



Department  
for Environment  
Food & Rural Affairs



Llywodraeth Cymru  
Welsh Government



Cyfoeth  
Naturiol  
Cymru  
Natural  
Resources  
Wales



Environment  
Agency

# delivering benefits through evidence

source

pathway

receptor

Spatial joint probability for FCRM and  
national risk assessment multivariate  
event modeller - user guide

Report – SC140002/R3

We are the Environment Agency. We protect and improve the environment.

Acting to reduce the impacts of a changing climate on people and wildlife is at the heart of everything we do.

We reduce the risks to people, properties and businesses from flooding and coastal erosion.

We protect and improve the quality of water, making sure there is enough for people, businesses, agriculture and the environment. Our work helps to ensure people can enjoy the water environment through angling and navigation.

We look after land quality, promote sustainable land management and help protect and enhance wildlife habitats. And we work closely with businesses to help them comply with environmental regulations.

We can't do this alone. We work with government, local councils, businesses, civil society groups and communities to make our environment a better place for people and wildlife.

**Published by:**

Environment Agency, Horizon House, Deanery Road,  
Bristol, BS1 9AH  
[www.environment-agency.gov.uk](http://www.environment-agency.gov.uk)

ISBN: 978-1-84911-413-4

© Environment Agency – April 2018

All rights reserved. This document may be reproduced with prior permission of the Environment Agency.

Email: [fcerm.evidence@environment-agency.gov.uk](mailto:fcerm.evidence@environment-agency.gov.uk)

Further copies of this report are available from our publications catalogue:  
<http://www.gov.uk/government/publications>

or our National Customer Contact Centre:  
T: 03708 506506

Email: [enquiries@environment-agency.gov.uk](mailto:enquiries@environment-agency.gov.uk)

**Author(s):**

Neil Hunter, Rob Lamb, Ross Towe, Sarah Warren,  
Elizabeth Wood

**Dissemination Status:**

Publicly available

**Keywords:**

Multivariate analysis, Joint probability, Event set, Scenarios, R, RStudio

**Research Contractor:**

JBA Consulting, South Barn, Broughton Hall, Skipton,  
North Yorkshire BD23 3AE  
Tel: +44 (0) 1756 799919

**Environment Agency's Project Executive:**

Stefan Laeger, National Modelling and Forecasting

**Theme and Project Manager:**

Sue Manson, Incident Management & Risk Modelling Theme

**Collaborator(s):**

Lancaster University, HR Wallingford

**Project Number:**

SC140002/R3

# Evidence at the Environment Agency

Scientific research and analysis underpins everything the Environment Agency does. It helps us to understand and manage the environment effectively. Our own experts work with leading scientific organisations, universities and other parts of the Defra group to bring the best knowledge to bear on the environmental problems that we face now and in the future. Our scientific work is published as summaries and reports, freely available to all.

This report is the result of research commissioned and funded by the Joint Flood and Coastal Erosion Risk Management Research and Development Programme. The Joint Programme is jointly overseen by Defra, the Environment Agency, Natural Resources Wales and the Welsh Government on behalf of all Risk Management Authorities in England and Wales:

<http://evidence.environment-agency.gov.uk/FCERM/en/Default/FCRM.aspx>.

You can find out more about our current science programmes at:

<https://www.gov.uk/government/organisations/environment-agency/about/research>.

If you have any comments or questions about this report or the Environment Agency's other scientific work, please contact [research@environment-agency.gov.uk](mailto:research@environment-agency.gov.uk).

Professor Doug Wilson  
**Director, Research, Analysis and Evaluation**

# Executive summary

A realistic assessment of flood risk includes understanding the potential for harm stemming from multiple hazards, such as combined inland and coastal flooding, or flooding in multiple locations, such as in recent regionally and nationally significant events. A quantitative assessment of the risk of these types of events requires joint probability analysis, which has been applied in the latest flooding scenarios developed within the Cabinet Office's National Risk Register (NRR) process, bringing together the latest assessments of hazards and threats to increase national resilience and preparedness.

The development of widespread flood risk scenarios considered within the NRR involved the application of new statistical methods using complex and specialist software. Working in close partnership with Defra and supply chain partners, those advances in statistical methodology have been translated into a practical tool – the Multivariate Event Modeller (MEM) – for the analysis of joint probability within flood risk management. The MEM is designed to be applied more readily by trained practitioners for joint probability analysis in operational flood risk management, with case studies and supporting guidance.

The MEM is a demonstration tool that allows practitioners to start analysing spatial or multivariate joint probability problems in more than 3 variables and combinations of up to 10 variables. It provides an alternative to use alongside existing methods, and does not supersede or take priority over other methods.

This user guide places the MEM within the context of existing good practice in joint probability analysis, listing some possible uses within flood and coastal risk management while noting some caveats and limitations to be borne in mind.

Instructions are given on how to install the required software and how to run the MEM tool. Guidance is provided on the process of using the tool and what each of the 6 tabs within the tool does. Numerous screenshots (some annotated) are included.

Five case studies offer examples of how the MEM tool can be used in a range of environmental applications.

The guide ends with a series of FAQs and a glossary.

## Disclaimer

Please note that this research tool is still in beta testing phase, as such neither the Environment Agency nor its contractors take any responsibility or liability for any outputs from this tool, or any decisions any user takes based on outputs obtained from using this tool. This is not a tool suitable for normal operational use without the use of other supporting evidence / specialist expertise at the present time.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	About the multivariate event modeller	1
1.2	Use of the MEM for joint probability analysis	2
1.3	Structure of the user guide	5
<b>2</b>	<b>Software installation</b>	<b>6</b>
2.1	Software requirements	6
2.2	Installation procedure	6
2.3	Opening the MEM	11
<b>3</b>	<b>Tool overview</b>	<b>13</b>
3.1	Overview tab	14
3.2	Input data tab	16
3.3	Build model tab	22
3.4	Marginal analysis tab	28
3.5	Joint probability analysis tab	32
3.6	Export tab	36
<b>4</b>	<b>Case study applications</b>	<b>38</b>
4.1	Case study 1: Validation based on a simulated dataset	38
4.2	Case study 2: Joint probability of 2 tributaries at Stoke Canon (hydrological case study)	48
4.3	Case study 3: Joint probability of Storm Desmond on the River Tyne (hydrological case study)	52
4.4	Case study 4: Joint probability of high flow on the River Ellen and high tidal surges at Workington (estuary case study)	61
4.5	Case study 5: Demonstrating a spatial joint analysis of offshore extremes for the north Wales coast	66
<b>5</b>	<b>FAQs</b>	<b>70</b>
	<b>References</b>	<b>72</b>
	<b>List of abbreviations</b>	<b>73</b>
	<b>Glossary</b>	<b>74</b>
Table 1.1	Comparison of MEM and older methods used for joint probability analysis	3
Table 1.2	Where the MEM could be used	4
Table 3.1	Example format for time series data sampling all observations	18
Table 3.2	Example format for observations sampled at peak surge events	19
Table 3.3	Example runtimes for different numbers of variables and years of simulated events	23
Table 3.4	Example head of a simulated event set	36
Table 4.1	Potential combinations of tributary flows that would have a joint AEP of 2% or of 1%	51
Table 4.2	Maximum observed flow values for Storm Desmond at each gauge	54
Table 4.3	Comparison of the median annual maximum flow values (QMED) with the 50% annual chance flow values from fitting a generalised Pareto distribution to the data (automatically done in the MEM) and	

	from fitting a generalised logistic distribution to the data (using the user-supplied marginal parameters)	55
Table 4.4	Joint probability for a range of combinations of AEPs	66
Table 4.5	Values associated with the maximum recorded skew surge event at both locations (JP26 and JP27)	68
Figure 2.1	Flow chart providing an overview of the installation procedure	6
Figure 2.2	Step 1 to install R for Windows	7
Figure 2.3	Step 2 to install R for Windows	7
Figure 2.4	Step 3 to install R for Windows (previous releases are also available from this page)	7
Figure 2.5	R components to install (default setting)	8
Figure 2.6	First step to install RStudio for Windows	8
Figure 2.7	Opening RStudio without a shortcut	9
Figure 2.8	RStudio on starting the application	9
Figure 2.9	First step to install Rtools	10
Figure 2.10	Components that should be selected when installing Rtools (for 64-bit system)	10
Figure 2.11	How to set the working directory (folder where the packages are saved) in RStudio	11
Figure 2.12	How to open the MEM tool in RStudio	12
Figure 3.1	Tool workflow	13
Figure 3.2	'Overview' tab	15
Figure 3.3	Initial view of 'Input data' tab if you have selected to create a new joint probability model	17
Figure 3.4	'Input data' tab after time series data sampling all observations (or a saved model which used this type of data) have been uploaded	20
Figure 3.5	'Input data' tab after observations sampled at peak surge events (or a saved model which used this type of data) have been uploaded	21
Figure 3.6	Initial view of the 'Build model' tab before creating a dependence structure	25
Figure 3.7	'Build model' tab once the dependence structure has been created	26
Figure 3.8	'Build model' tab once the event set has been simulated, or initial view of the tab if you have loaded a saved model	27
Figure 3.9	Initial view of 'Marginal analysis' tab	30
Figure 3.10	'Marginal analysis' tab if 'Input and use your own marginal parameters' is selected	31
Figure 3.11	Initial view of the 'Joint probability analysis' tab with data plotted on its original measurement scale	34
Figure 3.12	'Joint probability analysis' tab with data plotted on the standardised scale and the joint probability calculated for an event where each variable is set to a 1% chance	35
Figure 3.13	'Export results' tab	37
Figure 4.1	Scatter plot of the simulated AI and AD dataset	39
Figure 4.2	Joint probability results for the validation case study where a 1% chance has been applied to all 5 AD variables	42
Figure 4.3	Joint probability results for the validation case study where a 1% chance has been applied to all five AI variables	43
Figure 4.4	Joint probability results for the validation case study where a 1% chance has been applied to all 10 variables in the validation dataset	44
Figure 4.5	Joint probability results for the validation case study where a 1% chance has been applied to one of the AD variables and one of the AI variables	45
Figure 4.6	Joint probability results for the validation case study where a 1% chance has been applied to 2 AD variables	46
Figure 4.7	Joint probability results for the validation case study where a 1% chance has been applied to two AI variables	47
Figure 4.8	Location of flow gauges, with Stoke Canon visible towards the south-west of the inset map just north of Exeter	48
Figure 4.9	Relationship between daily mean flow at Wood Mill and Thorverton	49
Figure 4.10	AEP and associated values (calculated within the MEM) in the table on the 'Joint probability analysis' tab	50
Figure 4.11	Time series of daily mean data for Wood Mill and Thorverton	50
Figure 4.12	Location of the 3 gauging stations in the Tyne catchment for which the joint probability was calculated	53
Figure 4.13	Time series of daily maximum data for Bywell, Reaverhill and Haydon Bridge	54
Figure 4.14	AEP and associated values in the table on the 'Joint probability analysis' tab	55
Figure 4.15	Joint probability analysis for the peak flows on 5 December 2015 using the automatically determined marginal parameters	57
Figure 4.16	Joint probability analysis for the peak flows on the 5 <sup>th</sup> December using the user-supplied marginal parameters	58
Figure 4.17	Joint probability analysis for the peak flows on 6 December 2-15 using the automatically determined marginal parameters	59
Figure 4.18	Joint probability analysis for the peak flows on 6 December 2015 using the user-supplied marginal parameters	60
Figure 4.19	Location of the Bullgill flow gauging station and the Workington tide gauge with Maryport situated at the mouth of the River Ellen	61
Figure 4.20	Relationship between daily maximum fluvial flow at Bullgill and tide surge residual at Workington	62
Figure 4.21	AEP and associated values (calculated within the MEM) in the table on the 'Joint probability analysis' tab	63
Figure 4.22	Time series of daily maximum data for Bullgill and Workington	63
Figure 4.23	Joint probability analysis for the scenario where both gauges have an AEP of 10%	65
Figure 4.24	Location of JP26 and JP27	67
Figure 4.25	Joint probability analysis for the highest recorded skew surges at JP26 and JP27	69

# 1 Introduction

A realistic assessment of flood risk includes understanding the potential for harm stemming from multiple hazards, such as combined inland and coastal flooding, or flooding in multiple locations, such as in recent regionally and nationally significant events. A quantitative assessment of the risk of these types of events requires joint probability analysis, which has been applied in the latest flooding scenarios developed within the Cabinet Office's National Risk Assessment (NRA) process, bringing together the latest assessments of hazards and threats to increase national resilience and preparedness.

The joint probability of 2 or more variables being 'extreme' is relevant in flood and coastal risk management (FCRM) in various contexts, including

- assessing the likelihood of extreme peak flow events on multiple tributaries of a river to help in developing scenarios for a whole catchment model
- placing recent or historical floods in context by estimating the combined likelihood of extreme flows, water levels, rainfall, wave or wind observations at one or more locations
- assessing the likelihood of combinations of extreme river flows, storm surge and possibly other relevant variables
- modelling the chance of combinations of extreme conditions occurring together in related variables such as soil moisture content, rainfall accumulations and river flows

Previous Defra and Environment Agency guidance (Hawkes 2005) describes methods for joint probability calculations for certain pairs of variables, based on statistical models supported by specialist software (HR Wallingford and Lancaster University 2000).

More recently, new statistical methods have been developed that allow for a more general, data-driven analysis of the joint probability of extreme events in combinations of multiple variables (Heffernan and Tawn 2004). Applications include:

- regional and national scale assessments of the probability of widespread flooding in rivers (Lamb et al. 2010)
- joint probability analysis of extreme surge and waves at the coast (Gouldby et al. 2014)

These methods have been documented and tested through a series of research reports from a previous Defra and Environment Agency Flood and Coastal Erosion Risk Management project SC060088 'Risk of widespread flooding (spatial coherence)<sup>1</sup>.

## 1.1 About the Multivariate Event Modeller

The Multivariate Event Modeller (MEM) is a tool that implements the new methods with user-supplied datasets to estimate the joint probability of extreme events in combinations of up to 10 variables. The tool is designed for joint probability analysis of

---

<sup>1</sup> Available from <https://www.gov.uk/government/publications/the-risk-of-widespread-flooding-capturing-spatial-patterns-in-flood-risk-from-rivers-and-coasts>

extremes in time series data. Additionally, it handles a special case where data have been sampled for extreme surge events to illustrate the extension of coastal joint probability methods to incorporate spatial dependence.

The MEM is a demonstration tool that allows practitioners to start analysing spatial or multivariate joint probability problems in more than 3 variables. It provides an alternative to use alongside existing methods, and does not supersede or take priority over other methods.

It is a specialist tool aimed at those interested in joint probability assessment including hydrologists, hydraulic modellers and coastal engineers. It has the following applications.

- Extreme value analysis (any variables including fluvial, rainfall and coastal) – annual exceedance probability curves can be viewed on the ‘Marginal analysis’ tab for each variable.
- It can be used to understand and view the relationships between the largest values of a combination of variables in space and time.
- It can be used to calculate the dependence parameter Chi for a pair of variables, which can then be used in the joint probability desk study approach described in the Defra/Environment Agency technical report FD2308/TR2 (Hawkes 2005).
- It can be used to find the probability of an observed or hypothetical extreme event at multiple locations and/or in multiple variables (for example, to quantify the severity of an observed or a forecasted event).
- It can be used to find the encounter probabilities of an observed or hypothetical extreme event at multiple locations and/or in multiple variables (for example, the chance of observing a particular event over the lifetime of a mortgage).
- The simulated event set can be exported for use as an input to models. For example, it can be used in a catastrophe model or a SWAN model to analyse the damage/impact of an event or hydraulic models to test flood resilience measures using a variety of plausible extreme events.

## 1.2 Use of the MEM for joint probability analysis

Table 1.1 places the MEM within the context of existing good practice in joint probability analysis.

Table 1.2 lists in greater detail some of the uses within FCRM in which it is envisaged that the MEM could be applied. It also includes caveats and limitations to be borne in mind.



**Table 1.1 Comparison of MEM and older methods used for joint probability analysis**

<b>Method</b>	<b>Year established in practice</b>	<b>User skill-level</b>	<b>Potential types of application</b>	<b>Application track record</b>	<b>Characteristics</b>
Simplified approach (contour)	2005	Low (non-specialists)	Introductory 'Broad brush' analysis	Widely used in practice Quick No special software/hardware needed	Approximate Limited number of variables
JOIN-SEA (contour)	1998	Specialist	Scheme design  Flood risk analysis at different spatial scales and levels of complexity	Widely used in practice  Relatively quick	Robust software and method  Three variables Fixed dependence structure (no 'black swans')
JOIN-SEA (risk-based)	1998	Specialist	As above	Widely used in practice	As above plus impacts analysis
Heffernan and Tawn (2004) (contour)	2010	Specialist	NRA scenarios  Scheme design  Tributary inflows analysis  Flood risk analysis at different scales and complexities including high-dimensional problems	Emerging methods  Implemented in bespoke software  May require significant computer power and data storage	Theoretically robust for all types of dependence and over large ranges of spatial scale  Works for arbitrary combinations of variables  Can support time dependence
Heffernan and Tawn (2004) (risk-based)	2008	Specialist	Re/insurance catastrophe models  Scheme design  Flood risk analysis at	Well-established in insurance risk modelling  Not yet applied in	As above plus impacts analysis

Method	Year established in practice	User skill-level	Potential types of application	Application track record	Characteristics
			different spatial scales and complexities	FCRM engineering	
Heffernan and Tawn (2004) (MEM)	2017	Moderate	Supporting analysis for FCRM planning, scenario analysis and post-event analytics	New	As above for up to 10 variables New software No built-in treatment of time dependence

Notes: 'Contour' refers to the method of joint probability analysis based on the construction of joint probability contours, or design curves, relating to the chance of exceeding specified combinations of values in physical variables.  
'Risk-based' refers to methods in which the probability of a defined range of outcomes (such as increasing amounts of monetary damage) is assessed, based on the analysis of 2 or more input variables.

**Table 1.2 Where the MEM could be used**

#### What the MEM is intended for

- **General multivariate analysis** – can handle between 2 and 10 variables simultaneously, and variables can represent any concurrently sampled quantities (including river flow, water levels, rainfall, wind speed or surge). The method accounts for complex dependencies between variables, as observed in many hydrological records.
- **Visualising data** – interactive plots that allow the user to investigate the data easily.
- **Insight** – the MEM tool offers supporting evidence about dependence and joint probabilities (for example, the user can visualise how variables relate to each other in the extremes, and can understand and view the relationships between the largest values of a combination of variables).
- **Support for the project FD2308 joint probability methods** – calculates the dependence parameter Chi for any pair of variables, for use in the FD2308 joint probability 'desk study' approach. The derivation of Chi is based on the methods set out in Hawkes (2005).
- **Calibrating a joint probability model to match an existing analysis of annual flood probabilities** – the MEM can automatically fit a statistical distribution for the annual exceedance probabilities (AEPs) in each variable, using the peaks over threshold method. However, an externally derived model can be specified instead, defined in terms of a generalised logistic model for the AEPs. This enables the joint probability analysis to incorporate flood frequency curves derived using Flood Estimation Handbook methods.
- **Exporting simulated joint probability events** – the MEM can generate and export plausible concurrent extreme events using Monte Carlo simulation.

#### Applications where the MEM could provide supporting evidence, with interpretation

- **Probabilistic scenario generation** – the MEM outputs could be used as a basis for developing likelihood-based extreme event scenarios, although expert interpretation and judgement is recommended to ensure that outputs are used appropriately.
- **Support for setting inflows/boundary conditions for hydraulic model scenarios** – a possible use for the MEM is the development of extreme event scenarios for tributary inflows and/or estuaries in hydraulic models. It is necessary to give careful consideration to sampling uncertainties, the robustness of the input data and timing issues. It is recommended that the MEM be applied and compared with other methods, including the FD2308 joint probability method, local marginal analysis (for example, Flood Estimation Handbook methods) and historical precedents to ensure a robust analysis.

---

#### **Applications where the use of the MEM requires caution**

- **Detailed design calculations for individual locations** – more established and rigorously tested standard methods such as the Flood Estimation Handbook, the Revitalised Flood Hydrograph (ReFH) or Defra’s extreme sea level models should be used. However, the MEM may provide additional supporting evidence about spatial or multivariable joint events as discussed above.
- **Generation of events for coastal damage calculations** – the MEM implements a spatial extension of the joint probability analysis of storm surges, extreme wind and wave observations, but this is conditional only on the occurrence of an extreme surge, and therefore does not consider the full set of combinations of processes that could cause damage because of coastal flooding or waves.

---

## 1.3 Structure of the user guide

Section 2 provides instructions for installing the required software and running the MEM tool.

Section 3 guides you through the process of using the MEM tool and explains what each tab does.

Five case studies in Section 4 provide examples of how the MEM tool can be used in a range of environmental applications.

A series of FAQs is given in Section 5.

There is also a glossary of terms.

## 2 Software installation

This section explains how to install and run the MEM, providing an overview of the software requirements and how to install the various components.

### 2.1 Software requirements

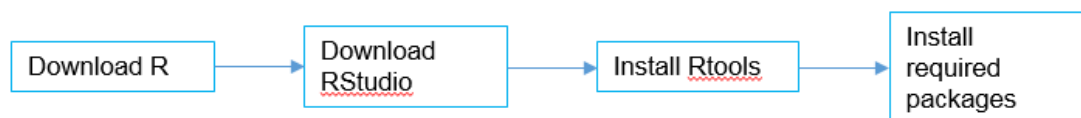
The following open source third-party components are required on the user's machine:

- R for Windows – the MEM has been tested on versions 3.2.0 to 3.3.3
- RStudio – the MEM has been tested on versions 0.99.467 to 1.0.136
- Rtools
- The R packages provided with the MEM tool (note that the MEM may not work if the versions of the R packages are different from those provided) – instructions for installing the packages and their dependencies are given in Section 2.2.4.

The MEM tool has been developed and tested using Microsoft® Windows 7 and has been verified to run on Windows 10.

### 2.2 Installation procedure

Figure 2.1 shows the order for installing the software.



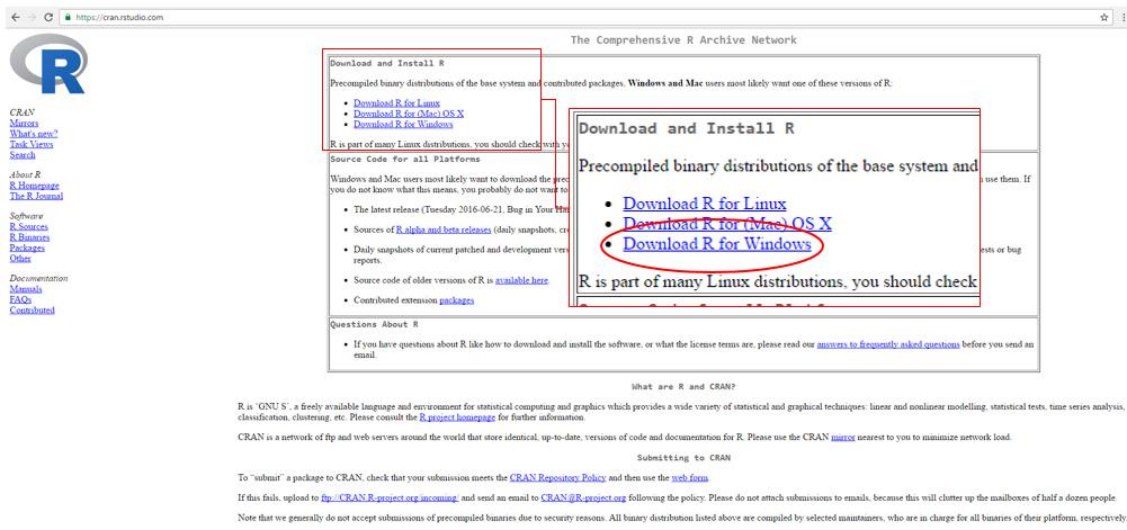
**Figure 2.1** Flow chart providing an overview of the installation procedure

#### 2.2.1 R for Windows

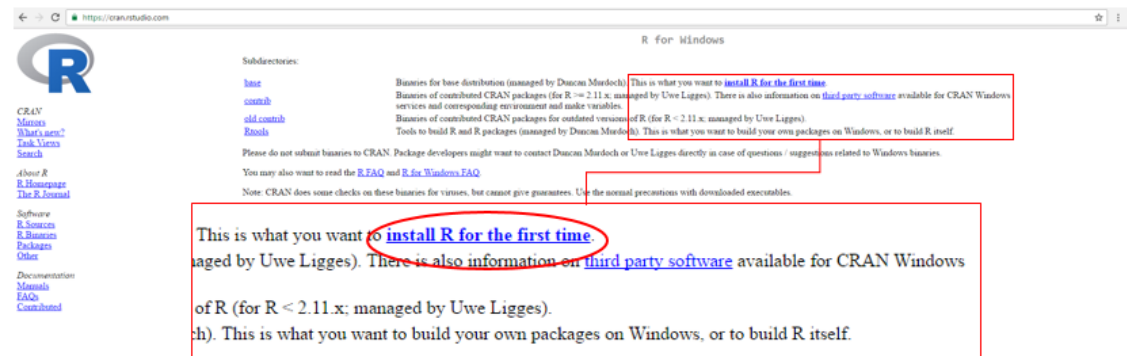
The latest version of R (and previous releases) can be downloaded from: <https://cran.rstudio.com/>

Make sure that you install the appropriate components for your hardware and operating system (that is, 64-bit components if installing on a 64-bit Windows system).

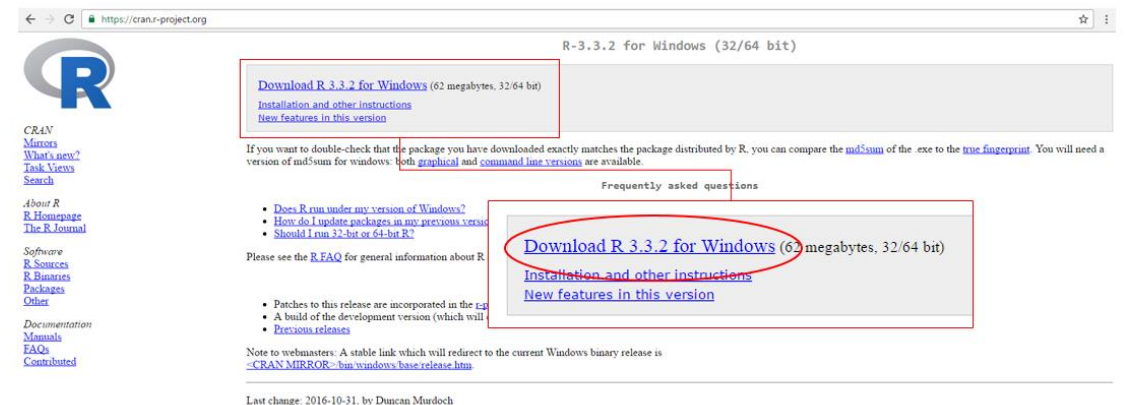
Figure 2.2, Figure 2.3 and Figure 2.4 show the first steps in the download. Click 'Download R for Windows' and then 'Install R for the first time'.



**Figure 2.2 Step 1 to install R for Windows**



**Figure 2.3 Step 2 to install R for Windows**



**Figure 2.4 Step 3 to install R for Windows (previous releases are also available from this page)**

Follow the installation procedure. You should be able to accept most of the default options, although you may wish to change the saving location.

Figure 2.5 shows the components to install on a 64-bit system. This should be the default setting.

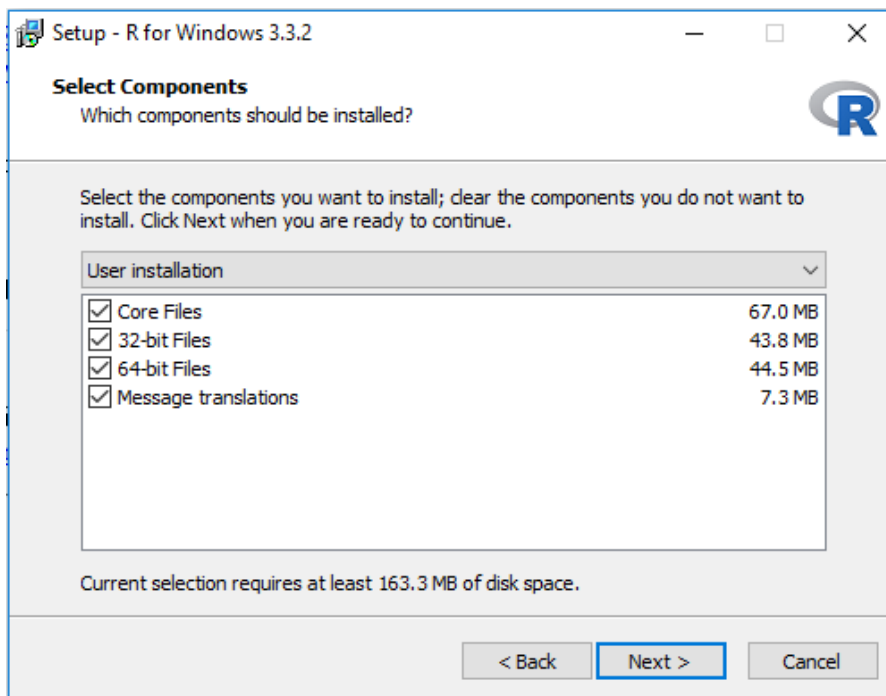


Figure 2.5 R components to install (default setting)

## 2.2.2 RStudio

The latest version of RStudio can be downloaded from:

<https://www.rstudio.com/products/rstudio/download/>

Select the Windows installer (Figure 2.6) and follow the installation process.

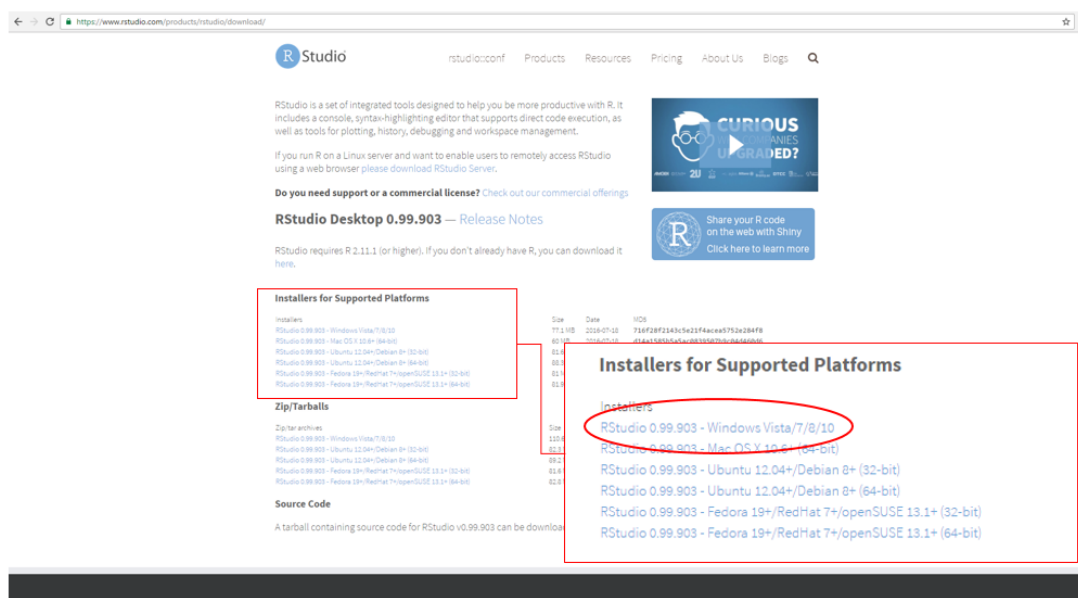


Figure 2.6 First step to install RStudio for Windows

Once installed, a shortcut should have been created which you can use to open RStudio. Otherwise you can find the application in the directory you selected for saving

RStudio under 'bin/rstudio' (Figure 2.7) (assuming the default settings have been followed).

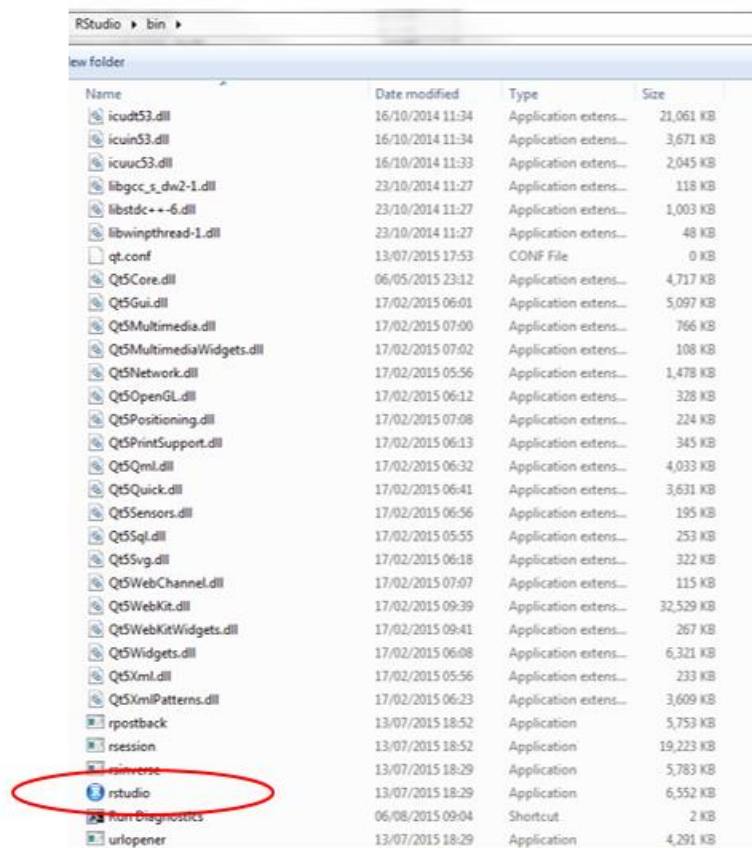


Figure 2.7 Opening RStudio without a shortcut

RStudio will look similar to Figure 2.8 when you open it, though there may also be a box in the top left quadrant.

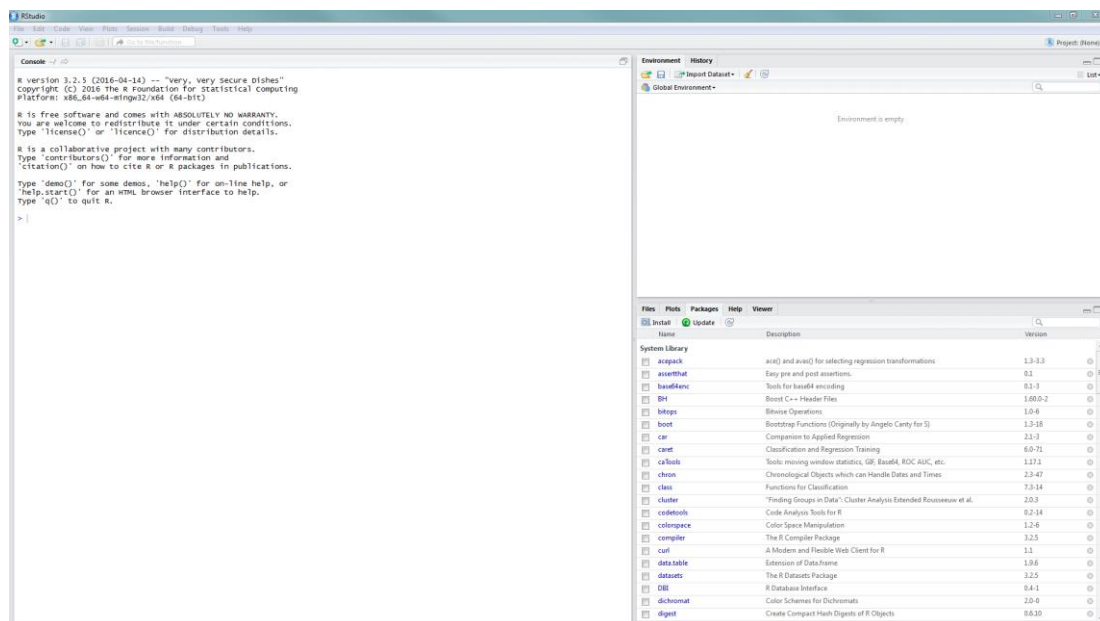


Figure 2.8 RStudio on starting the application

## 2.2.3 Rtools

Some of the packages need compiling when they are installed from the folder provided and this requires Rtools. This can be downloaded from:

<https://cran.r-project.org/bin/windows/Rtools/>

Select the version that is compatible with the version of R you have downloaded (Figure 2.9).

The current version of this file is recorded here: [VERSION.txt](#)

Download	R compatibility	Frozen?
<a href="#">Rtools34.exe</a>	R 3.3.x and later	No
<a href="#">Rtools33.exe</a>	R 3.2.x to 3.3.x	Yes
<a href="#">Rtools32.exe</a>	R 3.1.x to 3.2.x	Yes
<a href="#">Rtools31.exe</a>	R 3.0.x to 3.1.x	Yes
<a href="#">Rtools30.exe</a>	R 2.15.1 to R 2.9.x	Yes
<a href="#">Rtools215.exe</a>	R 2.14.1 to R 2.15.1	Yes
<a href="#">Rtools214.exe</a>	R 2.13.x or R 2.14.x	Yes
<a href="#">Rtools213.exe</a>	R 2.13.x	Yes
<a href="#">Rtools212.exe</a>	R 2.12.x	Yes
<a href="#">Rtools211.exe</a>	R 2.10.x or R 2.11.x	Yes
<a href="#">Rtools210.exe</a>	R 2.9.x or 2.1	
<a href="#">Rtools209.exe</a>	R 2.8.x or R	
<a href="#">Rtools208.exe</a>	R 2.7.x or R	
<a href="#">Rtools207.exe</a>	R 2.6.x or R	
<a href="#">Rtools206.exe</a>	R 2.6.x, R 2	

Download	R compatibility
<a href="#">Rtools34.exe</a>	R 3.3.x and later
<a href="#">Rtools33.exe</a>	R 3.2.x to 3.3.x
<a href="#">Rtools32.exe</a>	R 3.1.x to 3.2.x
<a href="#">Rtools31.exe</a>	R 3.0.x to 3.1.x

Figure 2.9 First step to install Rtools

Keep the default settings throughout the download, although you may wish to change the saving location. Figure 2.10 shows the components that should be selected for a 64-bit system.

Select Components  
Which components should be installed?

Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.

Package authoring installation

- R toolset 19.2 MB
- Cygwin DLLs 5.8 MB
- R 3.3.x+ 32 bit toolchain 401.8 MB
- R 3.3.x+ 64 bit toolchain 437.3 MB
- Extras to build 32 bit R: TCL/TK 12.8 MB
- Extras to build 64 bit R: TCL/TK 12.6 MB

Current selection requires at least 796.2 MB of disk space.

< Back Next > Cancel

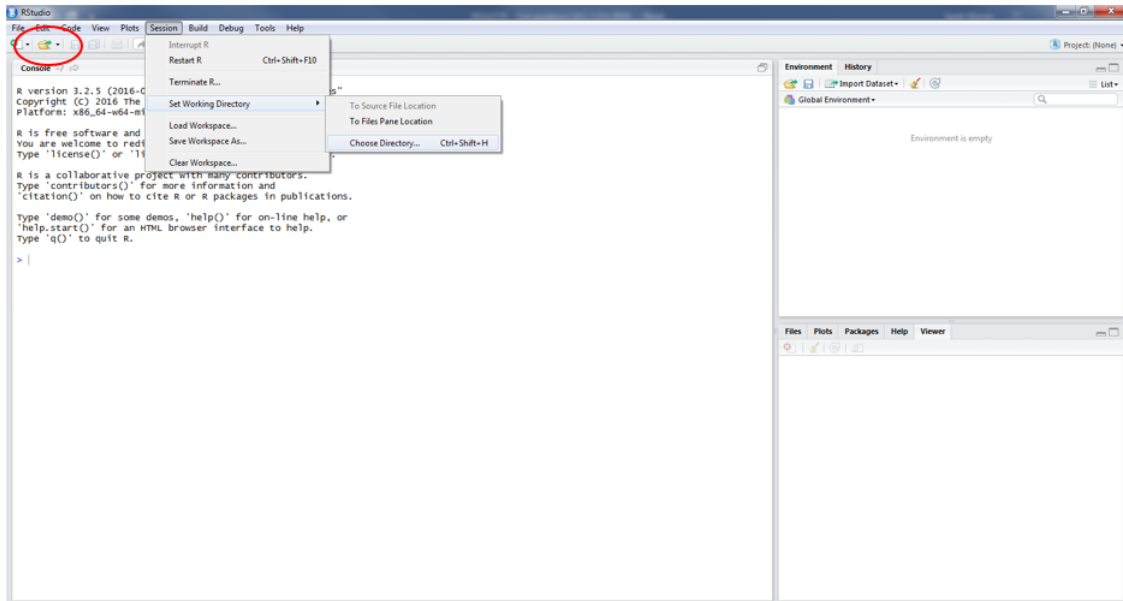
Figure 2.10 Components that should be selected when installing Rtools (for 64-bit system)



## 2.2.4 R packages

The required versions of the R packages are provided in the 'Packages' folder in the MEM tool and can be installed using the 'Install packages.R' script in the same folder. This can be opened from within RStudio by clicking the 'Open an existing file' button (circled in Figure 2.11) or using 'Ctrl+O'. You need to set the directory where the packages are saved (the 'Packages' folder) by navigating to the folder (see Figure 2.11). Instructions for running the rest of the code are given in the script.

These packages only need installing once per installation of R. It is recommended that they are installed on a new installation of R to ensure that there are no conflicts with packages that may have been installed previously.

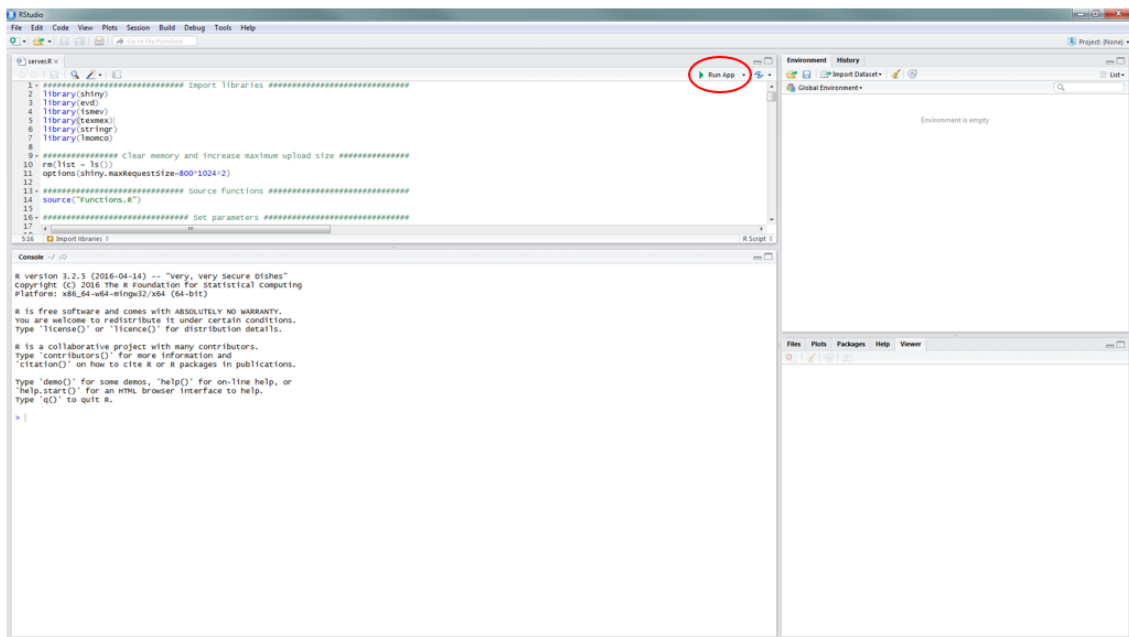


**Figure 2.11** How to set the working directory (folder where the packages are saved) in RStudio

## 2.3 Opening the MEM

The MEM tool is run from within RStudio. A blank workspace is recommended (that is, no other code has been run or data imported before running the tool).

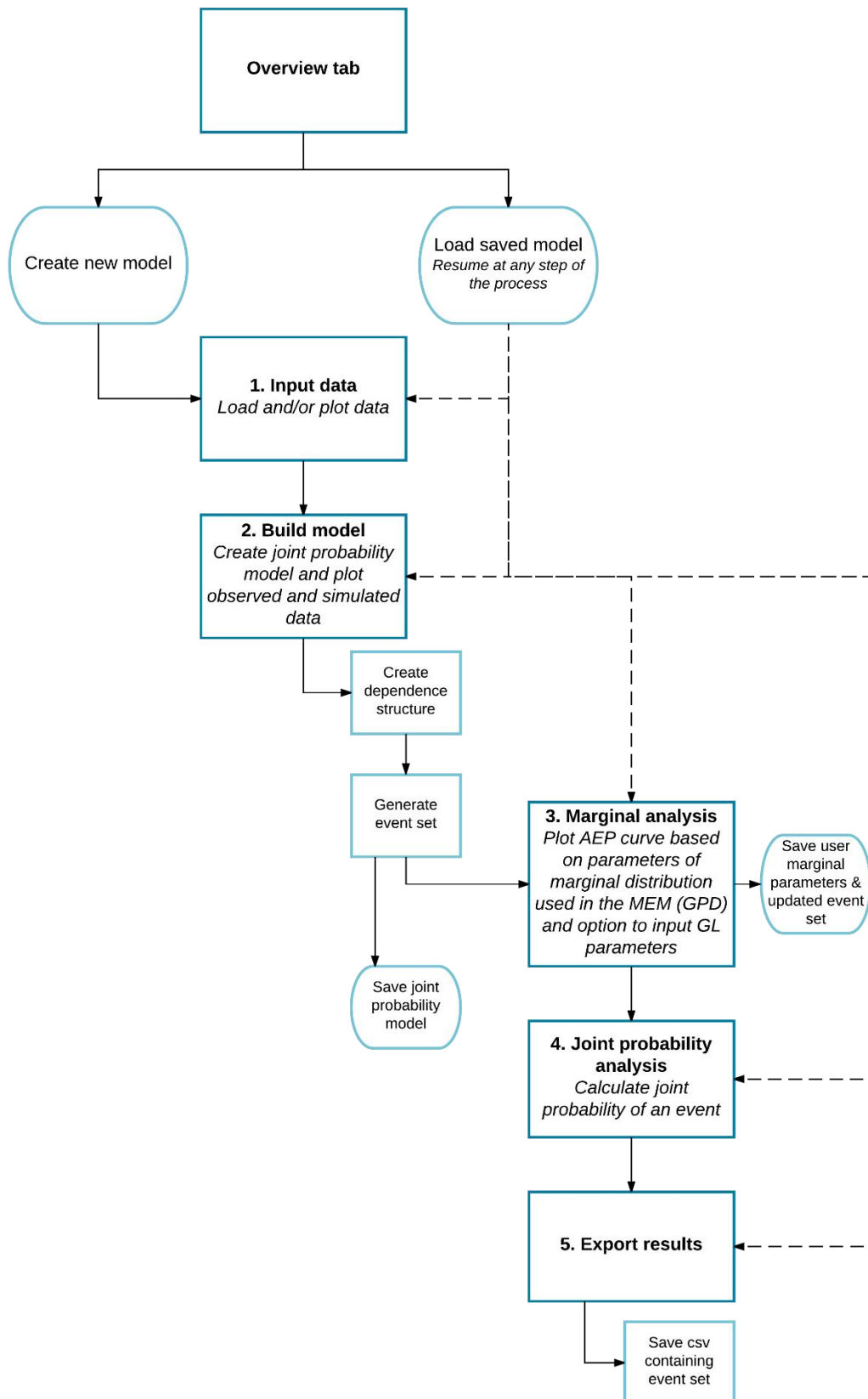
Either the 'server.R' or 'ui.R' file need opening using the 'Open an existing file' button (circled in Figure 2.11) or 'Ctrl+O'. You should then see a 'Run App' button at the top of the box containing the script (circled in Figure 2.12). Click this button to open the MEM tool.



**Figure 2.12 How to open the MEM tool in RStudio**

The MEM should open in a new RStudio viewer window where it can be used. Alternatively, the MEM can be run in an internet browser using the 'Open in Browser' button when the MEM loads (see Figure 3.2).

# 3 Tool overview



**Figure 3.1 Tool workflow**

Notes: GL = generalised logic; GPD = generalised Pareto distribution

The MEM contains 6 tabs:

- Overview
- Input data
- Build model
- Marginal analysis
- Joint probability analysis
- Export

The tabs can be worked through sequentially as shown in the workflow in Figure 3.1 and are described in further detail in this section.

After completing each tab, you can switch back to a previous step at any time and can return to the 'Build model' tab to simulate a new event set, for example, if you wish to simulate more years of data.

Plots may not render properly if switching between tabs while plots are loading so it is recommended that you stay on a tab until the plots have completely appeared.

## 3.1 Overview tab

The overview tab (Figure 3.2) introduces the MEM tool and its structure, providing information about each tab. There are 2 options on this tab:

- Start a new analysis with new data by clicking either the 'Create a new joint probability model' button or the '1 Input data' tab.
- Load a saved model to carry out further analysis using a previously created model and simulated event set (as saved as a binary .RDS file in the 'Build model' tab) by clicking on the 'Load saved model' button and then the 'Browse' button.

Once you select to load a saved model, the option to input new data will disappear. The MEM will need to be refreshed to bring back the option. When data have been loaded, all options to load data will disappear and the MEM will need to be refreshed to load new data.

You can start a fresh analysis at any time by pressing the 'Reload the Shiny application' button when using the RStudio viewer window or the browser refresh button (or F5) when using a web browser.

There is an option to open the MEM in a web browser (opens in your default browser). This will open a fresh version of the MEM which will not contain any information you may have already entered. The MEM must also be open in the RStudio viewer window to work in a web browser. Although instances of the MEM open in both the RStudio viewer window and a browser simultaneously are not linked to each other, carrying out 2 analyses at the same time could slow both processes down. It is therefore recommended that one analysis is carried out at a time in either the RStudio viewer window or in a web browser. A web browser may be faster at rendering plots.

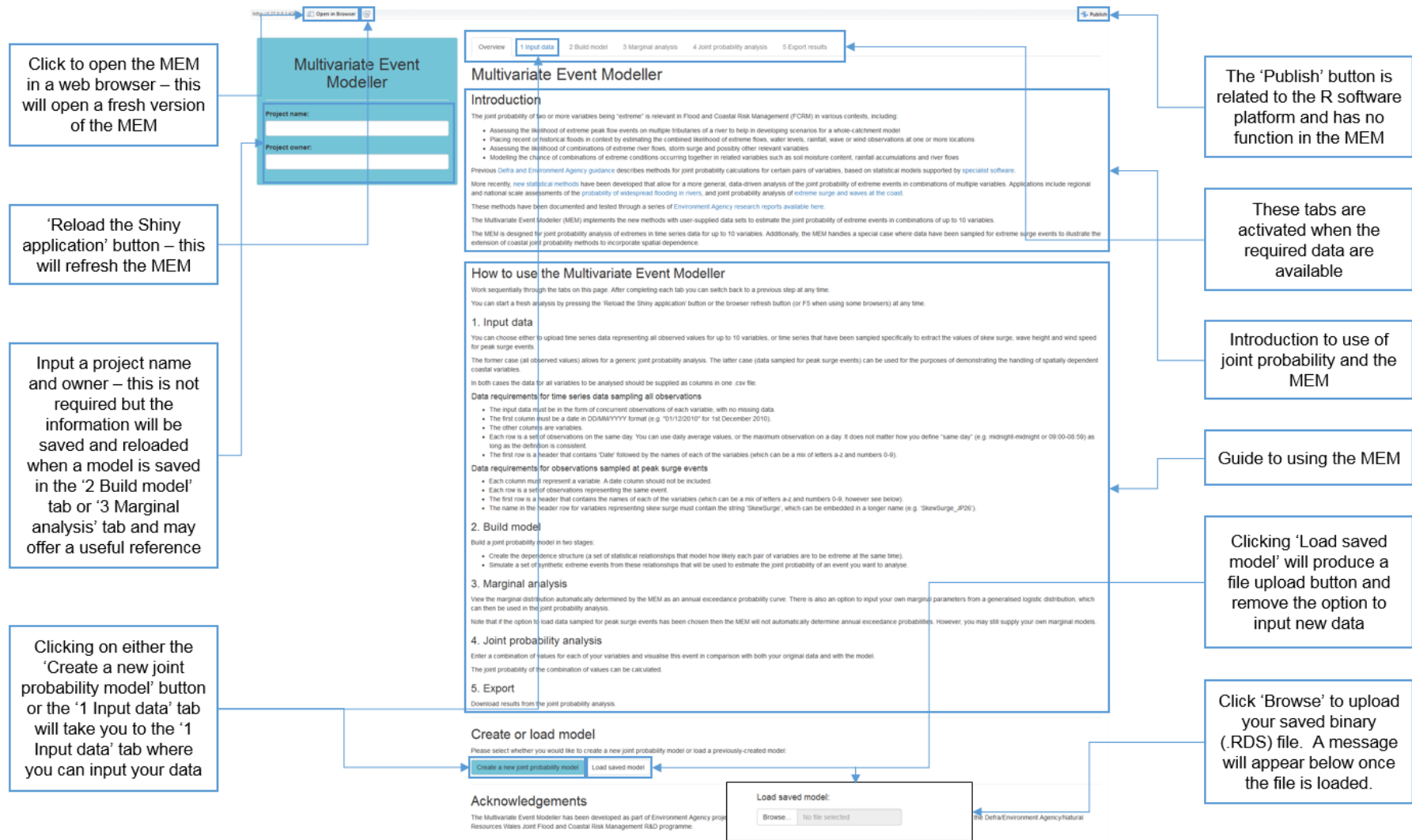


Figure 3.2 'Overview' tab

## 3.2 Input data tab

If you have selected to input new data, this tab will appear as in Figure 3.3 with the file upload button visible.

There is an option to upload either time series data sampling all observations or observations sampled at peak surge events. You need to select which you will be uploading before you upload the data, as the options will disappear once data have been loaded and the selection affects the plots and analysis in the MEM.

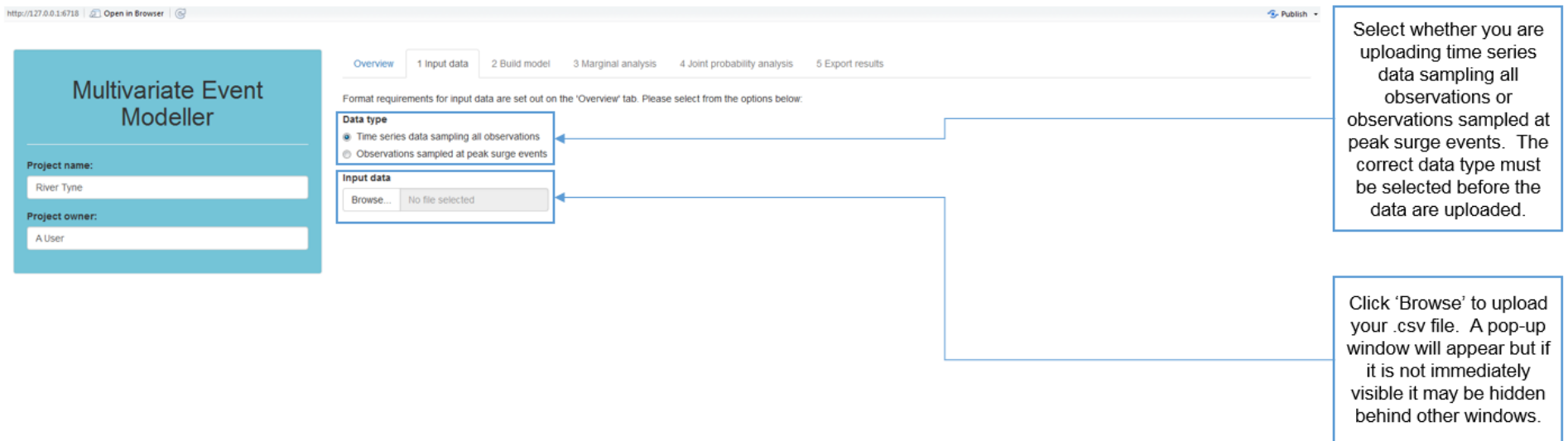
The MEM will try to identify if the uploaded data are inconsistent with the selected data type and may give a warning message but this is something to check if errors appear.

You will need to refresh the MEM to change your selection once you have uploaded data.

If time series data sampling all observations are uploaded, the MEM internally declusters each variable based on the runs method. This method is commonly used in threshold exceedance modelling and separates clusters wherever 7 consecutive data values fall below the threshold

(<https://www.rdocumentation.org/packages/evd/versions/2.3-2/topics/fpot>).

When the methodology was applied for the NRA scenarios, a minimum record length of 20 years was applied to each flow gauge.



**Figure 3.3** Initial view of 'Input data' tab if you have selected to create a new joint probability model

The 'Browse' button allows a .csv file to be uploaded containing data for up to 10 variables. Uploading data that are not in a .csv file may cause errors. If you wish to change your data, the MEM must be refreshed first.

The data must be in a specific format to be used in the MEM tool (see examples in Table 3.1 and Table 3.2).

The **data requirements for time series data sampling all observations** are as follows.

- The input data must be in the form of concurrent observations of each variable, with no missing data. Any missing periods of observations must be infilled to create a complete, continuous record for input to the MEM; see the FAQs in Section 5 for more information.
- The first column must be a date in DD/MM/YYYY format (for example, '01/12/2010' for 1 December 2010). It may be worth checking this by opening the .csv file in a text editor.
- The other columns are variables.
- Each row is a set of observations on the same day. You can use daily average values, or the maximum observation on a day. It does not matter how you define 'same day' (for example, midnight to midnight or 09:00 to 08:59) as long as the definition is consistent.
- The first row is a header that contains the names of each of the variables (which can be a mix of letters a–z and numbers 0–9). These names will be used in figures and tables throughout the MEM. If any of the names contain spaces, these will be replaced with a '.', as shown in the figures in this user guide for the River Tyne example, which uses data from Haydon Bridge. The MEM will also automatically add an 'X' to the front of variable names that start with a number.

**Table 3.1 Example format for time series data sampling all observations**

Date	Bywell	Reaverhill	Haydon Bridge
21/09/2002	7.273	3.252	3.448
22/09/2002	5.916	3.252	3.345
23/09/2002	5.703	3.109	3.178
24/09/2002	5.987	4.363	3.080
25/09/2002	6.882	3.700	2.952
26/09/2002	17.435	12.018	2.826
27/09/2002	29.260	12.018	2.703
28/09/2002	13.180	12.018	2.583
29/09/2002	13.180	8.851	2.437
30/09/2002	5.495	2.910	2.295

The **data requirements for observations sampled at peak surge events** are as follows.



- Each column must represent a variable. A date column should not be included and the MEM will add an index column.
- Each row is a set of observations representing the same event.
- The first row is a header that contains the names of each of the variables (which can be a mix of letters a–z and numbers 0–9, but see next bullet point). The comments on the naming convention from above still apply.
- The name in the header row for variables representing skew surge must contain the string ‘SkewSurge’, which can be embedded in a longer name (for example, ‘SkewSurge\_JP26’).

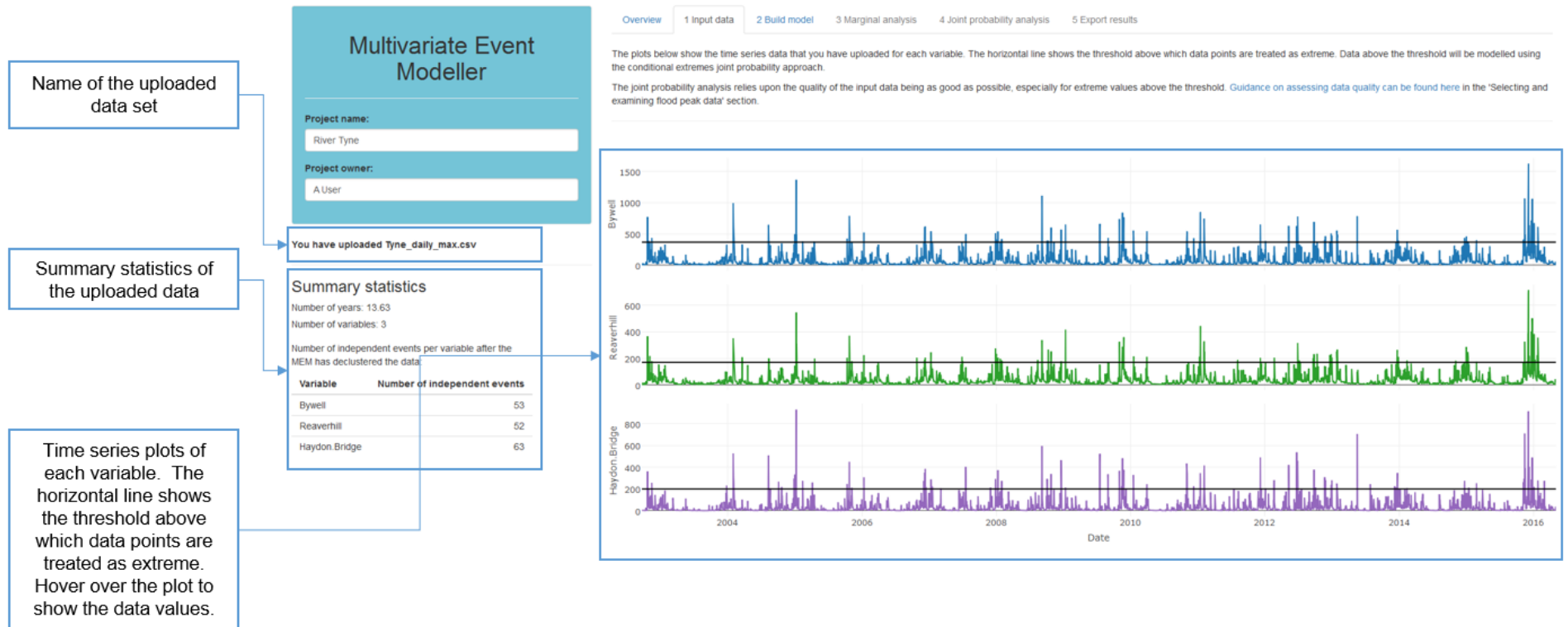
**Table 3.2 Example format for observations sampled at peak surge events**

<b>SkewSurge.JP26</b>	<b>WaveHeight.JP26</b>	<b>SkewSurge.JP27</b>	<b>WaveHeight.JP27</b>
0.234	1.787973	0.208963	2.409756
0.02363	1.815529	0.021606	0.523028
0.283888	1.750618	0.209	2.482906
0.090756	0.885693	0.079	1.0772
0.098469	1.206618	0.162	2.383084
0.359972	1.764836	0.315	2.343085
0.313159	1.472587	0.405	2.494383
0.097374	0.687542	0.28	1.280819
-0.0231	0.910051	0.127	1.978013
0.117	0.872204	0.197	1.777779

The joint probability analysis relies on the quality of the input data being as good as possible, especially for extreme values above the threshold. Guidance on assessing data quality can be found in the Environment Agency’s Flood Estimation guidelines (Environment Agency 2012).

Once data are uploaded, or if you have loaded a saved model, the tab will appear as in Figure 3.4 or Figure 3.5, depending on the type of data uploaded, and the next tab will be activated.

Summary statistics for the data are displayed in the left panel, with plots of the data in the main panel. If you have entered data for observations sampled at peak surge events (Figure 3.5), you will need to enter the number of years of data represented (that is, the number of years in the full dataset before the observations were sampled at peak surge events). The threshold used to define the extreme values is plotted as a horizontal line for each variable. The 97.5th percentile has been used in the MEM, following previous applications of the methodology, and data above the threshold will be modelled using the conditional extremes joint probability approach.



**Figure 3.4** 'Input data' tab after time series data sampling all observations (or a saved model which used this type of data) have been uploaded



**Figure 3.5** 'Input data' tab after observations sampled at peak surge events (or a saved model which used this type of data) have been uploaded

### 3.3 Build model tab

The main panel in this tab shows the relationships between each pair of variables in your uploaded data. The plot controls on the left allow you to choose whether to view your data on their original measurement scale (for example, water level or flow rate), or on a mathematically transformed scale that emphasises the extreme values and standardises across differences in the absolute physical scale ('standardised scale').

The right side of the tab guides you through the process of building a joint probability model. This is carried out in 2 stages.

- Create the dependence structure – a set of statistical relationships that model how likely each pair of variables are to be extreme at the same time.
- Simulate a set of synthetic extreme events from these relationships that will be used to estimate the joint probability of an event you want to analyse.

Figure 3.6 shows the initial view of the tab. The scatter plots pair up the data at the same date or time index (that is, there is no time lag); a time lag could be introduced by offsetting the date of the input data.

Figure 3.7 shows the tab once the dependence structure has been created, with the data plotted on a standardised scale. If the record length of the input data is not long enough or a variable lacks sufficient extreme events to permit robust analysis of the extremes, a warning will appear.

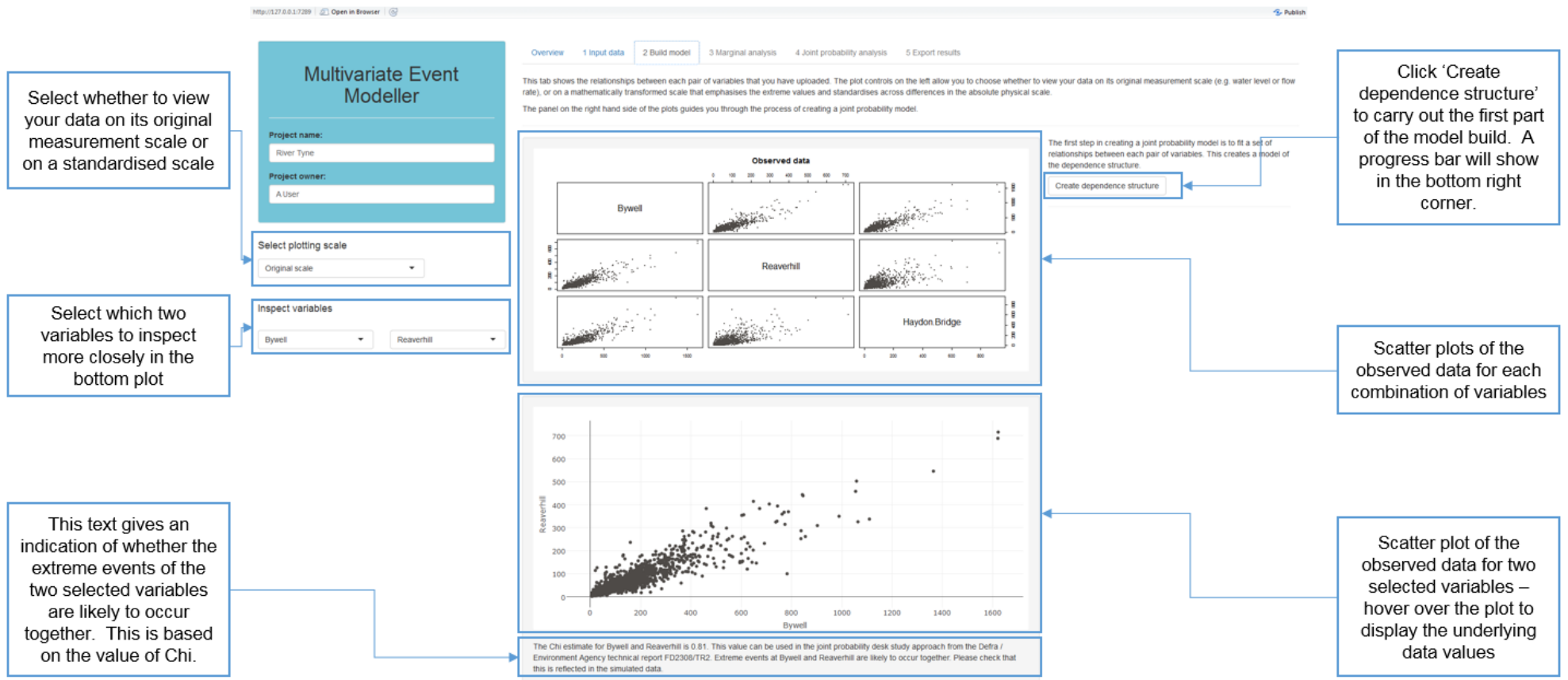
You can then select the number of years of events that you would like to simulate (this is a synthetic period of record). A larger simulated sample will give you a more precise analysis of joint probabilities, but there is a trade-off because a large simulation may also take some time. Simulating a minimum of 10,000 years of synthetic events is suggested. The maximum number of events that can be simulated will depend on the particular configuration of the user's computer system. Very large simulations may result in memory errors or unacceptably long processing times. If a larger number of events is required, it is recommended that the simulation size is increased incrementally by doubling the number of events to establish what is feasible.

Table 3.3 provides example runtimes for different numbers of variables and years of simulated events. Note that these will vary across machine specifications and be dependent on whether other processing is taking place simultaneously.

**Table 3.3 Example runtimes for different numbers of variables and years of simulated events**

<b>Number of variables</b>	<b>Record length</b>	<b>Number of years of simulated events</b>	<b>Runtime</b>	<b>PC specification</b>
2	52.5 years	20,000	Creating dependence structure: 7 seconds  Event simulation: 30 seconds (plus 15 seconds for updating the plots)	Windows 7  16.0GB RAM  64-bit operating system  Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
2	52.5 years	100,000	Creating dependence structure: 7 seconds  Event simulation: 2 minutes 40 seconds (plus 1 minute for updating the plots)	Windows 7  16.0GB RAM  64-bit operating system  Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
3	14 years	20,000	Creating dependence structure: 5 seconds  Event simulation: 2 minutes 55 seconds (plus 15 seconds for updating the plots)	Windows 7  16.0GB RAM  64-bit operating system  Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
3	14 years	50,000	Creating dependence structure: 5 seconds  Event simulation: 7 minutes 10 seconds (plus 30 seconds for updating the plots)	Windows 7  16.0GB RAM  64-bit operating system  Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
10	23 years	10,000	Creating dependence structure: 25 seconds  Event simulation: 2 minutes 20 seconds (plus 40 seconds for updating the plots)	Windows 7  16.0GB RAM  64-bit operating system  Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
10	23 years	20,000	Creating dependence structure: 25 seconds  Event simulation: 4 minutes (plus 45	Windows 7  16.0GB RAM  64-bit operating system

Number of variables	Record length	Number of years of simulated events	Runtime	PC specification
			seconds for updating the plots)	Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
10	23 years	50,000	Creating dependence structure: 25 seconds Event simulation: 12 minutes (plus 1 minute 45 seconds for updating the plots)	Windows 7 16.0GB RAM 64-bit operating system Processor: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz



**Figure 3.6** Initial view of the 'Build model' tab before creating a dependence structure



A warning message appears once the dependence structure has been created if the record length is too short to permit robust analysis of the extremes

Once the dependence structure has been created, the 'Simulate events' button appears with an option to select the number of years of synthetic events that you would like to simulate. This updates automatically when you load a saved model to the number of years used in that model.

A progress bar appears here when the dependence structure is being created and when the events are being simulated. There are two stages to the simulation progress bar: 'Simulating events' and 'Processing events'.

Figure 3.7 'Build model' tab once the dependence structure has been created



The screenshot shows the 'Multivariate Event Modeller' interface, specifically the 'Build model' tab. The interface includes a header with navigation tabs (Overview, 1 Input data, 2 Build model, 3 Marginal analysis, 4 Joint probability analysis, 5 Export results) and a main content area with several panels.

**Left Panel (Form):**

- Project name:** River Type, Project owner: A User
- Select plotting scale:** Original scale (dropdown)
- Display options:** Observed data (checked), Simulated data (checked)
- Inspect variables:** Bywell, Reaverhill (dropdowns)
- Data thinning:** Slider from 1 to 100, currently at 10. Text: "Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes."

**Top Right Panel (Text):**

This tab shows the relationships between each pair of variables that you have uploaded. The plot controls on the left allow you to choose whether to view your data on its original measurement scale (e.g. water level or flow rate), or on a mathematically transformed scale that emphasises the extreme values and standardises across differences in the absolute physical scale. The panel on the right hand side of the plots guides you through the process of creating a joint probability model.

**Center Panel (Plots):**

**Observed data and 10,000 years' worth of simulated events:** A 3x3 grid of scatter plots for Bywell, Reaverhill, and Haydon Bridge. The axes represent flow rate (m³/s) for each location.

**Large Scatter Plot:** A plot of Reaverhill vs Bywell flow rate. It shows a dense cloud of blue points (Simulated) and a smaller set of black points (Observed). The axes range from 0 to 2000 m³/s.

**Bottom Panel (Text):**

The Chi estimate for Bywell and Reaverhill is 0.81. This value can be used in the joint probability desk study approach from the Defra / Environment Agency technical report FD2305/TR2. Extreme events at Bywell and Reaverhill are likely to occur together. Please check that this is reflected in the simulated data.

**Right Panel (Form):**

- Create dependence structure:** Button
- Dependence structure created:** Section header
- Warning:** The input data is not a sufficiently long record to permit robust analysis of the extremes, or might be lacking sufficient extreme events (e.g. this is possible with a groundwater flow series).
- Enter number of years to simulate:** Input field with 10000, Simulate events button
- Event simulation completed:** Section header
- Save model:** Button
- Choose location for saved model:** Button
- Destination folder to save model:** Input field
- Filename for saved model:** Input field (example: "Model 1")
- Save model now:** Button

**Callout Boxes:**

- Top Left:** Select whether you would like to view your observed data, the simulated data or both
- Bottom Left:** Select the percentage of the simulated data to plot
- Top Right:** Option to save the model once the event set has been simulated
- Middle Right:** Opens a dialogue box to choose a destination folder for saving the model
- Bottom Right (Top):** Selected folder path appears here or you can enter a location manually (e.g. C:\temp)
- Bottom Right (Bottom):** Enter a filename for the saved model
- Bottom Center:** Save model (the file automatically saves as a binary file)

Figure 3.8 'Build model' tab once the event set has been simulated, or initial view of the tab if you have loaded a saved model

Once the simulation is complete (Figure 3.8), the remaining tabs will be activated and the plots will update to include the events simulated using the conditional extremes model (the plots will be greyed out while updating and this will take longer for larger event sets). The cloud of simulated points should look similar to the cloud of observed points, but should extend beyond the range of the observed dataset so as to include more extreme events than in the observations (see the plots in Figure 3.8).

The joint probability analysis will require some of the simulated data to be more extreme than events you wish to analyse. If this is not the case, you can repeat the simulation for a larger number of years.

Additional plotting controls also appear on the left to provide an option for selecting which data to view and to control thinning of the data. The larger event sets take longer to plot and so thinning the data randomly speeds up the plotting, although some detail will be lost in the plots. This is purely a visual performance optimisation and will not affect the calculations. There may be situations where a combination of substantial thinning, simulation size and marginal parameters lead to there being no simulated data to plot, in which case an error may be reported ('argument of length 0'). If this is encountered, try increasing the proportion of data plotted.

Testing a large simulation representing 100,000 years of data for 3 variables suggested that plotting 35% of the data could be very slow and plotting 50% of the data could take up to 15 minutes and not allow access to the MEM during this time. This applies to a mid-range laptop (of around 2015 vintage). Performance with a large simulated event set and observed dataset will depend on the hardware being used but increasing the plotting proportion incrementally is suggested. For large simulated event sets, it will probably be unnecessary to plot a large proportion of the data. The larger the simulated event set, the longer most processing within the MEM will take.

There is an option to save the model you have created. This applies when you have created a new joint probability model and when you have simulated a new event set after loading a previously saved model; it will not work if you load a previously saved model and try to save it again without simulating a new event set first (you can check that a new event set is being simulated as a progress bar will appear in the bottom right corner).

The input data (and number of years represented by observations sampled at peak surge events), dependence structure and simulated event set (including the number of years simulated) are saved as a binary file and this can be loaded back into the MEM on the overview tab so that you do not need to repeat the model build. If you save the model with the same name as an existing .RDS file, then the new file will overwrite the existing file.

## 3.4 Marginal analysis tab

This is an optional tab (Figure 3.9) where you can view the marginal distribution used in the joint probability analysis. The marginal analysis in the MEM tool fits a generalised Pareto distribution to peaks over threshold data for each variable. This tab also gives you the option to input your own marginal parameters for each variable from a generalised logistic distribution (for example, the marginal parameters obtained from a Flood Estimation Handbook flood frequency analysis).

The 2 choices of marginal analysis can be compared by plotting the AEP curves. Either of the 2 marginal analyses can be used in the 'Joint probability analysis' tab. If, however, the option to load data sampled for peak surge events has been chosen, then the MEM will not automatically determine AEPs. You may still supply your own marginal models, which can then be used in the 'Joint probability analysis' tab. In

this situation, no data will be plotted unless you have supplied your own marginal parameters and selected to plot them in the display options. In all cases, the plot will be empty between the user ticking the box to input their own marginal parameters and actually entering the parameters.

If you select to use your own marginal parameters, a table will appear where you can enter your location, scale and shape parameters for each variable. The 'scale' parameter must be greater than zero and you must input parameters for all variables as the MEM will use only the parameters in the table if you have selected to use your own marginals. If the table does not fully appear initially, try clicking in it and you should be able to enter all your parameters. It is possible to copy the data from a spreadsheet and paste it into the table, but be careful that the data are the correct number of dimensions and pasted into the top left cell – otherwise they may cause errors.

Once you have entered your parameters, you will need to update the simulation outputs from the 'Build model' tab (see Figure 3.10) to be consistent with the parameter values that you have specified. This results in the severity of the simulated event set being determined using your specified generalised logistic parameters. You will not be able to save the event set if you have selected to use your own marginal parameters but have not updated the simulation outputs. This updated event set will then be plotted in the 'Joint probability analysis' tab rather than the original simulated event set and will be used in the joint probability calculation. The generalised logistic parameters will also be used to convert between AEP (%) and value in the table in the 'Joint probability analysis' tab rather than the generalised Pareto distribution parameters. If you change the parameters, make sure that you update the simulation outputs again (you will see the progress bar appear in the bottom right corner).

If you update the simulation outputs, there is an option to save the updated event set you have created. This saves all the information that is saved on the 'Build model' tab, as well as the updated event set and your marginal parameters. This will work when you have inputted your own marginal parameters and updated the simulation outputs for the first time (including after loading a model that had previously been saved on the 'Build model' tab) and when you have updated your parameters and the simulation outputs again after loading a model previously saved on this tab. It will not work if you load a model previously saved on this tab and try to save it again without changing your parameters and updating the simulation output (you can check that the simulation outputs are being updated as a progress bar will appear in the bottom right corner). As with the save button on the 'Build model' tab, this is saved as a .RDS binary file which can be loaded back into the MEM on the 'Overview' tab so that you do not need to repeat these steps.

If you save the model with the same name as an existing .RDS file then the new file will overwrite the existing file.



Figure 3.9 Initial view of 'Marginal analysis' tab

## Multivariate Event Modeller

**Project name:**  
River Tyne

**Project owner:**  
A User

Select variable to plot  
Bywell

Display options

Automatically-determined marginals (Auto)

User-supplied marginals (User)

Overview 1 Input data 2 Build model 3 Marginal analysis 4 Joint probability analysis 5 Export results

On this tab you can view the marginal distribution used in the joint probability analysis. The automatic marginal analysis in the MEM fits a generalised Pareto distribution to peaks over threshold data for each variable. This tab additionally gives you the option to input your own marginal parameters from a generalised logistic distribution (e.g. the marginal parameters obtained from a Flood Estimation Handbook flood frequency analysis). The two choices of marginal analysis can be compared by plotting the annual exceedance probability curves. Either of the two marginal analyses can be used in the 'Joint probability analysis' tab.

Tick the box if you would like to enter your own marginal parameters from a generalised logistic distribution. If the box is unticked the existing marginal analysis will be used in the joint probability analysis.

Input and use your own marginal parameters

	Location	Scale	Shape
Bywell	751.00	179.00	-0.20
Reaverhill	318.00	77.00	-0.20
Haydon Bridge	468.00	112.00	-0.20

Click the button to update the simulation outputs to use your generalised logistic marginal parameters. You will need to click the button again if you change your marginal parameters.

The simulated data have been updated.

**Save model**

This panel can be used to save your marginal parameters and the updated simulation outputs as well as the outputs from the 'Build model' tab for loading back in later.

**Destination folder to save model:**

**Filename for saved model:**

AEP curves based on automatically-determined marginals and user-supplied marginals

Input generalised logistic parameters for all variables into the table

Click button to update simulation outputs after inputting your own parameters. A message will appear below the button once updated.

Save model including your marginal parameters and the updated simulation outputs. The panel layout is the same as on the 'Build model' tab.

Figure 3.10 'Marginal analysis' tab if 'Input and use your own marginal parameters' is selected

## 3.5 Joint probability analysis tab

On this tab (Figure 3.11 and Figure 3.12), you can assess the joint probability that a set of extreme values in each of your variables could be exceeded in any one event. You can define an event by entering values of either the percentage chance (AEP in %) or the physical measurement for each of your variables in the table. The other column will automatically update. If the data type selected was observations sampled at peak surge events, then only physical measurement values can be entered unless you have specified your own marginal parameters.

As with the marginal parameters table, if the table does not fully appear initially, try clicking in it and you should be able to enter your values. It is possible to copy the values from a spreadsheet and paste them into the table for one column; be careful that you are copying the correct number of rows and paste it into the top cell otherwise it may cause errors. Only paste data for one column as the other column will update automatically.

When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%. If you enter an AEP of 100% then the minimum values from the observed data will appear in the 'Value' column. If you enter a physical measurement value that corresponds to an AEP value that rounds to 100%, the AEP value will be replaced with 99.90% to avoid implying a deterministic prediction (that is, a probability of 1). The physical measurement values will be shown as lines on the plots (when viewing data on the original measurement scale) to allow the event to be visualised in comparison with your observed data and the simulated data. You can input 'NAs' for variables for which you do not have data or you do not wish to include in the joint probability calculation although there must be data for at least 2 variables. You can copy the values from the table if you would like to save them elsewhere.

For a joint probability of more than zero, simulated data must exist in the top right quadrant above the lines on the plot which represent the event (the joint probability region) in all combinations of variables. These plots only show data for a pair of variables at a time, so it is possible that simulated data may not exist above the user-defined event for the other variables, which would result in a probability of zero.

The grey areas that show on the plots when viewing the simulated data indicate the non-extreme regions for each pair of variables. The plots update based on the values in the table and this can take some time whenever you add a value to the table. There is therefore an option to stop all the data plotting while updating the table to speed up this process. Alternatively, the plots do not update when viewing the data on a standardised scale so, if you do not need to view the values on the plot while updating the table, it would be faster to update the table while the plotting scale is set to standardised and then switch back to the original measurement scale. The plots will show as greyed out while they are updating, indicating that the MEM is still processing.

The MEM will compare your specified event to the events simulated from the model and use this information to estimate the joint probability. It will do this by calculating the proportion of years in which simulated events are more extreme than the values you enter in the table. It is suggested that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set, although it is possible that no events will fall in the joint probability region even with a large simulation. A range of encounter probabilities is also calculated; these are the percentage chance of the user-defined event in the table occurring in the given number of years.

The plots and calculations in this tab will change depending on whether the marginal parameters and simulated event set are based on the default MEM settings or whether you have selected to use your own marginal parameters from a generalised logistic distribution (see the third paragraph in Section 3.4). If you change your preference by ticking or unticking the 'Input and use your own marginal parameters' check box in the 'Marginal analysis' tab, the plots and the table in the 'Joint probability analysis' tab will automatically update – although this can take a little time, particularly with large amounts of data.

It is worth checking that the titles on the plots have updated to reflect the change and that the numbers in the table have updated (the column that was last changed will stay the same and the other column will update). The plot titles will still update when the option to load data sampled for peak surge events has been chosen; this is to make it clear when the user marginals are being used, even though an automatically determined marginal model is not actually being used in this instance. The joint probability calculation will not automatically update if you change your preference for which marginals to use, so make sure that you click the 'Calculate joint probability' button again.

**Multivariate Event Modeller**

Project name: River Tyne  
Project owner: A User

Select plotting scale: Original scale

Display options:  
 Observed data  
 Simulated data  
 Disable automatic plot updating

Inspect variables:  
 Bywell | Reaverhill

Data thinning: 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.

	AEP (%)	Value
Bywell	99.99	4.03
Reaverhill	99.99	2.68
Haydon Bridge	99.99	0.67

Calculate joint probability

Option to switch off the plots when viewing the data on the original measurement scale. This enables the table to be edited more quickly as the lines on the plots update automatically as you update the table and the plots can take time to refresh. When the plots are switched off, only the maximum and minimum data values will plot, as well as the lines corresponding to the values in the table.

Plot controls act in the same way as in the 'Build model' tab

The plots show the same data as the 'Build model' tab. When displaying the simulated data, grey areas indicate the non-extreme regions for each pair of variables.

Enter AEPs or physical measurement values into the table and the other column will automatically update. Use NAs for gauges for which you have no data.

Click to calculate the joint probability of the event you have entered into the table. You will need to click again to recalculate the joint probability if you change any values.

The black vertical and horizontal lines reflect the values in the table when viewing data on the original measurement scale. Lines with values higher than the plotted data won't be visible on the top plot. In the bottom plot, the lines can be seen near the axes as the default values are set to the minimum of the observed data.

**Figure 3.11 Initial view of the 'Joint probability analysis' tab with data plotted on its original measurement scale**





Figure 3.12 'Joint probability analysis' tab with data plotted on the standardised scale and the joint probability calculated for an event where each variable is set to a 1% chance

## 3.6 Export tab

This tab allows you to save the simulated event set as a .csv file (see Figure 3.13).

If you have selected to use your own marginal parameters from a generalised logistic distribution (that is, the 'Input and use your own marginal parameters' box is ticked on the 'Marginal analysis' tab), then the updated simulated event set will be saved. The default simulated event set will otherwise be saved. You can save both event sets by having the 'Input and use your own marginal parameters' box ticked when saving and then unticking the box and saving with a different name. If you save an event set with the same name as a .csv file that already exists, the original file will be overwritten.

The saved values are on the original measurement scale. Table 3.4 is an example of the first part of a simulated event set with a column of synthetic dates and a column for each of your input variables. The years in the date column will range from one to the number of years of data you have simulated and each row in the .csv file is a simulated event.

**Table 3.4 Example head of a simulated event set**

<b>Date</b>	<b>Bywell</b>	<b>Reaverhill</b>	<b>Haydon Bridge</b>
00001-01-01	423.7249252	218.4063133	133.5631785
00001-01-01	583.3716338	199.043348	309.3908287
00001-01-01	495.548666	221.8355402	428.0669201
00001-01-01	595.4306263	288.324738	313.8903914
00001-01-01	721.896538	246.0130171	348.8399247
00001-01-01	836.2824443	425.2091315	445.2085728
00001-01-01	367.5379993	209.0812099	197.4843164

Notes: Synthetic dates are shown in the first column.  
The featured flow gauges are from Case study 3 (see Section 4.3).

The screenshot displays the 'Export results' tab of the Multivariate Event Modeller. On the left, a sidebar contains the application title and user information. The main panel features a navigation menu and a 'Save event set' section. This section includes a 'Choose location for saved model' button, a text input field with the path 'C:\temp', another text input field with the filename 'River\_Tyne\_Event\_Set', and a 'Save event set' button. Five callout boxes on the right provide detailed instructions for each of these elements.

**Multivariate Event Modeller**

Project name:  
River Tyne

Project owner:  
A User

Overview 1 Input data 2 Build model 3 Marginal analysis 4 Joint probability analysis 5 Export results

**Save event set**  
This panel can be used to save the event set. This will be the automatically-simulated event set unless you have selected to use user supplied marginal parameters, in which case it will be the updated event set which is consistent with your generalised logistic parameters. This is based on whether the 'Input and use your own marginal parameters' box is ticked on the 'Marginal analysis' tab.

Choose location for saved model

C:\temp

River\_Tyne\_Event\_Set

Save event set

Option to save the simulated event set (values are on the original measurement scale of your observed data)

Opens a dialogue box to choose a destination folder for saving the model

Selected folder path appears here or you can enter a location manually

Enter a filename for the saved event set

Save event set (the file automatically saves as a .csv file)

Figure 3.13 'Export results' tab

# 4 Case study applications

This section contains 5 case studies for illustration only to highlight the types of situations where the MEM tool/approach could be used.

The first of these is a validation case study which uses a simulated dataset containing 5 variables that are dependent and 5 variables that are independent.

Two case studies address hydrological applications of the MEM. The first uses the MEM to determine potential combinations of marginal probabilities (that is, probabilities at each gauge) that result in a specified joint probability, and the second uses the MEM to calculate the joint probability of the flows experienced during Storm Desmond in December 2015 at 3 gauges in the north-east.

There is also a case study showing how the MEM can be used for an estuary application and one demonstrating a coastal application where data were sampled at peak surge events prior to input into the MEM.

The data for these case studies are provided in the 'data' folder within the tool.

## 4.1 Case study 1: Validation based on a simulated dataset

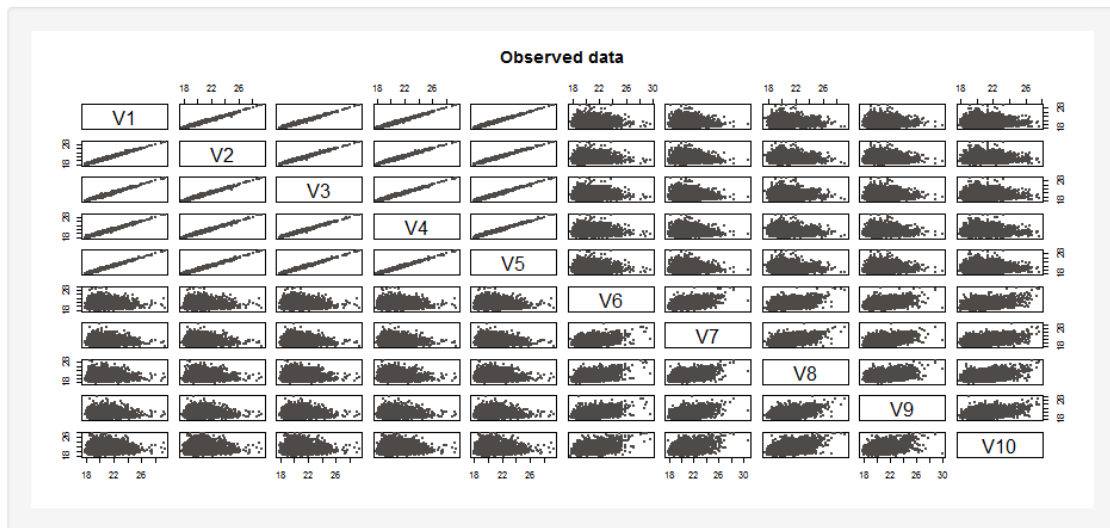
### 4.1.1 Context

A dataset containing 23 years of data for 10 variables was simulated based on known statistical distributions to provide a theoretical 'control case' for validating the MEM methods and results. Five of the variables are asymptotically independent (AI) (that is, the extremes should not occur together) and 5 are asymptotically dependent (AD) (that is, the extremes should occur together). The AI variables were simulated from a multivariate normal distribution with positive correlation. The AD variables are from a logistic distribution function.

### 4.1.2 Datasets

The data used in this case study are provided in 'Validation\_variables.csv'.

It can be seen from Figure 4.1 that V1–V5 are AD and V6–V10 are AI. The Chi values for pairs of the AD variables are very high (>0.9). However, they are much lower for pairs of the AI variables (<0.4) and are almost 0 for an AI and an AD variable together.



**Figure 4.1 Scatter plot of the simulated AI and AD dataset**

### 4.1.3 Analysis

For this analysis, 20,000 years of synthetic data were simulated. The automatically determined marginals (from a generalised Pareto distribution) were used in the joint probability analysis. The joint probability was calculated for events containing the AD variables, the AI variables and a combination of both sets of variables.

### 4.1.4 Results

#### *Event 1: Annual 1% chance for all 5 AD variables*

A 1% annual chance (an AEP of 0.01 or a return period of 100 years) was applied to each of the AD variables in the table on the 'Joint probability analysis' tab and NAs were applied to each of the AI variables. This allowed the calculation of the joint probability that the value associated with the 1% chance will be exceeded in any one event at all the 5 AD variables.

It can be seen from the lower plot in Figure 4.2 that there are simulated events in the joint probability region (circled in the figure) for the pair of variables plotted. The joint probability output text says that there are actually 11 simulated events that exceed the event entered into the table at all variables. Figure 4.2 also shows that the annual joint probability of the event is 1 in 1,818 or a 0.06% chance, and provides the encounter probabilities. There is a 2.7% chance of observing this event in 50 years.

Repeating this scenario using marginal threshold values of 0.5% (an AEP of 0.05 or a return period of 200 years) rather than 1% resulted in an annual joint probability of 1 in 5,000 (0.02% chance). As expected, this is a much lower probability than for the 1% chance event.

### *Event 2: Annual 1% chance for all 5 AI variables*

A 1% annual chance (an AEP of 0.01 or a return period of 100 years) was applied to each of the AI variables in the table on the 'Joint probability analysis' tab and NAs were applied to each of the AD variables. This allowed the calculation of the joint probability that the value associated with the 1% chance will be exceeded in any one event at all the 5 AI variables.

It can be seen from Figure 4.3 that there are simulated events in the joint probability region for the pair of variables plotted. The joint probability output text says that there are only 4 simulated events that exceed the event entered into the table at all variables. This low value would be expected as the extremes should not occur together in this dataset. Figure 4.3 also shows that the annual joint probability of the event is 1 in 5,000 or a 0.02% chance, and provides the encounter probabilities. There is only a 1% chance of observing this event in 50 years.

These numbers would decrease further for a more extreme event. For instance, when inputting a 0.5% chance for each of the AI variables, only one simulated event exceeds the user-supplied event, and zero would not be unexpected.

The joint probability of Event 2 is much lower than that of Event 1. This would be expected given that the extremes of the AD variables should occur together but the extremes of the AI variables should not occur together. The same pattern is found when assessing the joint probability of only 2 variables from the AD and AI data, in that the joint probability of the user-supplied event being exceeded in any one event is much higher for the AD variables than for the AI variables.

### *Event 3: Annual 1% chance for all 10 variables*

A 1% annual chance (an AEP of 0.01 or a return period of 100 years) was applied to each of the 10 variables in order to calculate the joint probability that the 1% chance values will be exceeded at all variables in any one event. There are no simulated events that exceed this event (Figure 4.4), which would be expected when combining the AI variables and the AD variables as their extremes would not be expected to occur together. A solution for this joint probability event would therefore require a larger simulation (that is, generation of a sample containing more than the equivalent of 20,000 years of data).

### *Event 4: Annual 1% chance for one AD variable and one AI variable*

A 1% annual chance was applied to one of the AD variables (V1) and one of the AI variables (V10), with the remaining variables containing NAs. Given the difference between the 2 variables, it is not surprising that there were no simulated events which exceeded the user-supplied event as the extremes are not likely to occur together (the Chi estimate between V1 and V10 is 0.01) (Figure 4.5).

### *Event 5: Annual 1% chance for 2 AD variables*

A 1% annual chance (an AEP of 0.01 or a return period of 100 years) was applied to 2 of the AD variables (V1 and V2) in the table on the 'Joint probability analysis' tab and NAs were applied to the remaining variables. This allowed the calculation of the joint probability that the value associated with the 1% chance will be exceeded in any one event at both V1 and V2.

This would be expected to have a higher joint probability than the case with 5 variables (Event 1) and Figure 4.6 shows that there are 12 simulated events that exceed the event entered into the table at all variables. Figure 4.6 also shows that the annual joint probability of the event is 1 in 1,667 or a 0.06% chance, which is marginally higher than the case with 5 variables. The encounter probabilities are also slightly higher for this case with 2 variables.

#### *Event 6: Annual 1% chance for 2 AI variables*

A 1% annual chance (an AEP of 0.01 or a return period of 100 years) was applied to 2 of the AI variables (V6 and V7) in the table on the 'Joint probability analysis' tab and NAs were applied to the remaining variables. This allowed the calculation of the joint probability that the value associated with the 1% chance will be exceeded in any one event at both V6 and V7.

This would be expected to have a higher joint probability than the case with 5 variables (Event 2) and Figure 4.7 shows that there are 8 simulated events that exceed the event entered into the table at all variables. Figure 4.7 also shows that the annual joint probability of the event is 1 in 2,500 or a 0.04% chance, so is twice as likely as the case with 5 variables. This is reflected in the encounter probabilities as well.



Figure 4.2 Joint probability results for the validation case study where a 1% chance has been applied to all 5 AD variables



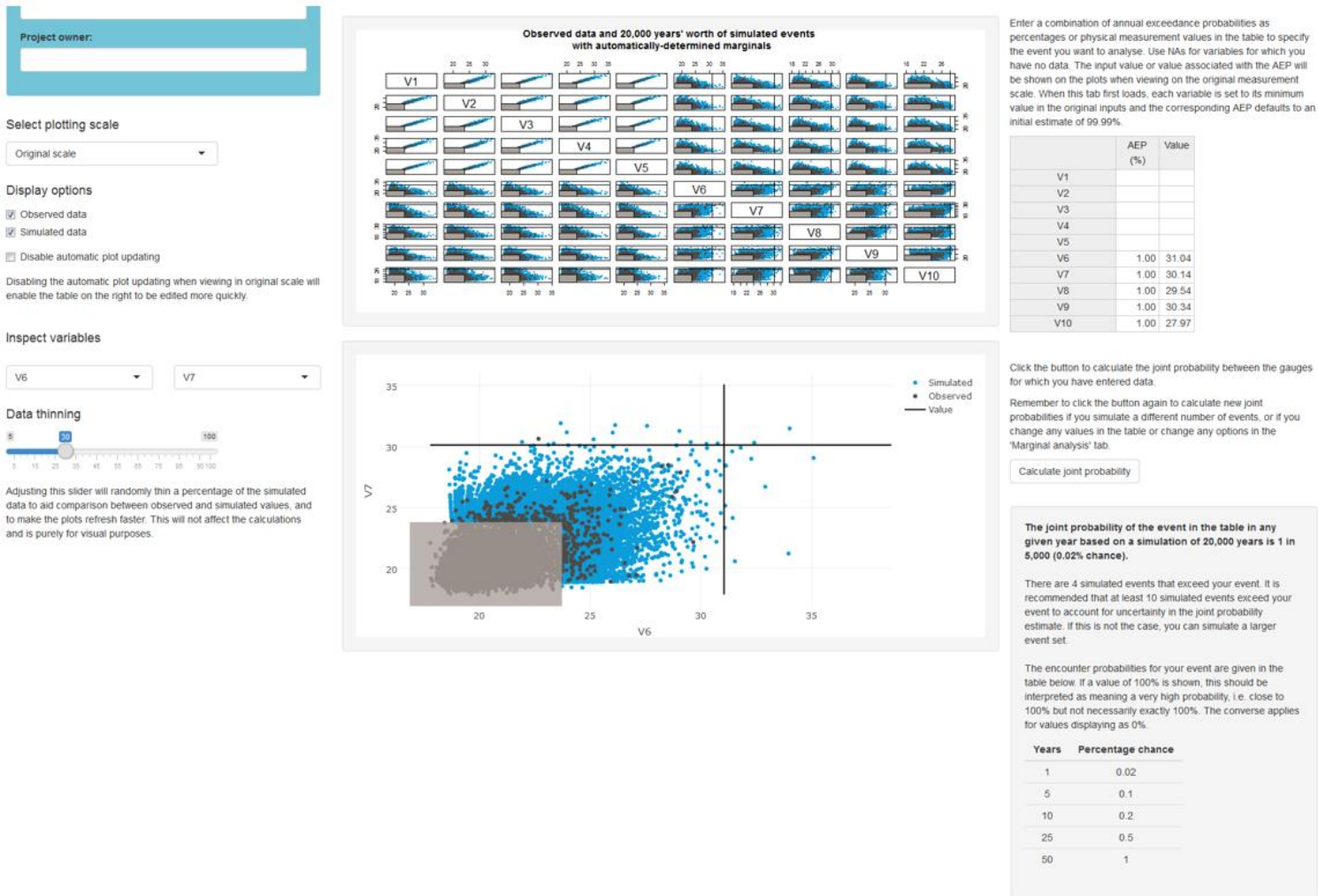


Figure 4.3 Joint probability results for the validation case study where a 1% chance has been applied to all 5 AI variables

Project owner:

Select plotting scale

Original scale

Display options

- Observed data
- Simulated data
- Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

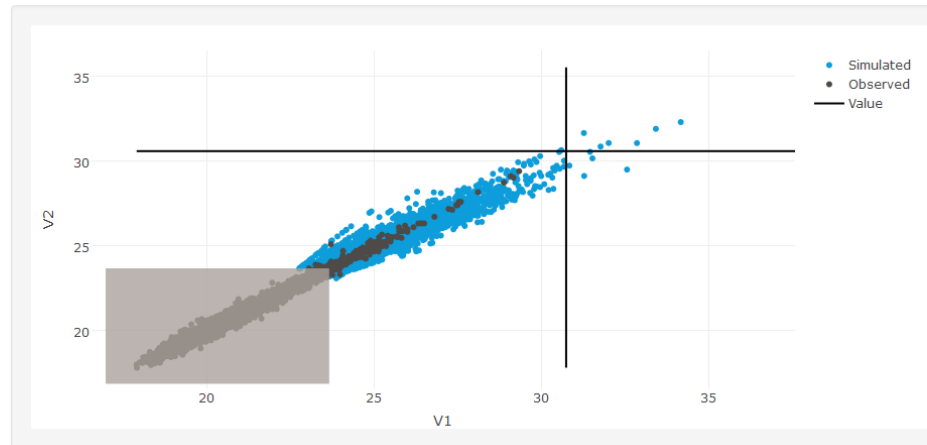
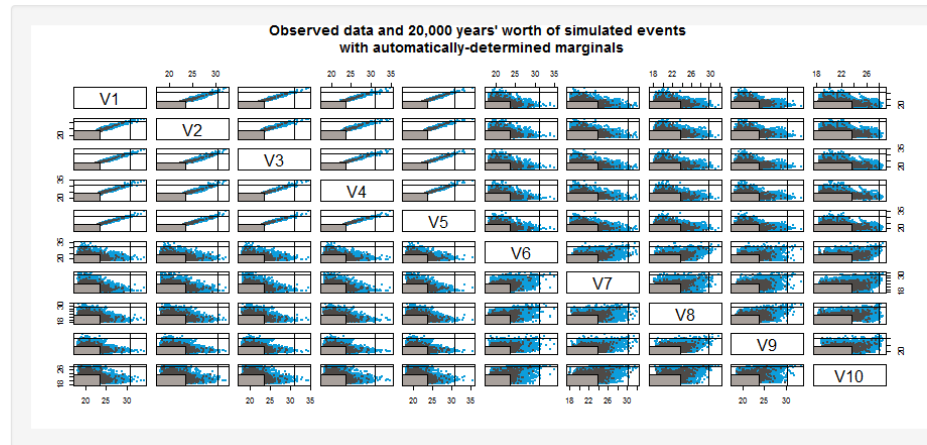
Inspect variables

V1 V2

Data thinning

5 30 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
V1	1.00	30.73
V2	1.00	30.58
V3	1.00	30.95
V4	1.00	31.01
V5	1.00	30.79
V6	1.00	31.04
V7	1.00	30.14
V8	1.00	29.54
V9	1.00	30.34
V10	1.00	27.97

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

**Error:** No events in the simulated event set are greater than your specified event. Please simulate a larger event set.

**Error:** No events in the simulated event set are greater than your specified event. Please simulate a larger event set.

**Figure 4.4** Joint probability results for the validation case study where a 1% chance has been applied to all 10 variables in the validation dataset

Project owner:

Select plotting scale

Original scale

Display options

- Observed data
- Simulated data
- Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

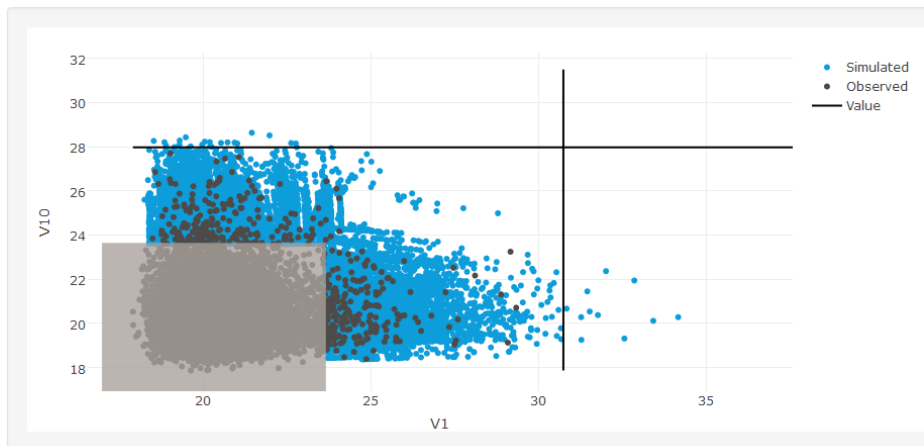
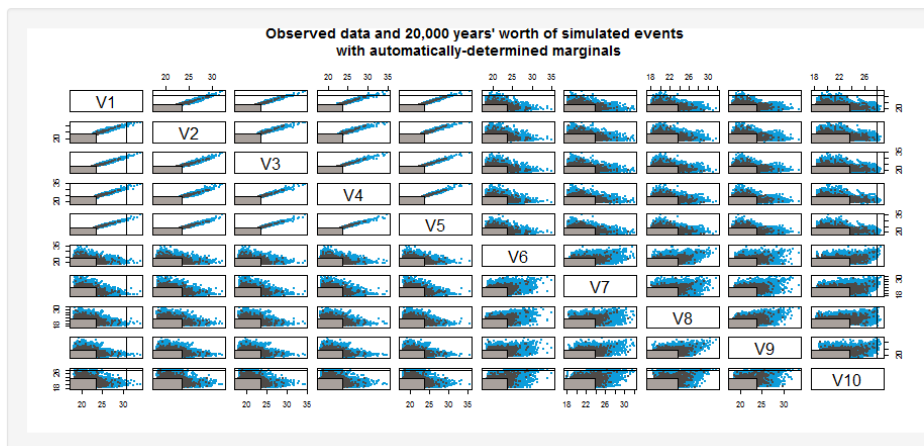
Inspect variables

V1 V10

Data thinning

5 30 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
V1	1.00	30.73
V2		
V3		
V4		
V5		
V6		
V7		
V8		
V9		
V10	1.00	27.97

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

**Error:** No events in the simulated event set are greater than your specified event. Please simulate a larger event set.

**Error:** No events in the simulated event set are greater than your specified event. Please simulate a larger event set.

**Figure 4.5** Joint probability results for the validation case study where a 1% chance has been applied to one of the AD variables and one of the AI variables

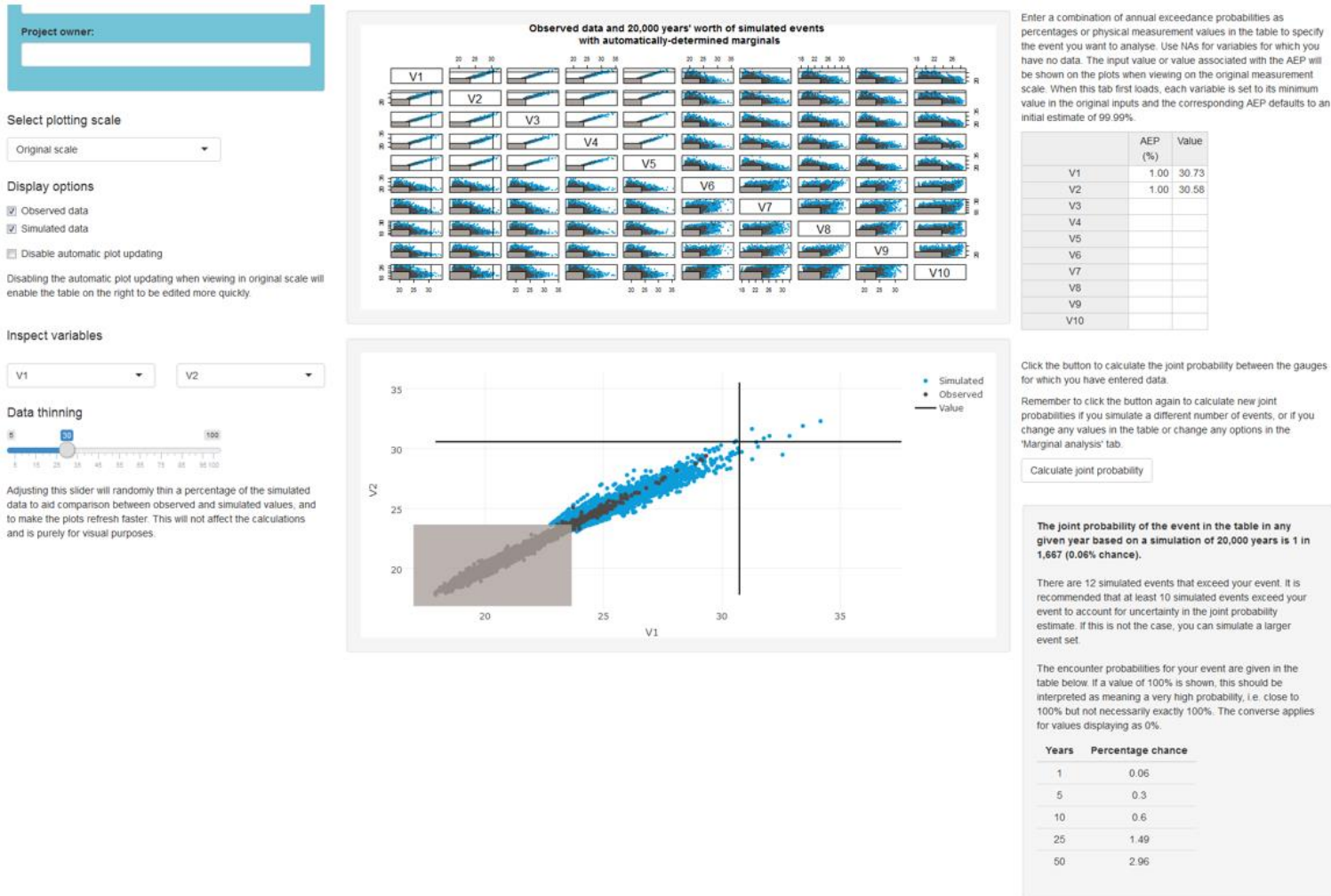


Figure 4.6 Joint probability results for the validation case study where a 1% chance has been applied to 2 AD variables

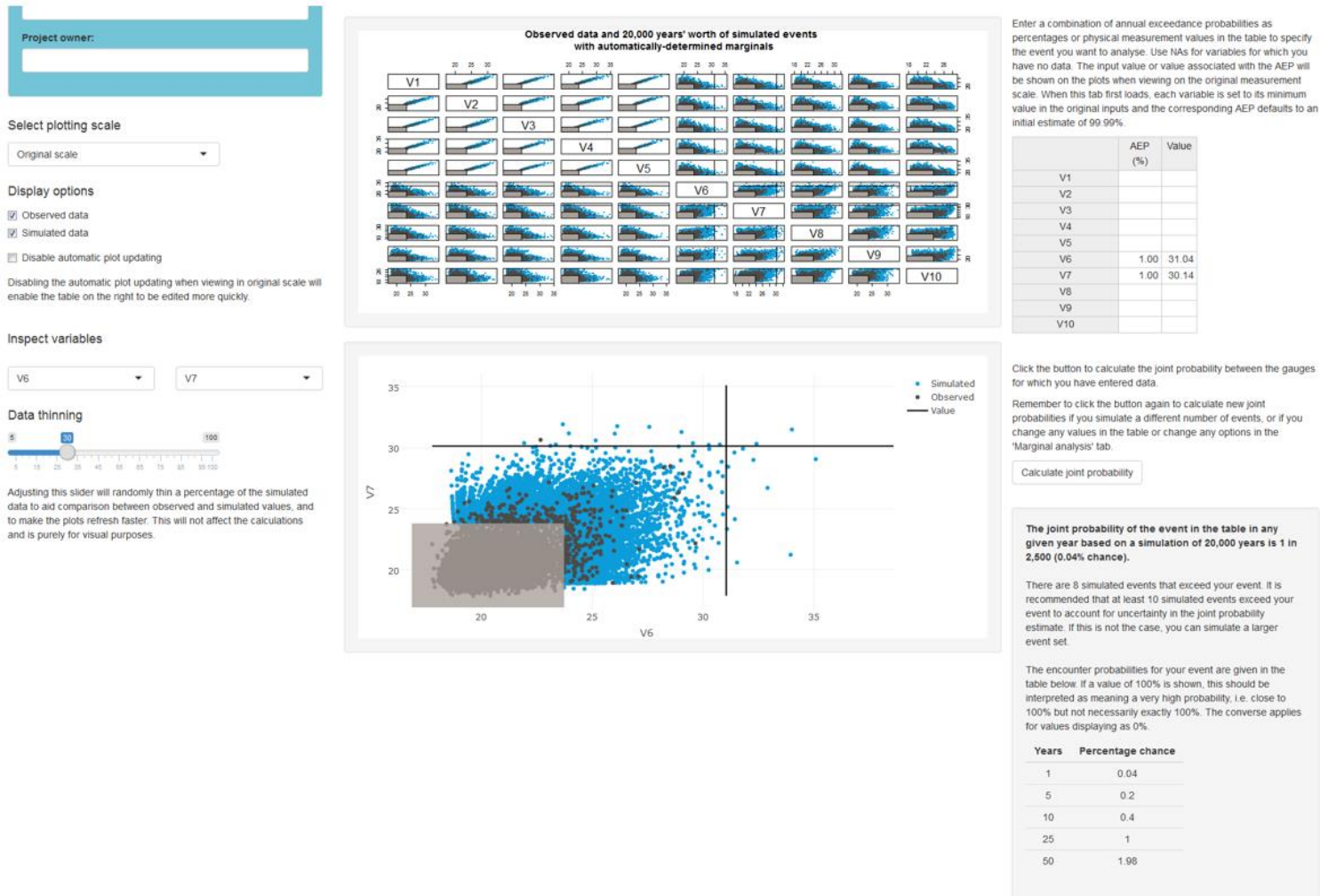
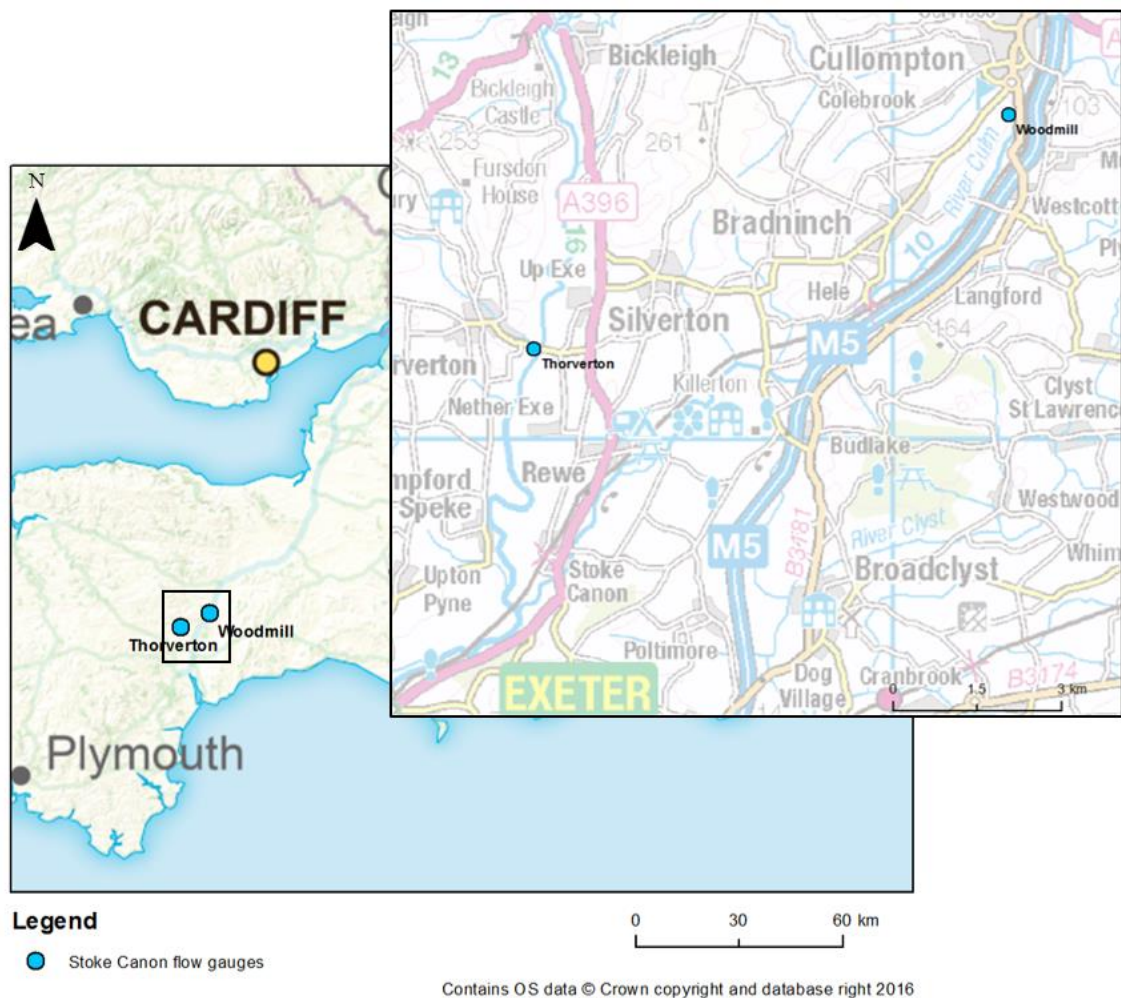


Figure 4.7 Joint probability results for the validation case study where a 1% chance has been applied to 2 AI variables

## 4.2 Case study 2: Joint probability of 2 tributaries at Stoke Canon (hydrological case study)

### 4.2.1 Context

Stoke Canon is situated at the confluence of the Rivers Exe and Culm in Devon (Figure 4.8) and is at risk of flooding from either or both rivers. During a study to improve hydrological and hydraulic models for Stoke Canon, a joint probability approach was necessary to determine flood flows on the 2 rivers that combine to produce floods with a range of specified AEPs.



**Figure 4.8** Location of flow gauges, with Stoke Canon visible towards the south-west of the inset map just north of Exeter

The combinations in the resulting joint probability matrix were simulated in a hydraulic model to determine the joint probability scenario that generated the most severe flooding event.

Suitable scenarios dominated by one of the rivers were chosen to test a range of options. This is a common hydrological application and this case study demonstrates how the MEM could be used to carry out the joint probability analysis.

## 4.2.2 Datasets

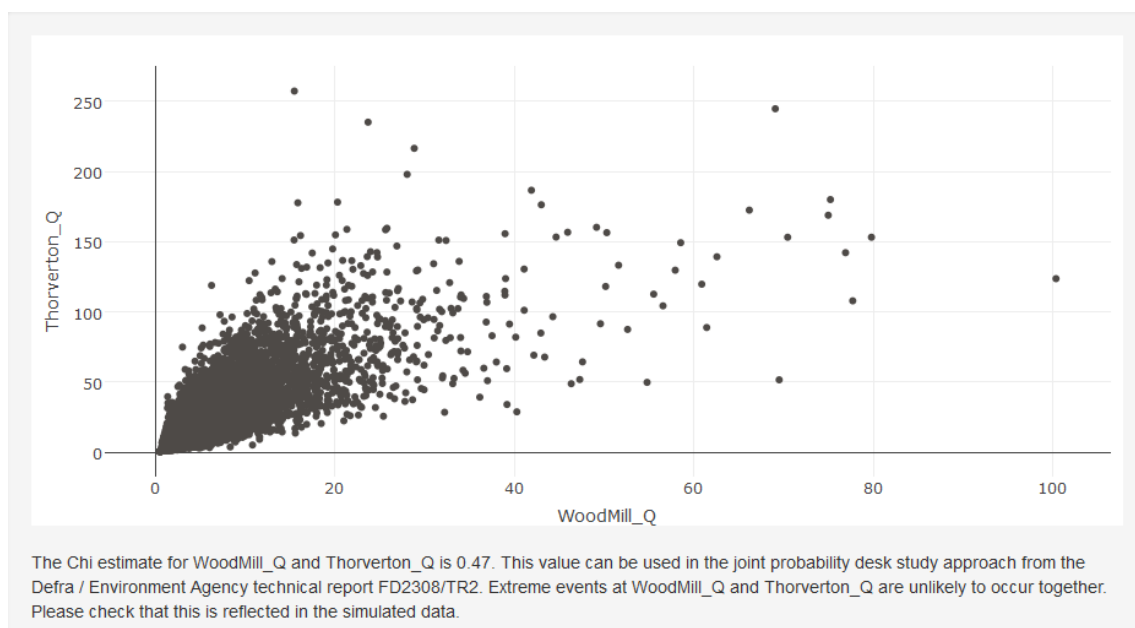
The data used in this case study are provided in 'Stoke\_Canon\_daily\_mean\_flow.csv'.

Two gauging stations with long records of flow were used in the joint probability analysis:

- Thorverton on the River Exe (1 May 1956 to 4 August 2014)
- Wood Mill on the River Culm (29 January 1962 to 4 August 2014)

The data provided were initially sampled at hourly or 15-minute time intervals. The daily mean flow values were calculated for both gauges (daily maxima could also have been used), checking that there were no missing data, and combined into one file with daily time steps starting on the later of the 2 start dates (29 January 1962). The resulting file was in the format shown in Table 3.1.

There is little dependence between these 2 gauges, with a Chi value of 0.47, although it can be seen that some of their extreme events occur together (Figure 4.9).



**Figure 4.9 Relationship between daily mean flows at Wood Mill and Thorverton**

## 4.2.3 Analysis

Over the 52.5 years of data, there were a total of 205 independent peak events that exceeded the 97.5th percentile flow for Wood Mill and 149 for Thorverton.

A 20,000-year event set was simulated and the automatically determined marginal parameters were used in the analysis. The values associated with specified return periods were viewed alongside the observed data to check that the marginal analysis looked sensible for this data.

The joint probability design scenarios of interest had the following AEPs: 1 in 2, 1 in 10, 1 in 20, 1 in 25, 1 in 50, 1 in 75, 1 in 100 and 1 in 1,000. There are any number of potential combinations of probabilities that could produce these design return periods, and different combinations can be tested in the MEM.

There is a range of combinations of interest, including scenarios dominated by one of the rivers and scenarios where both rivers have similar probabilities. This case study focused on the 1 in 50 AEP (2%) scenario and the 1 in 100 AEP (1%) scenario. A probability was selected for one of the gauges and then the other value was estimated to obtain the required scenario.

#### 4.2.4 Results

The joint probability table was used to check whether the marginal AEPs and values calculated within the MEM (Figure 4.10) looked sensible in relation to the observed data (the daily mean data are plotted in Figure 4.11).

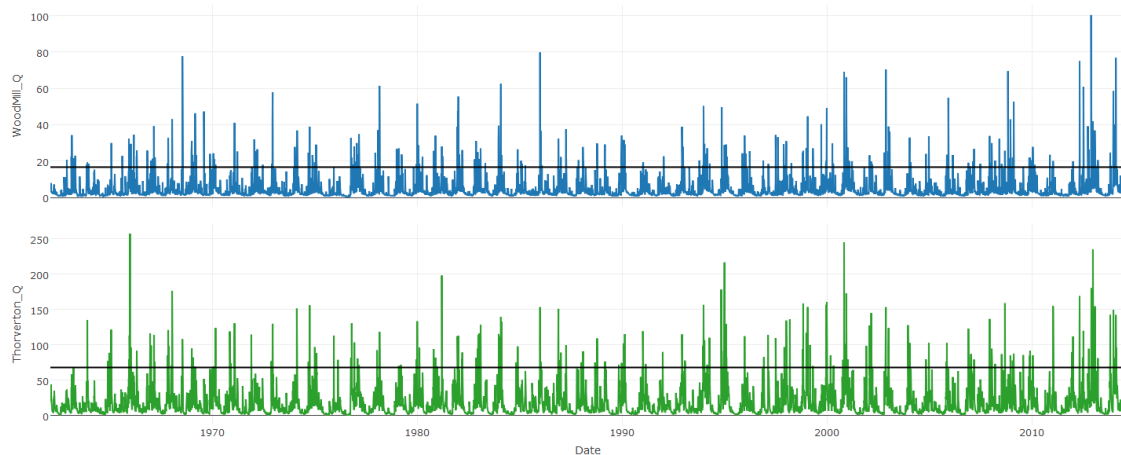
For the 1% annual chance (100-year return period), the value for Wood Mill is slightly higher than all the observed values in the 52.55-year dataset and the value for Thorverton is just below the maximum observed value. This is reasonable.

For the 2% annual chance (50-year return period), the value for Wood Mill is exceeded once in the record and the value for Thorverton is exceeded twice in the record. This is also reasonable.

(a)	AEP (%)	Value	(b)	AEP (%)	Value
WoodMill_Q	1.00	108.51	WoodMill_Q	2.00	95.43
Thorverton_Q	1.00	251.82	Thorverton_Q	2.00	236.46

**Figure 4.10 AEP and associated values (calculated within the MEM) in the table on the ‘Joint probability analysis’ tab**

Notes: The values associated with a 1% chance are given in (a) and those associated with a 2% chance are given in (b).



**Figure 4.11 Time series of daily mean data for Wood Mill and Thorverton**

Notes: The 97.5th percentile used as the threshold above which events are classed as extreme is shown as a black line at a value of  $16.7\text{m}^3\text{s}^{-1}$  for Wood Mill and  $67.8\text{m}^3\text{s}^{-1}$  for Thorverton.

The 50% annual chance value (equivalent to the median annual maximum flow, QMED) calculated by the MEM was also compared to the QMED calculated outside the MEM by finding the median annual maximum value. The values were very similar. For Wood Mill, the 50% chance value from the MEM is  $39.1\text{m}^3\text{s}^{-1}$  and the calculated QMED



value is  $36.6\text{m}^3\text{s}^{-1}$ . For Thorverton, the 50% chance value from the MEM is  $129\text{m}^3\text{s}^{-1}$  and the calculated QMED value is  $127.6\text{m}^3\text{s}^{-1}$ .

Table 4.1 provides some potential combinations for design floods with a joint AEP of 2% and 1%. The selected scenarios are comparable with those in the hydrology study, which found:

- scenarios that were dominated by the River Culm (at Wood Mill)
- scenarios that were dominated by the River Exe (at Thorverton)
- scenarios where the 2 tributaries had similar probabilities

Where different, the values identified in that work using the joint probability desk study approach developed in the Defra/Environment Agency project FD2308 (Hawkes, 2005) are also given in Table 4.1.

In general, the marginal AEPs at Wood Mill and Thorverton are higher than those obtained using the desk study approach to produce the same joint probability. The joint AEP calculated by the MEM of the marginal probabilities from the desk study approach is generally lower. However, the estimates derived using the alternative methodologies do not differ by orders of magnitude.

**Table 4.1 Potential combinations of tributary flows that would have a joint AEP of 2% or of 1%**

Target 'design' scenario AEP (return period)	MEM joint probability of combination at Wood Mill and Thorverton (by trial and error)	AEP (return period)	
		Wood Mill (Culm)	Thorverton (Exe)
2% (50-year) (Culm-dominated)	2.005% (50-year)	2% (50-year)	28% (3.6-year) 13.3% (7.5-year)
2% (50-year) (neither dominated)	1.985% (50-year)	5% (20-year) 3.3% (30-year)	6.5% (15.4-year) 6.7% (15-year)
2% (50-year) (neither dominated)	2.02% (50-year)	6.2% (16.1-year) 5% (20-year)	5% (20-year) 4% (25-year)
2% (50-year) (Exe-dominated)	2.015% (50-year)	23% (4.3-year) 10% (10-year)	2% (50-year)
1% (100-year) (Culm-dominated)	1.005% (100-year)	1% (100-year)	21.1% (4.7-year) 6.7% (15-year)
1% (100-year) (neither dominated)	1% (100-year)	2.5% (40-year) 1.25% (80-year)	3.7% (27-year) 4% (25-year)
1% (100-year) (neither dominated)	1% (100-year)	3% (33.3-year) 1.54% (65-year)	3.3% (30.3-year) 3.3% (30-year)

Target 'design' scenario AEP (return period)	MEM joint probability of combination at Wood Mill and Thorverton (by trial and error)	AEP (return period)	
		Wood Mill (Culm)	Thorverton (Exe)
1% (100-year) (neither dominated)	1% (100-year)	3.3% (30.3-year) <b>2.5% (40-year)</b>	3% (33.3-year) <b>2% (50-year)</b>
1% (100-year) (Exe-dominated)	1% (100-year)	17.3% (5.8-year) <b>5% (20-year)</b>	1% (100-year)

Notes: The values in red are those identified in the original hydrology study using the joint probability desk study approach developed in Defra/Environment Agency project FD2308 (Hawkes 2005) (where different).

### 4.3 Case study 3: Joint probability of Storm Desmond on the River Tyne (hydrological case study)

#### 4.3.1 Context

Storm Desmond affected northern England and Scotland at the beginning of December 2015. High flows and flooding were experienced in the north-east with maximum flows on the Tyne observed on 5 and 6 December.

The joint probability of the flows associated with Storm Desmond at 3 gauging stations in the Tyne catchment (Figure 4.12) was considered as part of a broader study to look into the influence of Kielder Reservoir on the joint response from the North Tyne and South Tyne.



**Figure 4.10** Location of the 3 gauging stations in the Tyne catchment for which the joint probability was calculated

### 4.3.2 Datasets

The data used in this case study are provided in 'Tyne\_daily\_max\_flow.csv'.

Flow data were provided for 3 gauging stations at 15-minute resolution. These were:

- Bywell (River Tyne, downstream of the confluence of the North Tyne and South Tyne)
- Reaverhill (North Tyne)
- Haydon Bridge (South Tyne)

The data were checked for missing values and for missing dates. Across the gauges, there were 13.5 years with no missing data (21 September 2002 to 3 May 2016) and this section of each record was used in the joint probability analysis.

The daily maximum values were extracted for all 3 gauges and combined into one file with daily time steps. The resulting file was in the format shown in Table 3.1.

The daily maximum flow is plotted in Figure 4.13. It can be seen that the largest peaks often occur together, particularly at Bywell and either of the other gauges (the Chi values for these combinations are both greater than 0.6). The Chi estimate for Reaverhill and Haydon Bridge is lower at 0.45, and so extreme events at these 2 gauges are less likely to occur together.



**Figure 4.11 Time series of daily maximum data for Bywell, Reaverhill and Haydon Bridge**

Notes: The 97.5th percentile used as the threshold above which events are classed as extreme is shown as a black line at a value of  $367.1\text{m}^3\text{s}^{-1}$  for Bywell,  $169.6\text{m}^3\text{s}^{-1}$  for Reaverhill and  $200.1\text{m}^3\text{s}^{-1}$  for Haydon Bridge.

The maximum observed flow values from 4 December to 6 December at each gauge are given in Table 4.2.

**Table 4.2 Maximum observed flow values for Storm Desmond at each gauge**

Date	Bywell flow ( $\text{m}^3\text{s}^{-1}$ )	Reaverhill flow ( $\text{m}^3\text{s}^{-1}$ )	Haydon Bridge ( $\text{m}^3\text{s}^{-1}$ )
4 December 2015	744.203	393.991	305.289
5 December 2015	1620.776	688.470	914.598
6 December 2015	1622.048	715.521	603.764

### 4.3.3 Analysis

Over the 13.5 years of data, there were a total of 53 independent events for Bywell, 52 for Reaverhill and 63 for Haydon Bridge, based on the 97.5th percentile.

Given that the event being analysed is one of the highest in the record and so likely to have a low chance of occurring, a 50,000-year event set was simulated. The automatically determined marginal parameters were initially used in the analysis. The parameters from a generalised logistic distribution were also calculated as a comparison. The AEP curves on the 'Marginal analysis' tab were viewed alongside the observed data to check that the marginal analysis looked sensible for these data.

The joint probability of the flow values on 5 December and 6 December was calculated using both the automatically determined marginal parameters (from a generalised Pareto distribution) and the user-supplied marginal parameters (from a generalised logistic distribution).

### 4.3.4 Results

The AEP curve on the 'Marginal analysis' tab and the joint probability table were used to check whether the marginal AEPs and values (Figure 4.14) looked sensible in relation to the observed data. The daily maximum data are plotted in Figure 4.13.

For the 1% annual chance (100-year return period), the value at all gauges is higher than all the observed values in the 13.5-year dataset. This could be expected for such a short record.

For the 2% annual chance (50-year return period), the value at all gauges is also slightly higher than the observed values, which is also reasonable.

Flow values using automatically determined marginals			Flow values using user-supplied marginals		
<b>(a)</b>			<b>(b)</b>		
	AEP (%)	Value		AEP (%)	Value
Bywell	1.00	1871.46	Bywell	1.00	2099.62
Reaverhill	1.00	912.92	Reaverhill	1.00	898.13
Haydon.Bridge	1.00	1245.61	Haydon.Bridge	1.00	1311.83
<b>(c)</b>			<b>(d)</b>		
	AEP (%)	Value		AEP (%)	Value
Bywell	2.00	1699.59	Bywell	2.00	1805.23
Reaverhill	2.00	799.46	Reaverhill	2.00	771.49
Haydon.Bridge	2.00	1109.95	Haydon.Bridge	2.00	1127.63

**Figure 4.12 AEP and associated values in the table on the 'Joint probability analysis' tab**

Notes: Flow values associated with a 1% annual chance are given in (a) and (b) and those associated with a 2% annual chance are given in (c) and (d). The AEP-value relationship determined from the automatic application of a generalised Pareto distribution was used to calculate the flow values in (a) and (c). The relationship determined from the parameters of a generalised logistic distribution was used to calculate the flow values in (b) and (d).

The 50% annual chance value (equivalent to the median annual maximum flow, QMED) calculated by the MEM was also compared with the QMED calculated outside the MEM by finding the median annual maximum value (Table 4.3).

The values were very similar, particularly for Reaverhill and Haydon Bridge using the automatically determined marginal parameters. The 50% annual chance flow value for Bywell was closer to QMED when using the user-supplied marginal parameters. Overall, the 50% chance flow values calculated using the automatically determined marginals are closer to the QMED values.

**Table 4.3 Comparison of the median annual maximum flow values (QMED) with the 50% annual chance flow values from fitting a generalised Pareto distribution to the data (automatically done in the MEM) and from fitting a**

**generalised logistic distribution to the data (using the user-supplied marginal parameters)**

Site (variable)	QMED (median annual maximum value) (m <sup>3</sup> s <sup>-1</sup> )	50% annual chance (2-year return period) flow (m <sup>3</sup> s <sup>-1</sup> )	
		Using automatically determined marginal parameters	Using user-supplied marginal parameters
Bywell	773.8	811.6	751
Reaverhill	337.2	337.7	318
Haydon Bridge	490.6	490.6	468

*Joint probability of peak flows on 5 December 2015*

**Using automatically determined marginals**

Inputting the peak flows for 5 December 2015 (Table 4.2) into the table on the 'Joint probability analysis' tab, with the selection to use the automatically determined marginals, gave an annual joint probability of 1 in 140 (0.7% chance). The large simulation ensured that numerous simulated events (357) exceeded the event on 5 December 2015 at all gauges to help constrain uncertainty in the results. The results are shown in Figure 4.15.

**Using user-supplied marginals**

Carrying out the same analysis, but using the user-supplied marginals, gave an annual joint probability of 1 in 64 (1.6% chance). An even greater number of simulated events (784) exceeded the event on 5 December 2015 at all gauges. The results are shown in Figure 4.14.

*Joint probability of peak flows on 6 December 2015*

**Using automatically determined marginals**

Inputting the peak flows for 6 December 2015 (Table 4.2) with the selection to use the automatically determined marginals gave an annual joint probability of 1 in 105 (0.96% chance). Again, there were plenty of simulated events (480) that exceeded the event on 6 December 2015 at all gauges. The results are shown in Figure 4.17.

**Using user-supplied marginals**

Inputting the peak flows for 6 December 2015 (Table 4.2) with the selection to use the user-supplied marginals gave an annual joint probability of 1 in 51 (1.9% chance). There were 978 simulated events that exceeded the event on 6 December 2015 at all gauges. The results are shown in Figure 4.16.

Project owner:

Select plotting scale

Original scale

Display options

- Observed data
- Simulated data
- Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

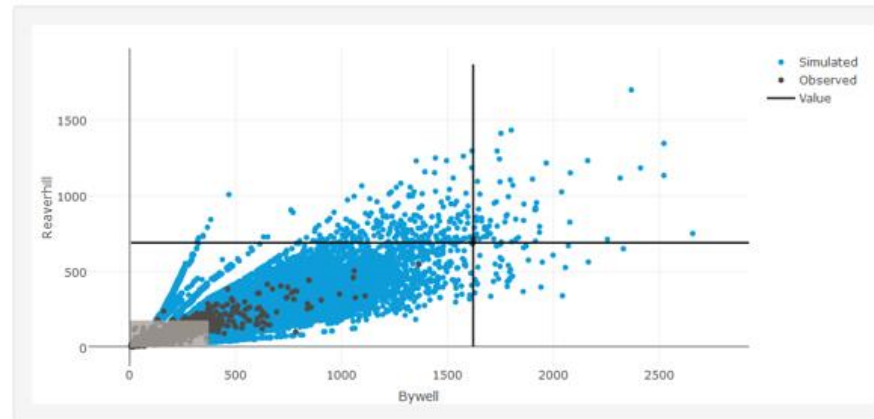
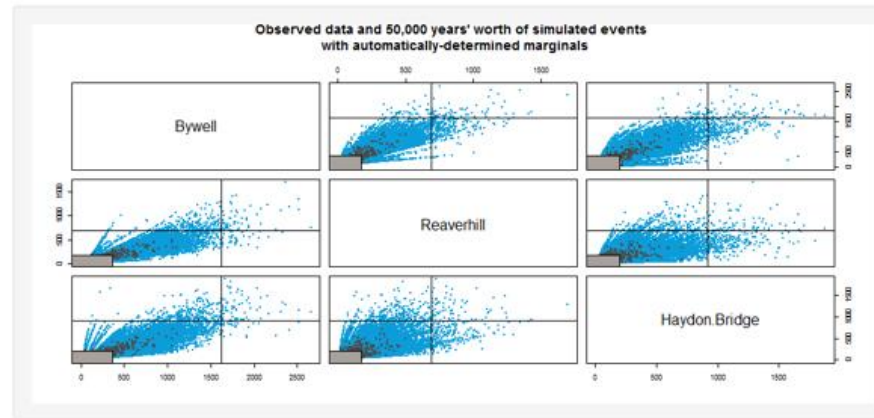
Inspect variables

Bywell Reaverhill

Data thinning

5 10 100 500 1000

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
Bywell	2.70	1620.78
Reaverhill	4.10	688.47
Haydon Bridge	5.60	914.60

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

The joint probability of the event in the table in any given year based on a simulation of 50,000 years is 1 in 140 (0.712% chance).

There are 357 simulated events that exceed your event. It is recommended that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set.

The encounter probabilities for your event are given in the table below. If a value of 100% is shown, this should be interpreted as meaning a very high probability, i.e. close to 100% but not necessarily exactly 100%. The converse applies for values displaying as 0%.

Years	Percentage chance
1	0.71
5	3.51
10	6.9
25	16.36
50	30.04

Figure 4.13 Joint probability analysis for the peak flows on 5 December 2015 using the automatically determined marginal parameters

Project owner:

Select plotting scale  
 Original scale

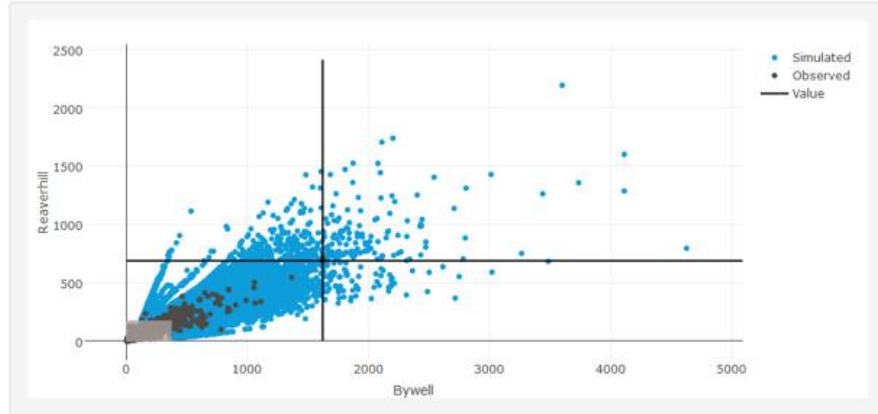
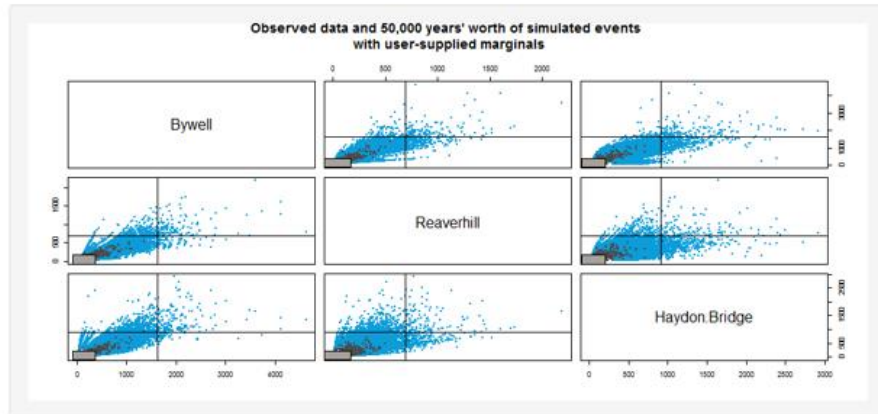
Display options  
 Observed data  
 Simulated data  
 Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

Inspect variables  
 Bywell Reaverhill

Data thinning  
 5 10 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
Bywell	3.20	1620.78
Reaverhill	3.30	688.47
Haydon Bridge	5.10	914.60

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

**The joint probability of the event in the table in any given year based on a simulation of 50,000 years is 1 in 64 (1.562% chance).**

There are 784 simulated events that exceed your event. It is recommended that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set.

The encounter probabilities for your event are given in the table below. If a value of 100% is shown, this should be interpreted as meaning a very high probability, i.e. close to 100% but not necessarily exactly 100%. The converse applies for values displaying as 0%.

Years	Percentage chance
1	1.56
5	7.57
10	14.57
25	32.54
50	54.49

Figure 4.14 Joint probability analysis for the peak flows on the 5 December 2015 using the user-supplied marginal parameters



Project owner:

Select plotting scale

Original scale

Display options

- Observed data
- Simulated data
- Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

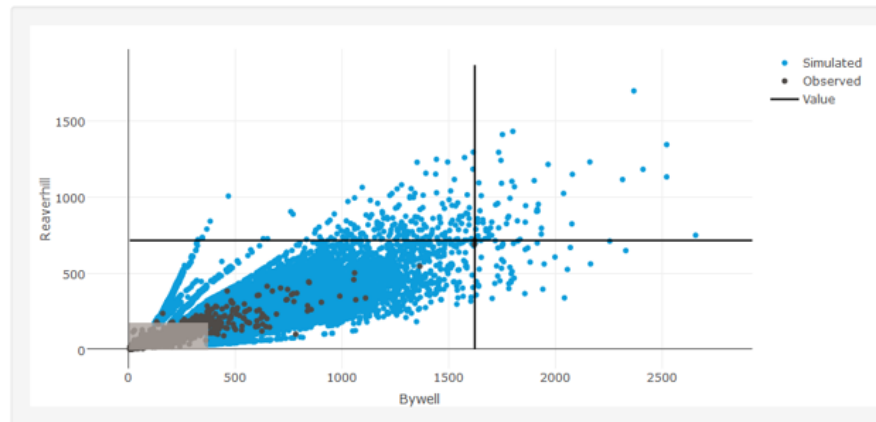
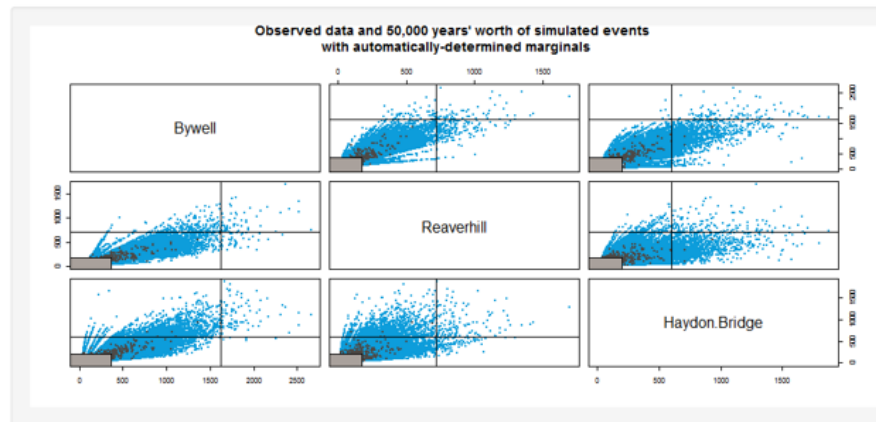
Inspect variables

Bywell      Reaverhill

Data thinning

5 10 25 35 45 55 75 85 90 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
Bywell	2.70	1622.05
Reaverhill	3.40	715.52
Haydon Bridge	29.30	603.76

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

**The joint probability of the event in the table in any given year based on a simulation of 50,000 years is 1 in 105 (0.956% chance).**

There are 480 simulated events that exceed your event. It is recommended that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set.

The encounter probabilities for your event are given in the table below. If a value of 100% is shown, this should be interpreted as meaning a very high probability, i.e. close to 100% but not necessarily exactly 100%. The converse applies for values displaying as 0%.

Years	Percentage chance
1	0.96
5	4.69
10	9.16
25	21.35
50	38.14

Figure 4.15 Joint probability analysis for the peak flows on 6 December 2015 using the automatically determined marginal parameters

Project owner:

Select plotting scale

Original scale

Display options

- Observed data
- Simulated data
- Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

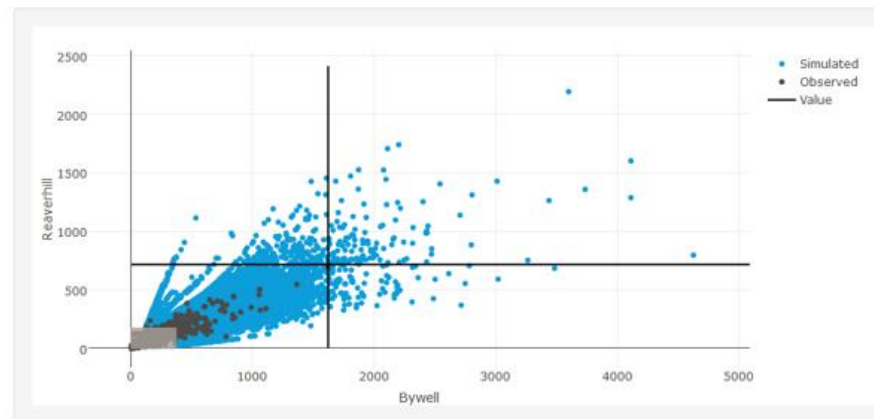
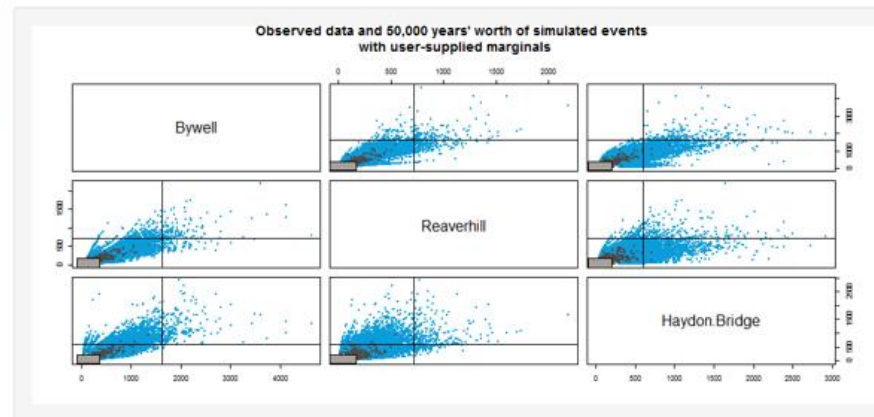
Inspect variables

Bywell Reaverhill

Data thinning

0 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
Bywell	3.20	1622.05
Reaverhill	2.80	715.52
Haydon Bridge	25.20	603.76

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

**The joint probability of the event in the table in any given year based on a simulation of 50,000 years is 1 in 51 (1.946% chance).**

There are 978 simulated events that exceed your event. It is recommended that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set.

The encounter probabilities for your event are given in the table below. If a value of 100% is shown, this should be interpreted as meaning a very high probability, i.e. close to 100% but not necessarily exactly 100%. The converse applies for values displaying as 0%.

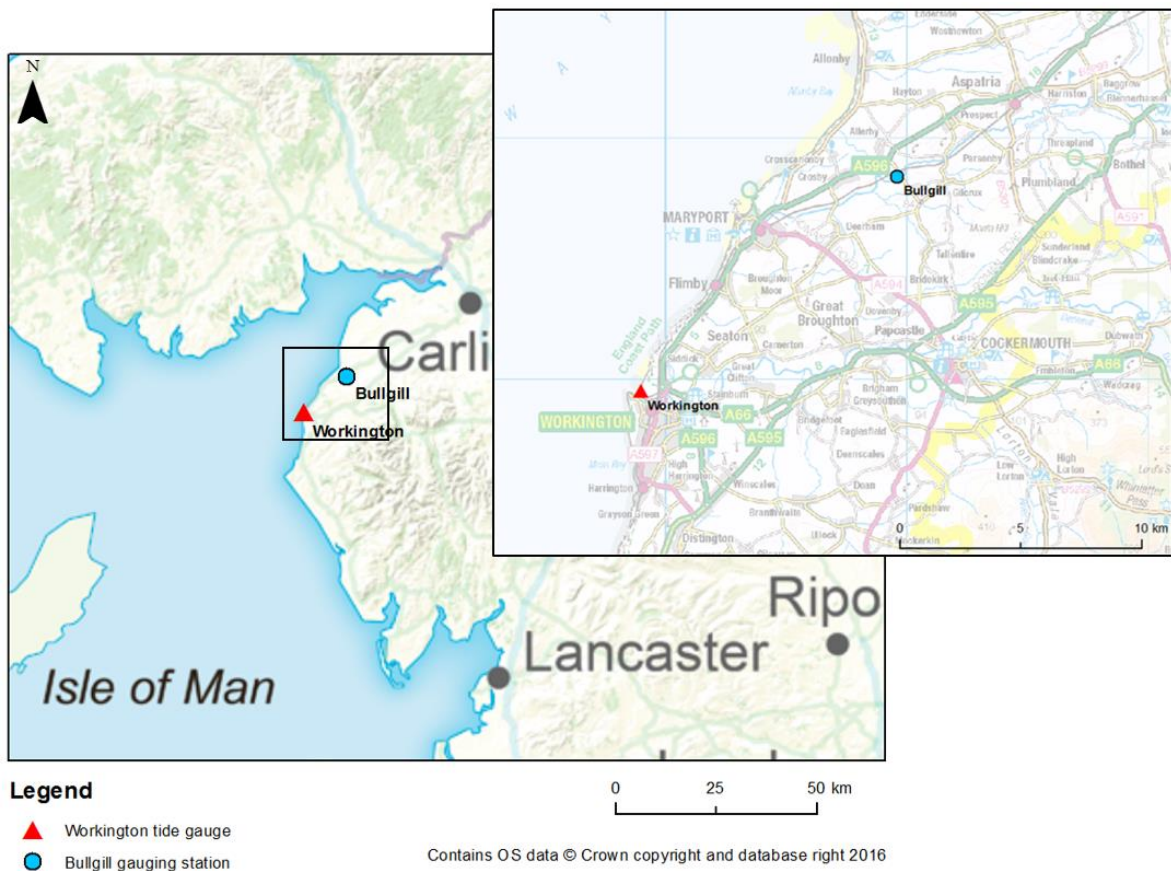
Years	Percentage chance
1	1.95
5	9.36
10	17.84
25	38.82
50	62.57

Figure 4.16 Joint probability analysis for the peak flows on 6 December 2015 using the user-supplied marginal parameters

## 4.4 Case study 4: Joint probability of high flow on the River Ellen and high tidal surges at Workington (estuary case study)

### 4.4.1 Context

The River Ellen in Cumbria flows out into the sea at Maryport (Figure 4.19). An assessment is required of the joint probability of the fluvial–tidal conditions in order to develop flood scenarios that can then be run through a hydraulic model. This case study demonstrates how the MEM could be used to investigate the impact of joint probability for flood risk at Maryport.



**Figure 4.17** Location of the Bullgill flow gauging station and the Workington tide gauge with Maryport situated at the mouth of the River Ellen

### 4.4.2 Datasets

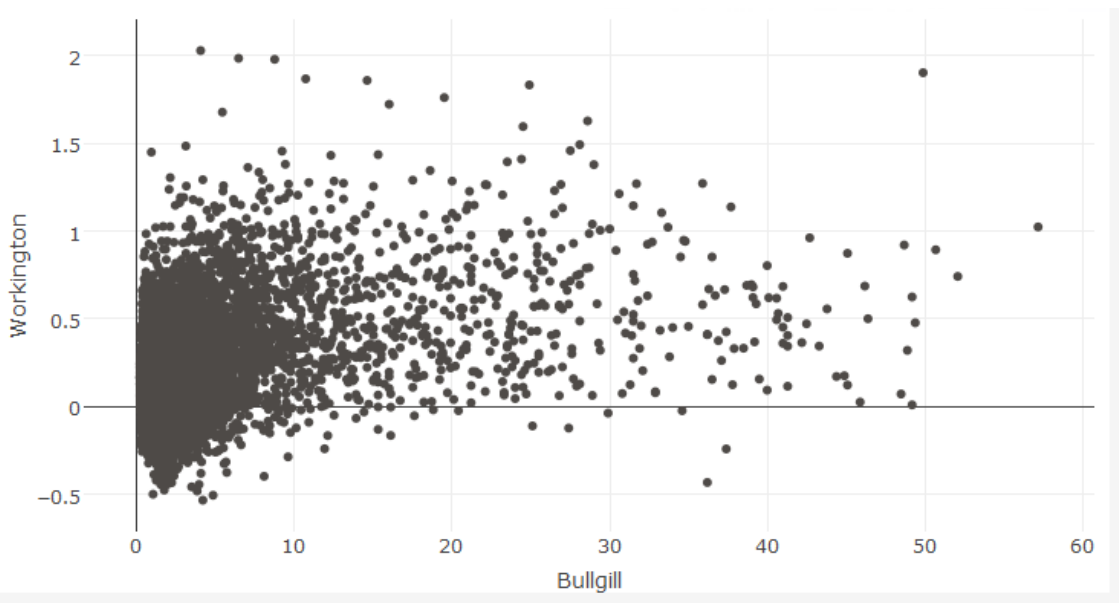
The data used in this case study are provided in 'Bullgill\_and\_Workington\_daily\_max.csv'.

The nearest Class A tide gauge recording 15-minute surge values is at Workington, 8.5km down the coast from Maryport. Daily maximum values of surge residual were

extracted.<sup>2</sup> Missing data and data flagged as ‘improbable’ were infilled with non-extreme data, so that they are not used in the analysis. The period of record for this dataset is 5 February 1992 to 30 September 2016.

The only flow gauging station on the River Ellen is at Bullgill. The record starts on 1 January 1976 and ends on 22 November 2016. The daily maximum flows were extracted and patches of missing data were found to exist until September 1993, which is close to the start date of the tidal data. Therefore, data before this date were removed and the Bullgill data were combined with the Workington data into one .csv file with daily time steps starting on 7 September 1993 and finishing on 30 September 2016. The resulting file was in the format shown in Table 3.1.

Figure 4.18 shows the relationship between the 2 datasets. It can be seen that they are not highly dependent; the Chi value is only 0.08 and so extreme events are not likely to occur together.



**Figure 4.18 Relationship between daily maximum fluvial flow at Bullgill and tide surge residual at Workington**

#### 4.4.3 Analysis

Over the 23 years of data, there were a total of 107 independent events for Bullgill and 85 for Workington, based on the 97.5th percentile.

A 10,000-year event set was simulated and the automatically determined marginal parameters were used in the analysis. The values associated with specified return periods were viewed alongside the observed data to check that the marginal analysis looked sensible for this data.

Given that there is little dependence between the 2 gauges, the joint probability was calculated for a range of combinations of marginal AEPs to help understand whether joint probability is an important contributor to risk at Maryport.

---

<sup>2</sup> The data were supplied by the British Oceanographic Data Centre as part of the function of the National Tidal & Sea Level Facility, hosted by the Proudman Oceanographic Laboratory and funded by the Environment Agency and the Natural Environment Research Council.

#### 4.4.4 Results

The joint probability table was used to check whether the marginal AEPs and values calculated within the MEM (Figure 4.21) looked sensible in relation to the observed data (the daily maximum data are plotted in Figure 4.22).

For the 1% annual chance (100-year return period), the value for Bullgill is slightly higher than all the observed values in the 23-year dataset (the maximum recorded value is  $57.2\text{m}^3\text{s}^{-1}$  and there are no other values above  $50\text{m}^3\text{s}^{-1}$ ). The value for Workington is also just above the maximum observed value (2.03m). These are both reasonable.

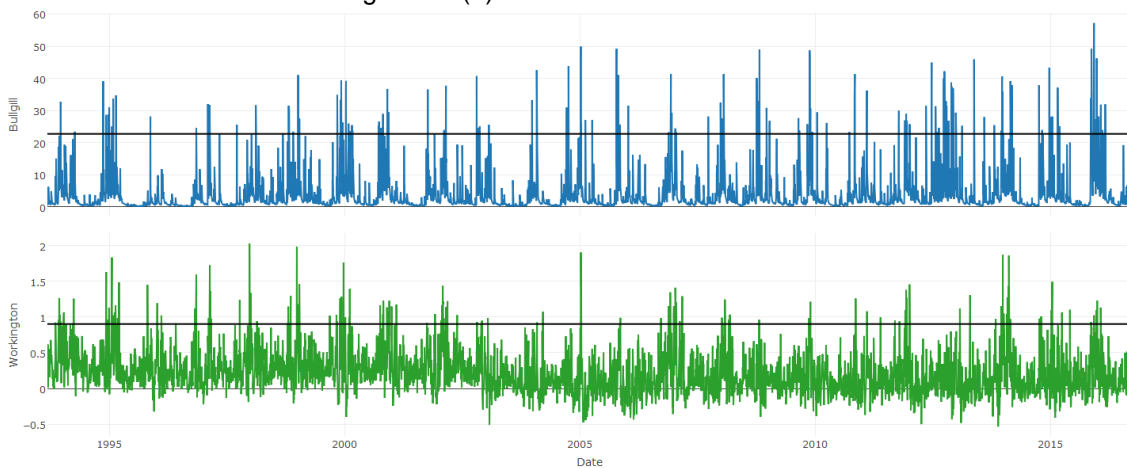
For the 2% annual chance (50-year return period), the value for Bullgill is exceeded once in the record. The value for Workington is slightly higher than the maximum value in the record and so these are also reasonable.

Figure 4.21 shows an apparent shift in the Workington data towards the end of 2002. The metadata state that the gauge was relevelled in 2002 and, on investigation, there is a difference in the average value of the periods before and after the releveling of 0.18m. This would need to be adjusted for use in a real application.

(a)	AEP (%)	Value	(b)	AEP (%)	Value
Bullgill	1.00	57.55	Bullgill	2.00	56.01
Workington	1.00	2.29	Workington	2.00	2.18

**Figure 4.21 AEP and associated values (calculated within the MEM) in the table on the ‘Joint probability analysis’ tab**

Notes: The values associated with a 1% chance are given in (a) and those associated with a 2% chance are given in (b).



**Figure 4.22 Time series of daily maximum data for Bullgill and Workington**

Notes: The 97.5th percentile used as the threshold above which events are classed as extreme is shown as a black line at a value of  $22.7\text{m}^3\text{s}^{-1}$  for Bullgill and 0.9022m for Workington.

For both gauges, the 50% annual chance value calculated by the MEM was compared with the median annual maximum value calculated outside the MEM. These values should be equivalent and they were found to be very similar. For Bullgill, the 50% chance flow from the MEM is  $40.65\text{m}^3\text{s}^{-1}$  and the calculated median annual maximum

flow, QMED, is  $40.7\text{m}^3\text{s}^{-1}$ . For Workington, the 50% chance value for surge residual from the MEM is 1.42m and the calculated median annual maximum value is 1.44m.

Project owner:

Select plotting scale  
 Original scale

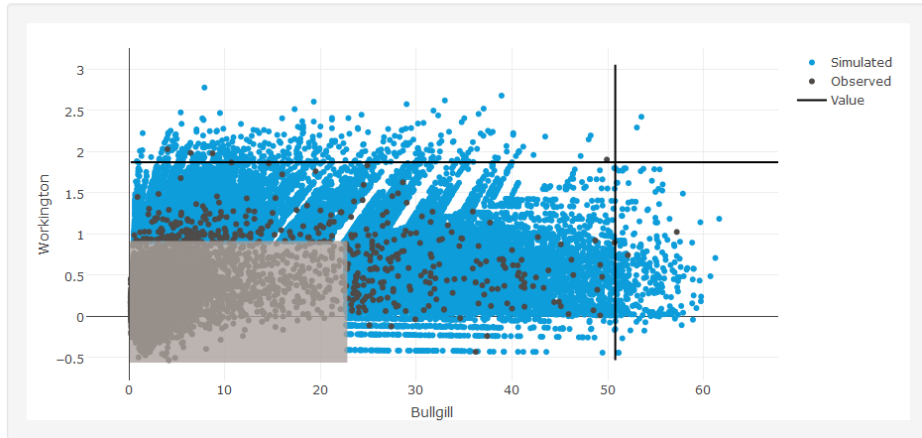
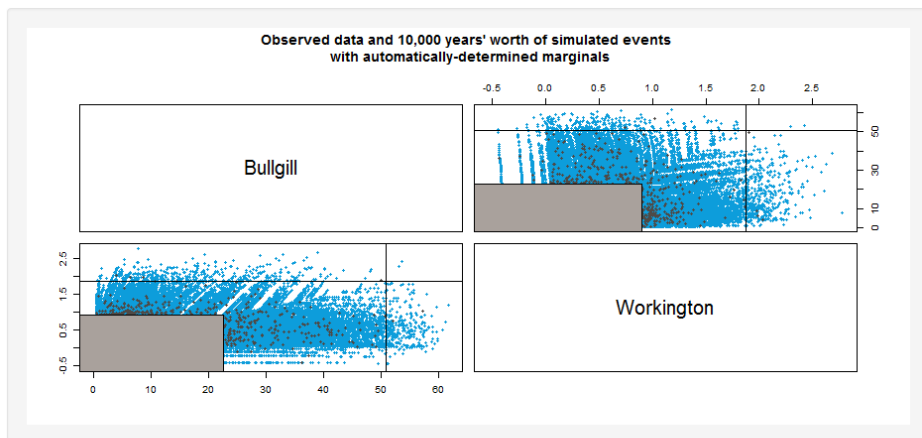
Display options  
 Observed data  
 Simulated data  
 Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

Inspect variables  
 Bullgill  
 Workington

Data thinning  
 5 30 100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of annual exceedance probabilities as percentages or physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. The input value or value associated with the AEP will be shown on the plots when viewing on the original measurement scale. When this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%.

	AEP (%)	Value
Bullgill	10.00	50.78
Workington	10.00	1.87

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

**The joint probability of the event in the table in any given year based on a simulation of 10,000 years is 1 in 909 (0.11% chance).**

There are 11 simulated events that exceed your event. It is recommended that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set.

The encounter probabilities for your event are given in the table below. If a value of 100% is shown, this should be interpreted as meaning a very high probability, i.e. close to 100% but not necessarily exactly 100%. The converse applies for values displaying as 0%.

Years	Percentage chance
1	0.11
5	0.55
10	1.09
25	2.71
50	5.35

Figure 4.19 Joint probability analysis for the scenario where both gauges have an AEP of 10%

Figure 4.23 shows the joint probability analysis for the scenario where both Bullgill and Workington have an AEP of 10%. The annual joint probability of the event is 1 in 909 or 0.11%. These 2 marginal AEPs, which are not particularly extreme in themselves, already give an annual joint probability of just over 1 in 1,000, which corresponds to an extreme planning assumption.

Given that it is unlikely that planning is required for a more extreme scenario, combinations of AEPs that are larger than 10% were investigated to gain a better understanding of how joint probability is influencing risk at Maryport. The plots do not suggest that joint probability is going to be a major factor. Combinations of AEPs at the 2 gauges and their joint probabilities are given in Table 4.4. The values indicate that joint probability is not an important factor for risk at Maryport with the joint AEPs being much lower than those of Bullgill and Workington individually, with even combinations of relatively frequent river and surge peaks (for example, AEPs of 50% and 25% respectively) having a joint probability of less than 1% in any year.

**Table 4.4 Joint probability for a range of combinations of AEPs**

AEP (return period)		
Bullgill flow gauge (River Ellen)	Workington tide gauge	Joint probability
10% (10-year)	10% (10-year)	0.11% (909-year)
10% (10-year)	25% (4-year)	0.31% (323-year)
10% (10-year)	50% (2-year)	0.51% (196-year)
25% (4-year)	10% (10-year)	0.24% (417-year)
25% (4-year)	25% (4-year)	0.63% (159-year)
25% (4-year)	50% (2-year)	1.21% (83-year)
50% (2-year)	10% (10-year)	0.43% (233-year)
50% (2-year)	25% (4-year)	0.85% (118-year)
50% (2-year)	50% (2-year)	1.87% (53-year)

## 4.5 Case study 5: Demonstrating a spatial joint analysis of offshore extremes for the north Wales coast

### 4.5.1 Context

This case study uses a subset of data from offshore variables (surge, wind and wave height) based on extracted peak surge levels. It is intended to demonstrate the handling of spatially dependent coastal variables within a joint probability analysis.

### 4.5.2 Datasets

The data used in this case study are provided in 'Offshore\_extremes\_north\_Wales.csv'.



A dataset was provided by HR Wallingford containing skew surge, wave height and wind speed at 2 locations covering north Wales (JP26 and JP27) (Figure 4.24).

The wave and wind data are hindcast from the WaveWatch III model and these data are 3-hourly. The water level dataset, from which the skew surge was derived, contains values every 15 minutes.

HR Wallingford declustered the data following a similar approach to JOIN-SEA, but using skew surge rather than high tide and extracting the concurrent wind and wave values (after lag).

The declustered dataset was derived from 34.4 years of data and this value has been entered into the MEM on the 'Input data' tab.



**Figure 4.20 Location of JP26 and JP27**

### 4.5.3 Analysis

A 100,000-year event set was simulated, given that the probability of very extreme events is being investigated. Marginal models were not automatically determined by the MEM given that the observations were sampled at peak surge events – and so it is not possible to fit a model that truly represents the data. No user-supplied marginal models were specified and so the analysis was based on physical measurement values.

Given recent experiences of unprecedented events, the joint probability of the largest observed surges at the 2 locations was calculated to find the chance of the 2 locations concurrently experiencing an event that is larger than any previously recorded. This was then extended to find the chance of that event coming in combination with the wind speeds and wave heights that were associated with those observed large surge events.

Table 4.5 shows the maximum recorded skew surge values at the 2 locations in the observed dataset and their associated wave heights and wind speeds (at the same location).

**Table 4.5 Values associated with the maximum recorded skew surge event at both locations (JP26 and JP27)**

Variable	Observed value in the data sampled at the most extreme peak surge event for each location
<b>Maximum skew surge event at JP26</b>	
SkewSurge.JP26	1.774m
WaveHeight.JP26	3.666m
WindSpeed.JP26	20.589ms <sup>-1</sup>
<b>Maximum skew surge event at JP27</b>	
SkewSurge.JP27	1.039m
WaveHeight.JP27	4.042m
WindSpeed.JP27	16.834ms <sup>-1</sup>

#### 4.5.4 Results

Entering the highest observed values for skew surge at both locations (Figure 4.25) gave an annual joint probability of 1 in 1,053 (0.095%). The probability of experiencing an extreme skew surge event larger than any previously recorded in both locations is therefore low.

Because the observations have been sampled conditional on an extreme surge event occurring, there is only a small reduction in the joint probability when considering the concurrent maximum wind speed and wave height in each location alongside the largest observed surge (that is, entering all values in Table 4.5 into the table on the 'Joint probability analysis' tab). The annual joint probability of this event is 1 in 1,176 (0.085%).

By comparison, when each location is examined individually, the annual joint probability of the concurrent values at the maximum skew surge event over the 3 variables is 1 in 191 (0.523%) for JP26 and 1 in 297 (0.337%) for JP27. There is therefore a noticeable difference between the chance of the joint events occurring individually at each location and together at both locations.

Project owner:

Select plotting scale

Standardised scale

Display options

Observed data

Simulated data

Disable automatic plot updating

Disabling the automatic plot updating when viewing in original scale will enable the table on the right to be edited more quickly.

Inspect variables

SkewSurge.JP26

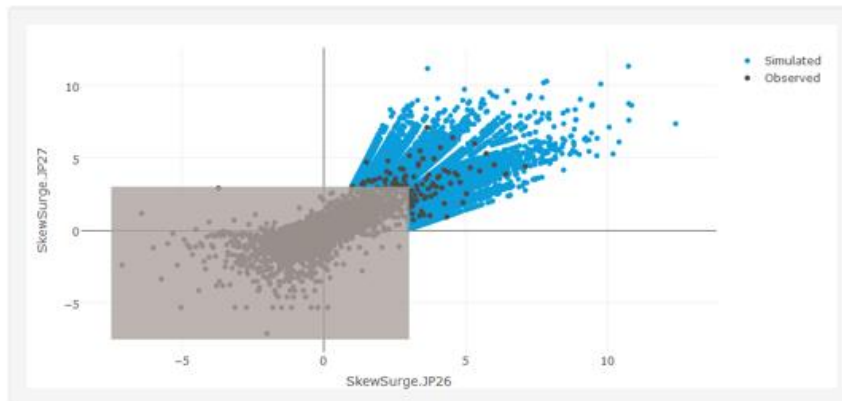
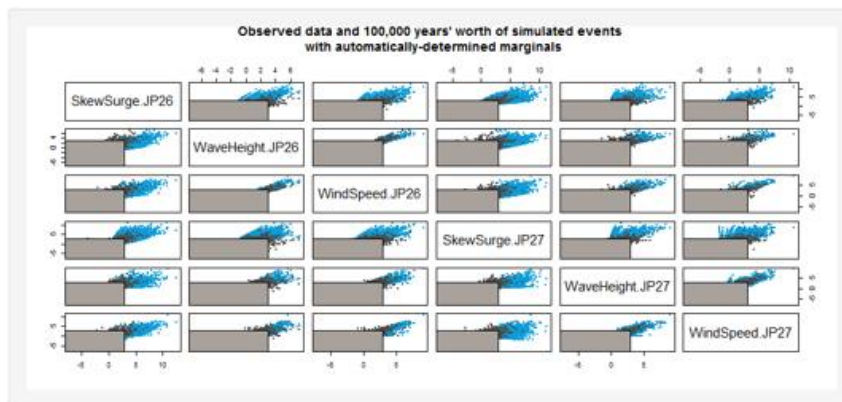
SkewSurge.JP27

Data thinning

10

100

Adjusting this slider will randomly thin a percentage of the simulated data to aid comparison between observed and simulated values, and to make the plots refresh faster. This will not affect the calculations and is purely for visual purposes.



Enter a combination of physical measurement values in the table to specify the event you want to analyse. Use NAs for variables for which you have no data. If you have specified a marginal model then you can also see and set the corresponding AEPs (when this tab first loads, each variable is set to its minimum value in the original inputs and the corresponding AEP defaults to an initial estimate of 99.99%).

	AEP (%)	Value
SkewSurge.JP26		1.77
WaveHeight.JP26		
WindSpeed.JP26		
SkewSurge.JP27		1.04
WaveHeight.JP27		
WindSpeed.JP27		

Click the button to calculate the joint probability between the gauges for which you have entered data.

Remember to click the button again to calculate new joint probabilities if you simulate a different number of events, or if you change any values in the table or change any options in the 'Marginal analysis' tab.

Calculate joint probability

The joint probability of the event in the table in any given year based on a simulation of 100,000 years is 1 in 1,053 (0.095% chance).

There are 95 simulated events that exceed your event. It is recommended that at least 10 simulated events exceed your event to account for uncertainty in the joint probability estimate. If this is not the case, you can simulate a larger event set.

The encounter probabilities for your event are given in the table below. If a value of 100% is shown, this should be interpreted as meaning a very high probability, i.e. close to 100% but not necessarily exactly 100%. The converse applies for values displaying as 0%.

Years	Percentage chance
1	0.1
5	0.47
10	0.95
25	2.35
50	4.64

Figure 4.21 Joint probability analysis for the highest recorded skew surges at JP26 and JP27

## 5 FAQs

### **Q: Why should the input data not contain missing values and how can I deal with missing data?**

The joint probability method requires pairs of concurrent observations to be compared, otherwise no information is available about dependence. Time series with missing data will therefore require those periods to be infilled (for example, by hydrological continuous simulation modelling or regression models) before use in the joint probability analysis.

### **Q: Why is there a limitation of 10 variables?**

The choice of 10 dimensions was agreed as a compromise between allowing a flexible exploration of joint probabilities and the computational demands of working with higher dimensional spaces, which increase dramatically with additional dimensions.

### **Q: Can the MEM provide a range of combinations for a specified joint probability?**

This is not a feature of the MEM and would require much more involved computation, since a very large number of combinations of marginal values could exist for any one, specified joint probability.

### **Q: Can I specify an AEP at one site and simulate a scenario at the other gauges?**

The MEM does not provide a scenario based on an AEP being specified at one site. There are likely to be many potential scenarios and these can be explored outside the MEM using the exported simulated event set. The relevant column can be sorted (for example, in Microsoft® Excel) and simulated events above the required threshold can be selected.

### **Q: What if I get an unrealistic joint probability?**

An unrealistic joint probability value may occur where extreme events are not easily defined, such as in chalk catchments, where an extreme event may last for months. This affects the dependence structure and hence the joint probability. The MEM may not be suitable for this type of data (also see next question).

### **Q: Can the MEM handle all types of hydrological data?**

The MEM can handle a number of hydrological datasets; it assumes that, for a given variable, the duration of extreme events will last for a maximum of a week. This choice of window of a week is consistent with the analysis performed in the spatial joint probability for FCRM and NRA work. If events typically last longer than a week, the MEM will be insufficient in capturing the extremal behaviour of these time series.

An example of a hydrological time series that would require further interpretation for use with the MEM is data from chalk catchments; important characteristics of these data are that the catchments are responding slowly and events are hard to define from solely analysing the hydrograph. If the MEM is used to model data from a chalk catchment, the resulting joint probability calculation may be inappropriate or not a good descriptor of the particular sources of flood risk in the catchment. A likely reason for this is that the return periods for the data are poorly estimated or that 'extreme events' are not well-defined.

One potential solution for modelling chalk catchments is to de-trend the data before performing a joint probability analysis (for example, removing any seasonality from the data).

If the MEM is used, it is recommended that the user supplies their own marginal parameters as the marginals determined automatically within the MEM tool may be incorrect.

**Q: Can I introduce a time lag to some of the variables (for example, for river gauges that tend to peak more than a day apart)?**

There are no settings in the MEM that automatically apply a time lag to any of the variables, although a fixed time lag could be introduced by offsetting the date of the input data.

# References

- ENVIRONMENT AGENCY, 2012. *Operational Instruction 197\_08: Flood Estimation Guidelines*. Bristol: Environment Agency.
- GOULDY, B., MÉNDEZ, F.J., GUANCHE, Y., RUSEDA, A. AND MINGUEZ, R., 2014. A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis. *Coastal Engineering*, 88, 15-26.
- HAWKES, P.J., 2005. *Use of joint probability methods in flood management. a guide to best practice*. R&D Technical Report FD2308/TR2. London: Department of Environment, Food and Rural Affairs.
- HEFFERNAN, J.E. AND TAWN, J.A., 2004. A conditional approach for multivariate extreme values (with discussion). *Statistical Methodology B*, 66 (3), 497-546.
- HR WALLINGFORD AND LANCASTER UNIVERSITY, 2000. *The joint probability of waves and water levels: JOIN-SEA Version 1.0. User manual*. Report TR 71, originally dated November 1998, re-issued with minor amendments in final form in May 2000. Wallingford: HR Wallingford.
- LAMB, R., KEEF, C., TAWN, J., LAEGER, S., MEADOWCROFT, I., SURENDRAN, S., DUNNING, P. AND BATSTONE, C., 2010. A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, 4 (4), 323-336.

# List of abbreviations

AD	asymptotically dependent
AEP	annual exceedance probability
AI	asymptotically independent
FCRM	flood and coastal risk management
MEM	Multivariate Event Modeller
NRA	National Risk Assessment
NRR	National Risk Register

# Glossary

<b>Annual exceedance probability (AEP)</b>	Probability that a value is exceeded in any given year. Can also be expressed as a percentage (or percent chance), which is 100 multiplied by the AEP. Return period is defined as 1/AEP.
<b>Catastrophe model</b>	Model used to estimate losses due to a catastrophic event such as widespread flooding.
<b>Generalised logistic distribution</b>	Three-parameter statistical distribution used to represent annual maximum data and as a basis for analysis of annual maximum flows in the Flood Estimation Handbook.
<b>Generalised Pareto distribution</b>	Statistical distribution used to represent peaks over threshold data.
<b>Joint probability</b>	The probability of 2 or more events occurring together at the same time.
<b>QMED</b>	Median annual maximum flow.
<b>SWAN model</b>	SWAN (Simulating WAVes Nearshore) is a third generation wave model, developed at Delft University of Technology, which computes random, short-crested wind-generated waves in coastal regions and inland waters ( <a href="http://swanmodel.sourceforge.net">http://swanmodel.sourceforge.net</a> ).
<b>WaveWatch III model</b>	Third generation wave model – described at <a href="http://polar.ncep.noaa.gov/waves/wavewatch/">http://polar.ncep.noaa.gov/waves/wavewatch/</a> .



**Would you like to find out more about us  
or about your environment?**

**Then call us on**

**03708 506 506** (Monday to Friday, 8am to 6pm)

**email**

**enquiries@environment-agency.gov.uk**

**or visit our website**

**www.gov.uk/environment-agency**

**incident hotline 0800 807060** (24 hours)

**floodline 0345 988 1188 / 0845 988 1188** (24 hours)

Find out about call charges ([www.gov.uk/call-charges](http://www.gov.uk/call-charges))



**Environment first:** Are you viewing this on screen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don't forget to reuse and recycle if possible.