estimate 95% upper CI
## 2-year return level        55.74692  56.71380     57.68069
## 20-year return level       69.56843  89.43914    109.30985
## 100-year return level      67.59093 181.75923    295.92753
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`75.662, 107`.

```
#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit_POT_GEV)
P
```

```
## [1] 0.9111923 0.9694368
```

```
AEP<-1-P
AEP
```

```
## [1] 0.08880766 0.03056321
```

```
RP<-1/AEP #return period (years)
RP ## [1] 11.26029 32.71908
```

Flood Frequency Estimation- Shipston

Zora van Leeuwen

11/02/2020

1.2 Introduction

A GEV Distribution is estimated for the AMAX series downloaded from the National River Flow Archive https://nrfa.ceh.ac.uk/ using the MLE method.

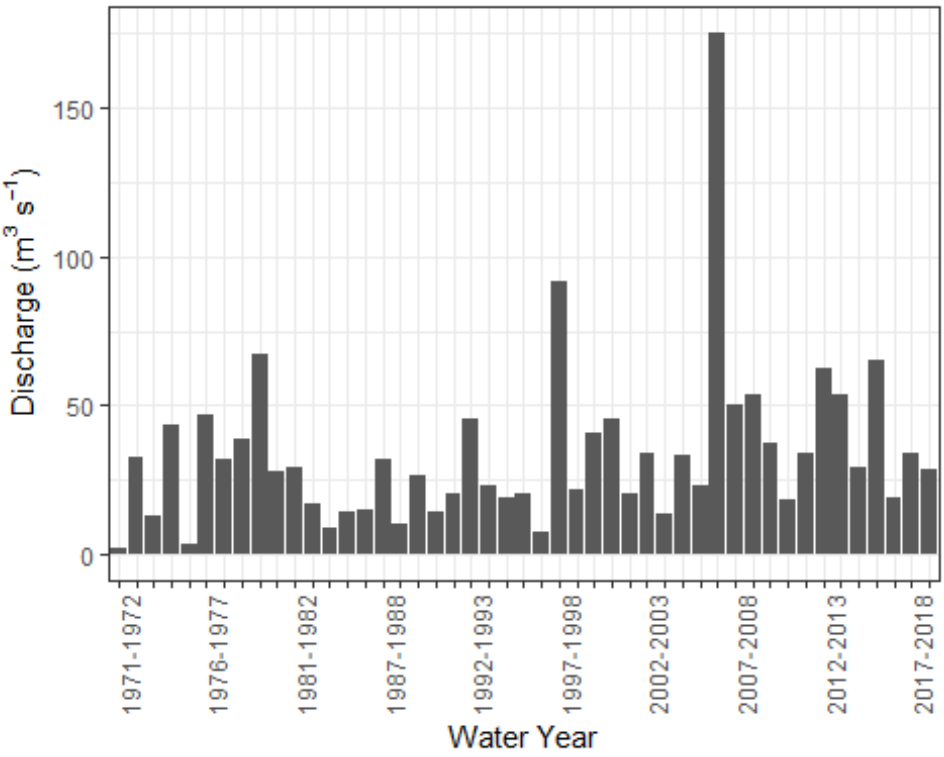
Goodness of fit plots are produced to evaluate the fit of the data along with goodness of fit measures such as the Negative log-likelihood, AIC and BIC. Standard Error Estimates are also provided. The fitted parameters are used to calculate the Annual Exceedance Probability (AEP) and Return Period (RP), the inverse of the AEP for a given level.

1.3 AMAX Data

Data file: `Input/Other catchments/Shipston/54106 Stour at Shipston AMAX.csv`

```
#setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
AMAX_data<-read.csv(params$filename)
head(AMAX_data)
```

```
##   Rank Water.Year       Date  Time Stage..m. Flow..m3.s.   Rating
## 1   46  1971-1972 24/04/1972 13:15     0.522       1.790 In Range
## 2   20  1972-1973 07/12/1972 03:15     2.984      32.460  Extrap.
## 3   41  1973-1974 16/01/1974 21:30     1.864      12.980 In Range
## 4   12  1974-1975 09/03/1975 14:00     3.287      43.470  Extrap.
## 5   45  1975-1976 26/09/1976 10:45     0.784       3.455 In Range
## 6    9  1976-1977 14/06/1977 19:30     3.368      46.780  Extrap.
##            Source Ref Comments
## 1 Digital Archive  2b       NA
## 2 Digital Archive  2d       NA
## 3 Digital Archive  2c       NA
## 4 Digital Archive  2d       NA
## 5 Digital Archive  2c       NA
## 6 Digital Archive  2d       NA
```



Plot of the AMAX series which is used to fit the GEV distribution

The plot shows the AMAX series contained in the `Input/Other catchments/Shipston/54106 Stour at Shipston AMAX.csv` dataset.

1.4 Fit GEV to AMAX series

The Package extRemes is used (Eric Gilleland, Richard W. Katz (2016). extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1-39. doi:10.18637/jss.v072.i08). The estimated shape parameter < 0, indicating a Weibull distribution.

```
library(extRemes)

fit<-fevd(AMAX_data$Flow..m3.s.,type="GEV")
fit
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Estimation Method used: MLE"
##
##  Negative Log-Likelihood Value:  205.6712
##
##  Estimated parameters:
## location    scale    shape
## 22.12814 14.88673  0.17633
##
##  Standard Error Estimates:
```

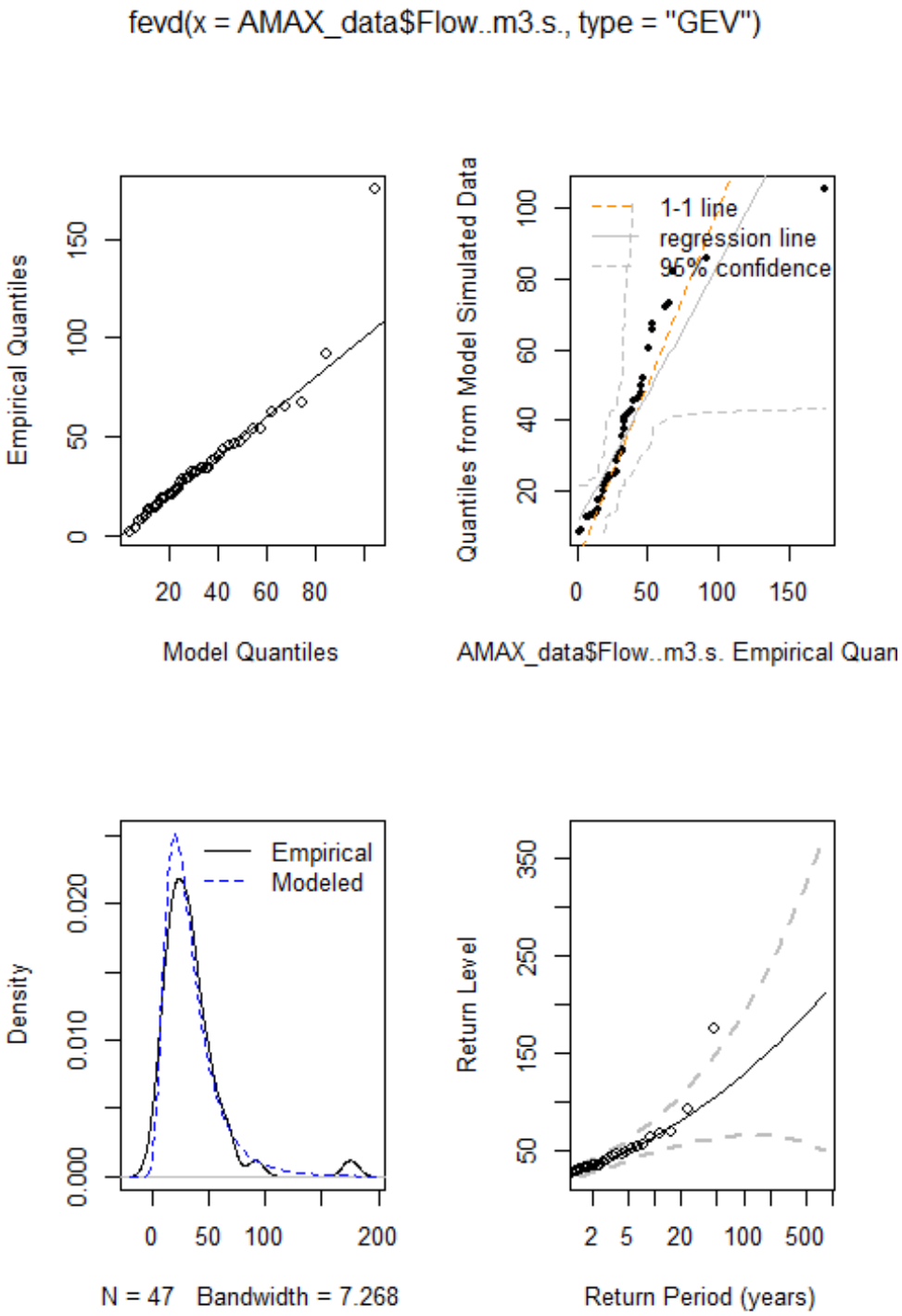
```
##  location     scale     shape
## 2.4334090 1.8928589 0.1071306
##
##  Estimated parameter covariance matrix.
##              location        scale        shape
## location  5.92147918  2.471555682 -0.073617874
## scale     2.47155568  3.582914874 -0.008288524
## shape    -0.07361787 -0.008288524  0.011476975
##
##  AIC = 417.3423
##  BIC = 422.8928
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI Estimate 95% upper CI
## location  17.35874566 22.12814   26.8975335
## scale     11.17679574 14.88673   18.5966663
## shape     -0.03364222  0.17633    0.3863022
```

Plotting the diagnostic plots shows a reasonable fit is achieved.

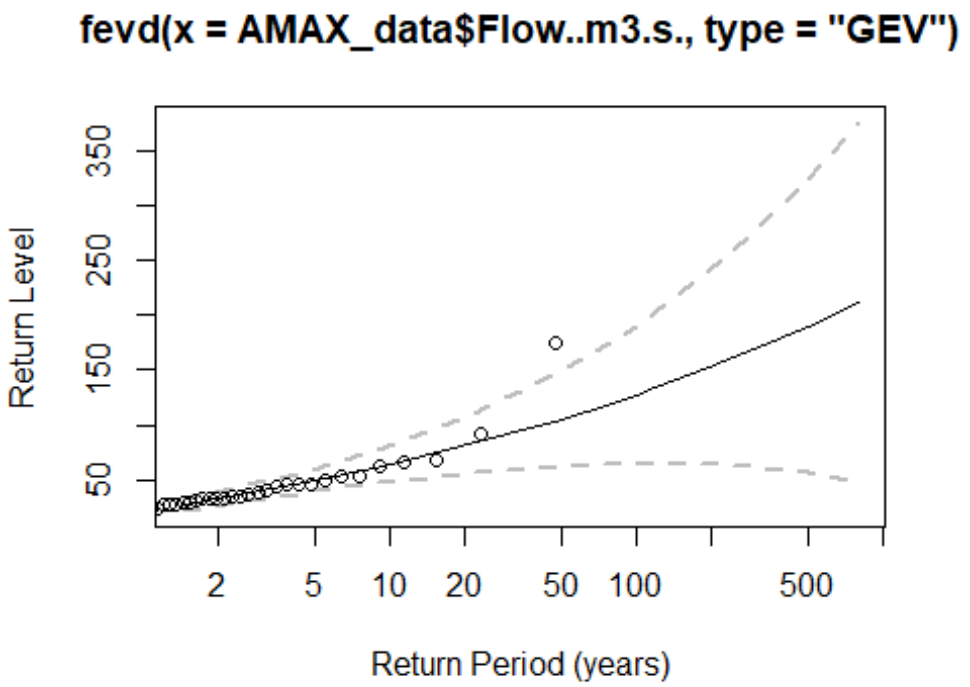
```
plot(fit)
```



AMAX

Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit,type="rl")
```




```
return.level(fit,do.ci=TRUE)

## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##                         95% lower CI  Estimate 95% upper CI
## 2-year return level        22.16610  27.76449     33.36288
## 20-year return level       55.37738  80.23991    105.10244
## 100-year return level      65.16271 127.70056    190.23841
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`43.509`.

```
level_from_GEV_fun<-function(fit,level){
  #Get parameters
  u<-fit$results$`par`[["location"]]
  d<-fit$results$`par`[["scale"]]
  e<-fit$results$`par`[["shape"]]
  z<-level

  #calculate from Gumbel distribution equation
  a<-e*((z-u)/d)#intermediate
  b<-(1+a)^(-1/e)#intermediate
  P<-exp(-b)# Probability

  return(P) #Return the probability
```

```
}

#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit)
P

## [1] 0.7573151

AEP<-1-P
AEP

## [1] 0.2426849

RP<-1/AEP #return period (years)
RP

## [1] 4.12057
```

For small return period events (<5 years) the distribution obtained using the AMAX and POT series are different (see http://evidence.environment-agency.gov.uk/FCERM/en/FluvialDesignGuide/Chapter2.aspx?pagenum=4). Hence, Fit a Generalised Pareto Distribution to the POT data instead.

1.5 Fit distribution to POT data

##POT Data Data file: `Input/Other catchments/Shipston/54106 Stour at Shipston POT.csv`

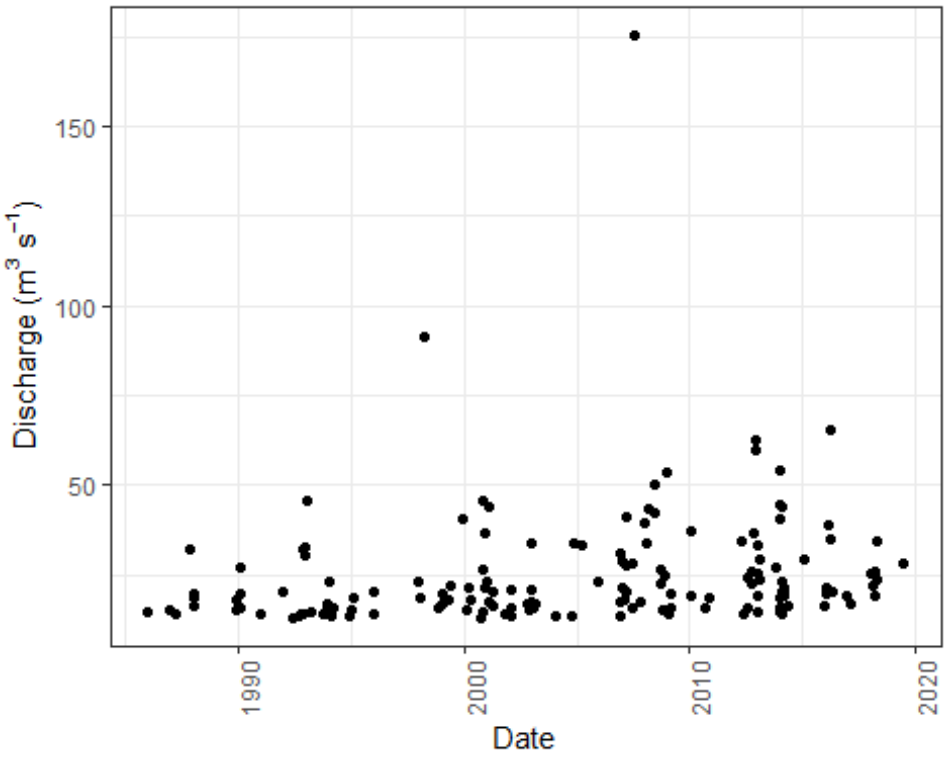
```
#setwd("D:/OneDrive - University of Leeds/Fragility_Curves")
POT_data<-read.csv(params$filename_POT)
POT_data$Date<-as.Date(POT_data$Date, format="%d/%m/%Y")

head(POT_data)

##   Rank       Date  Time Stage..m. Flow..m3.s.   Rating          Source Ref
## 1  137 1986-01-10 15:00     2.022      14.312 In Range Digital Archive  2c
## 2  124 1986-12-15 18:45     2.142      15.295 In Range Digital Archive  2c
## 3  144 1987-04-05 02:30     1.979      13.959 In Range Digital Archive  2c
## 4  148 1987-04-07 19:45     1.962      13.819 In Range Digital Archive  2c
## 5   34 1987-11-19 22:15     2.965      31.836  Extrap. Digital Archive  2d
## 6   80 1988-01-06 19:45     2.535      19.816 In Range Digital Archive  2c
##   Comments
## 1       NA
## 2       NA
## 3       NA
## 4       NA
## 5       NA
## 6       NA

summary(POT_data$Flow..m3.s.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.12   15.69   20.07   25.02   28.64  175.33
```



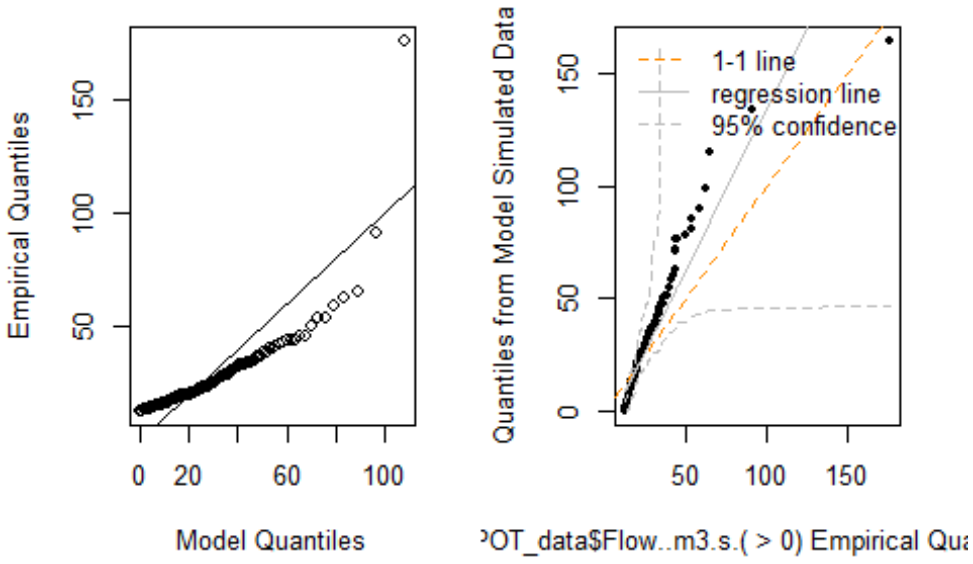
First, try to fit a Generalised-Pareto Distribution

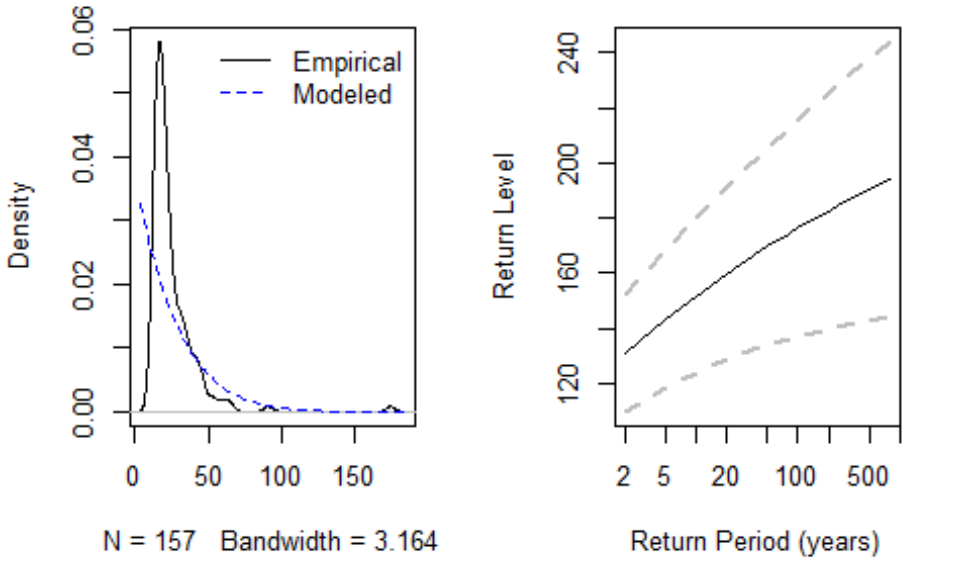
```
#estimate number of years in POT series
library(zoo)
years<-as.yearmon(max(POT_data$Date))-as.yearmon(min(POT_data$Date))

#Terrible fit with the GP
fit_POT_GP<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GP",span=years)

plot(fit_POT_GP)
```

vd(x = POT_data$Flow..m3.s., threshold = 0, type = "GP", span = year





POT GP Diagnostic plots for estimated parameter fit

GP Distribution does not fit the data, hence the GEV distribution is used again.

```
#Try GEV instead
fit_POT_GEV<-fevd(POT_data$Flow..m3.s.,threshold=0,type="GEV",span=years)
fit_POT_GEV
```

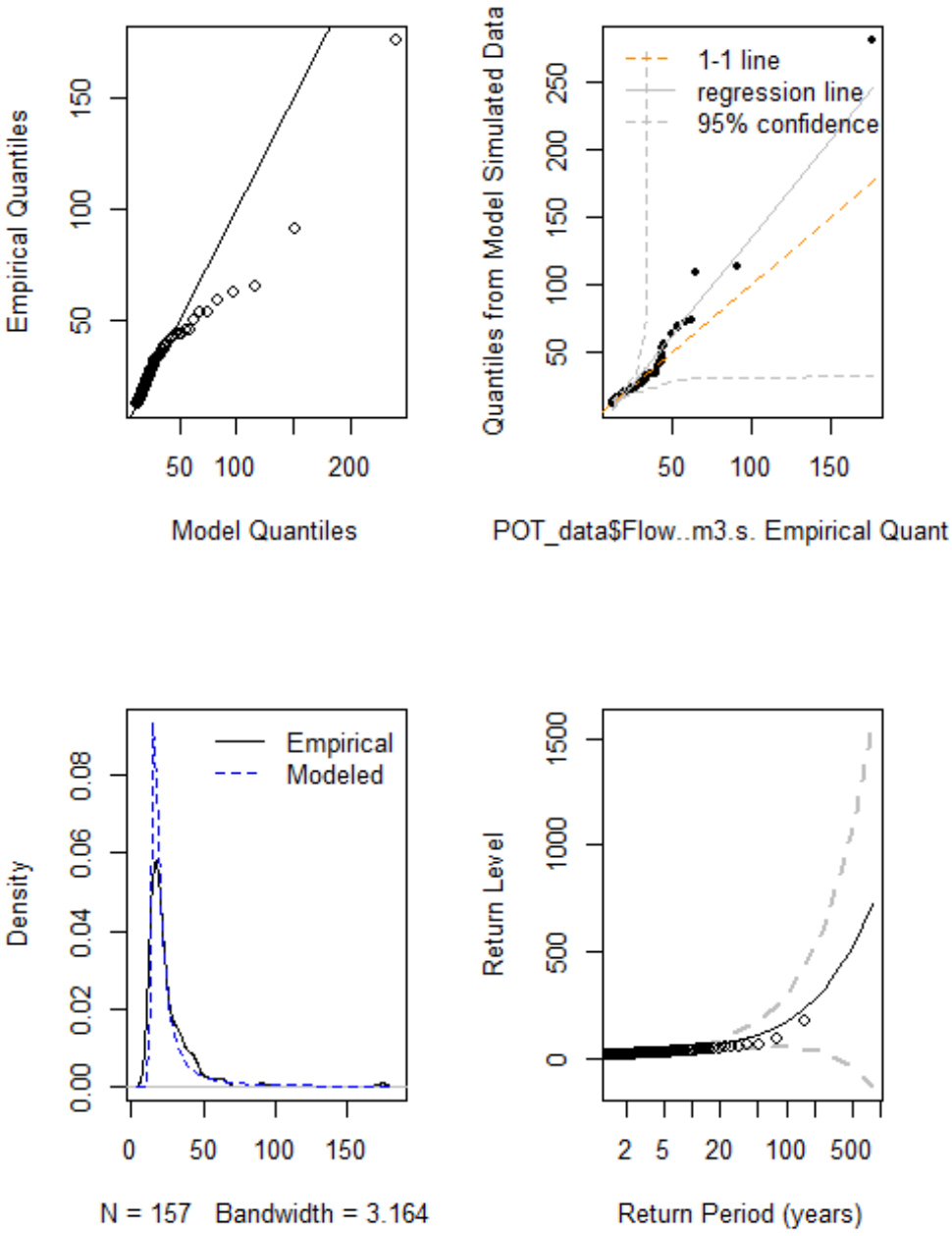
```
##
## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  549.1664
##
##
##  Estimated parameters:
##   location      scale      shape
## 17.2606931  4.6584833  0.6983897
##
##  Standard Error Estimates:
##  location     scale     shape
## 0.4606097 0.4910527 0.1146866
##
##  Estimated parameter covariance matrix.
##             location        scale        shape
## location  0.21216128  0.186217113 -0.020025740
## scale     0.18621711  0.241132719 -0.001890565
## shape    -0.02002574 -0.001890565  0.013153005
##
##  AIC = 1104.333
##
##  BIC = 1113.502
```

```
ci(fit,type="parameter")
```

```
## fevd(x = AMAX_data$Flow..m3.s., type = "GEV")
##
## [1] "Normal Approx."
##
##          95% lower CI Estimate 95% upper CI
## location  17.35874566 22.12814   26.8975335
## scale     11.17679574 14.88673   18.5966663
## shape     -0.03364222  0.17633    0.3863022
```

```
plot(fit_POT_GEV)
```

d(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = yea

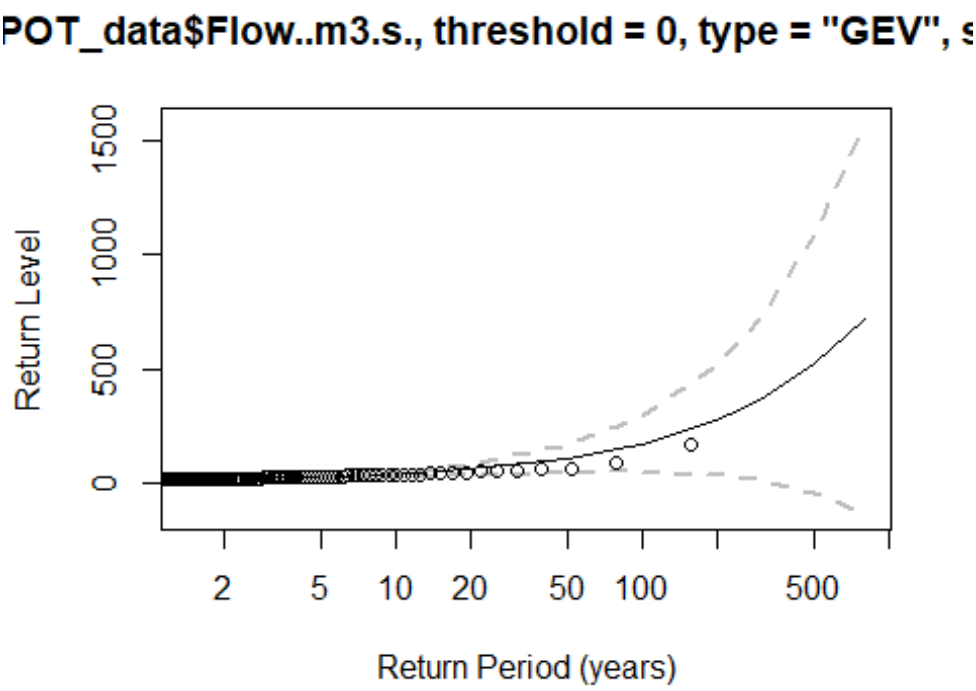


POT GEV Diagnostic plots for estimated parameter fit

The GEV Distribution produces a reasonable fit at lower return periods.The estimated shape parameter is >0, indicating a Frechet distribution is likely.

##POT Results The parameter fit and 95% confidence intervals shown in the return level plot are used to calculate the AEP and RP for the levels of interest.

```
plot(fit_POT_GEV,type="rl")
```




```
return.level(fit_POT_GEV,do.ci=TRUE)

## fevd(x = POT_data$Flow..m3.s., threshold = 0, type = "GEV", span = years)
## [1] "Normal Approx."
##                         95% lower CI  Estimate 95% upper CI
## 2-year return level        17.97292  19.20651     20.44010
## 20-year return level       41.31508  63.68213     86.04919
## 100-year return level      52.93414 176.32118    299.70822
```

The GEV distribution equation is used to calculate the AEP and RP for the required return levels:`43.509`.

```
#Apply the function to the levels of interest
P<-sapply(params$levels,level_from_GEV_fun,fit=fit_POT_GEV)
P

## [1] 0.903306

AEP<-1-P
AEP

## [1] 0.096694

RP<-1/AEP #return period (years)
RP

## [1] 10.3419
```