Defra / Environment Agency
Flood and Coastal Defence R&D Programme

JOINT PROBABILITY: DEPENDENCE MAPPING AND BEST PRACTICE

R&D Interim Technical Report
FD2308/TR1

July 2003

Authors:
Peter Hawkes
Cecilia Svensson

Statement of use

This document provides information for Defra and Environment Agency Staff about dependence and the use of joint probability methods, and constitutes an Interim R&D output from the Joint Defra / Environment Agency Flood and Coastal Defence R&D Programme. This interim report will be extended, and will be published in December 2004.

- Keywords: flood risk, dependence, joint probability, waves, sea level, surge, river flow, swell, mapping.

- Name, address and contact details of the research contractor: Dr Peter Hawkes, HR Wallingford, Howbery Park, Wallingford, Oxon OX10 8BA, email pjh@hrwallingford.co.uk.

- Client project manager: Dr Suresh Surendran, Environment Agency, Kings Meadow House, Reading RG1 8DQ, email suresh.surendran@environment-agency.gov.uk.

Contract Statement

This work was funded by the Department for the Environment, Food and Rural Affairs, under project code FD2308 (Joint probability: dependence mapping and best practice). Availability of this report implies no endorsement by the Department of the report's results or conclusions. The Project Manager for HR Wallingford was Dr Peter Hawkes; the analysis was undertaken for HR Wallingford by Sun Weidong; both were assisted by colleagues Ben Gouldby and Paul Sayers. The HR Wallingford job number was CBS0706. The Project Manager for CEH Wallingford was Dr Cecilia Svensson, assisted by colleague Dr David Jones. The Project Manager for the Proudman Oceanographic Laboratory was David Blackman, assisted by colleague Dr Ian Vassie. The Client Project Manager was Dr Suresh Surendran of the Environment Agency's Environmental Policy Centre for Risk and Forecasting at Reading. The project was undertaken under the joint Defra / Agency research theme Risk Evaluation and Understanding of Uncertainty, headed by Ian Meadowcroft.

Acknowledgements

The authors are grateful for comments provided by Michael Owen, and for his chairmanship of the open meeting during the industry consultation. Others involved in the industry consultation are listed in Appendix 1. The authors would particularly like to acknowledge Max Beran for his contributions during the consultation stage and Ian Meadowcroft for his comments on the draft report. Acknowledgement here or in the appendix does not imply endorsement of this report's comments and conclusions.

No new source data were generated during the present project. The sea level and surge data were made available by the Proudman Laboratory under licence for use by CEH Wallingford and HR Wallingford during this project. The wave and swell data were made available by the UK Met Office under a licence not specific to this project for use by HR Wallingford. The daily precipitation data were made available by the UK Met Office under a licence not specific to this project for use by CEH Wallingford.

EXECUTIVE SUMMARY

Joint probability analysis predicts the probability of occurrence of events in which two or more partially dependent variables simultaneously take high or extreme values. Several different environmental variables are potentially important in design and assessment of flood and coastal defences, for example waves, tide, surge, river flow, rainfall, swell and wind. This report summarises dependence between key pairs of variables around England, Wales and Scotland in a form suitable for use in simplified joint exceedence analysis methods. Confidence in the dependence estimates is indicated and in a way that could be used in design calculations. The main strands of the work described in this report were to:

- involve and consult the wider industry on their joint probability requirements, intended to increase the chance of appropriate take-up of methods and results, and to identify any gaps in the research programme;
- bring together recent joint probability work at HR Wallingford, CEH Wallingford and the Proudman Oceanographic Laboratory;
- extend it where necessary to the whole of England, Scotland and Wales, analysing and mapping dependence for several variable-pairs relevant to flood and coastal defence, addressing the perceived problem of lack of appropriate data for use in joint probability work.

The variable-pairs analysed and reported are:

- wave height & sea level;
- wave height & surge;
- tide & surge;
- river flow & surge;
- precipitation & surge;
- wind-sea & swell.

An accompanying best practice report will provide clear and relevant notes on when, where and how to apply joint probability methods and results, addressing the issue of reluctance to adopt methods poorly understood outside a fairly small group of specialists.

CONTENTS

EXECUTIVE SUMMARY **iii**
GLOSSARY **ix**

1. Introduction **1**
1.1 Background 1
1.2 Aims of the project 2
1.3 Outline of the project 3
1.4 Related documents 6

2. The Project Data Sets **10**
2.1 Sea level and surge 10
2.2 Waves and swell 12
2.3 River flow 12
2.4 Precipitation 12

3. Methods for Dependence Analysis **19**
3.1 Outline of methods and how used 19
3.2 Method used for waves, swell, sea level and surge 19
3.3 Method used for river flow, precipitation and surge 26
3.4 Method used for tide and surge 28
3.5 Simplified method for joint probability extremes 31
3.6 Relationship between methods 37

4. Results of Dependence Analysis **38**
4.1 Format of results 38
4.2 Results for waves, swell, sea level and surge 38
4.3 Results for river flow, precipitation and surge 52
4.4 Results for tide and surge 65

5. Concluding Remarks on Use of the Dependence Results **69**
5.1 Introduction 69
5.2 Additional information required from the user 69
5.3 Simplified method 69
5.4 Proposed software tool 70
5.5 Monte Carlo simulation method 70

6. References **72**

Tables

Table 2.1 The sea level and surge data sets used in dependence analysis with waves 10

Table 2.2 General information about the 130 river flow stations with daily mean flow records in the period 1963-2001 13

Table 2.3 General information about the 44 rain gauges with daily data in the period 1963-2001 16

Table 3.1 Combinations of two variables for a joint exceedence return period of 1 year 32

Table 3.2 Combinations of two variables for a joint exceedence return period of 5 years 32

Table 3.3 Combinations of two variables for a joint exceedence return period of 20 years 33

Table 3.4 Combinations of two variables for a joint exceedence return period of 100 years 33

Table 3.5 Combinations of two variables for a joint exceedence return period of 500 years 34

Table 3.6 Adjustment factors needed to apply Tables 3.1-3.5 to one record per day data 34

Table 3.7 Worked example of the simplified method 36

Table 3.8 Approximate relationship between dependence measures for a return period of 100 years 37

Table 4.1 Correlation coefficient (ρ, wave height & sea level): all wave directions combined 40

Table 4.2 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is higher 41

Table 4.3 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is lower 42

Table 4.4 Correlation coefficient (ρ, wave height & surge) all wave directions combined 43

Table 4.5 Correlation coefficient (ρ, wind-sea H_s & swell H_s) all wave directions combined 43

Table 4.6 Dependence measure, χ, between daily mean river flow and daily maximum surge: 5% significance level and 90% confidence intervals of χ 54

Table 4.7 Dependence measure, χ, between daily mean river flow and daily maximum surge occurring at high tide: 5% significance level and 90% confidence intervals of χ 57

Figures

Figure 2.1 Tide measurement and wave model points 11

Figure 2.2 Locations of the 130 river flow stations in Great Britain 17

Figure 2.3 Locations of the 44 precipitation gauges in Great Britain 18

Figure 3.1 Joint density contours for the bi-variate normal dependence structure when the marginal distributions are standard normal 25

Figure 4.1 Correlation coefficient (ρ, wave height & sea level): all wave directions combined 44

Figure 4.2 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is higher 45

Figure 4.3 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is lower 46

Figure 4.4 Correlation coefficient (ρ, wave height & surge): all wave directions combined 47

Figure 4.5 Correlation coefficient (ρ, $Hs_{wind\text{-}sea}$ and Hs_{swell}): all wave directions combined 48

Figure 4.6 Weymouth joint exceedence curves 50

Figure 4.7 Portpatrick joint exceedence curves 50

Figure 4.8 Weymouth JOIN-SEA method vs simplified method 51

Figure 4.9 Portpatrick JOIN-SEA method vs simplified method 51

Figure 4.10 Dependence between river flow and daily maximum surge 60

Figure 4.11 Dependence between river flow and daily maximum surge occurring at high tide 61

Figure 4.12 Dependence between precipitation and daily maximum surge in catchments draining to the British east coast 62

Figure 4.13 Dependence between precipitation and daily maximum surge in catchments draining to the British south and west coasts 63

Figure 4.14 The tide-surge interaction parameter for the upper and mid-tide bands 67

Figure 4.15 Relative levels of tide-surge interaction 68

Appendices

Appendix 1 Record of the industry consultation meeting at HR Wallingford on 30 May 2002

Appendix 2 Dependence measure used for river flow, precipitation and surge

Appendix 3 Comparison of dependence measures

GLOSSARY

Amplification factor $q_s(d)$ (in context)
Parameter used in the tide-surge interaction analysis, indicative of surge magnitude

Analytical approach (in context)
Referring to a statistically rigorous joint probability analysis

Bi-Variate Normal (BVN)
A two-variable distribution, where each variable has a Normal distribution, and there is a linear dependence relationship between the two

Chi (χ) (in context)
A dependence measure applied in this report to rainfall & surge and to river flow & surge

Climate change (in context)
Referring to the impact that future climate change might have on the dependence between variable-pairs

Colour coding (in context)
Referring to the five colours used to indicate different ranges of dependence, namely none, modestly correlated, well correlated, strongly correlated and super correlated

Confidence interval
The range, specified in terms of upper and lower bounds, within which the true answer is thought to lie, for a specified level of confidence

Correlation
A linear form of dependence

Correlation coefficient (ρ)
A measure of correlation for which -1 would indicate complete negative dependence, 0 would indicate independence and +1 would indicate complete positive dependence

Correlation factor (CF) (in context)
A measure of dependence used in the simplified method of joint probability analysis, indicating probability of occurrence relative to the independent case

Dependence
Referring to the numerical relationship between variables and the extent to which one can be predicted solely from a knowledge of the other(s)

Dependence measures (CF, χ, ρ)
Different numerical measures of dependence used in this report

Desk study approach (in context)
A particular joint probability analysis approach described in the accompanying best practice report

Direction sector (in context)
Referring to the way in which data sets incorporating waves can be approximately separated into types, specified by wave direction, prior to joint probability analysis

Distance from Wick (d)
A measure of geographical location used in the tide-surge interaction analysis, indicating clockwise distance from Wick in Scotland

Environmental variables
Variables representing weather (precipitation and wind), sea surface (waves, sea level surge and swell) or river conditions (flow)

Events (in context)
Referring to noteworthy occurrences, often identified objectively as records in which a threshold of interest is exceeded amongst the environmental or response variables

Exceedence probability
Probability (between zero and one) that a particular value of a variable will be exceeded

Extreme
An unusually high value of a variable, rarity usually specified in terms of return period or exceedence probability

Higher dependence sector (in context)
Referring to the sector in which dependence is expected to be highest when data sets are separated by wave direction sector (actual direction depends on location)

Independence
The complete lack of dependence between two variables, even if time lag is permitted, as for example between two dice

JOIN-SEA
Method and programs used by HR Wallingford for joint probability analysis

Joint density
The probability that two related variables will simultaneously lie in specified ranges (the equivalent of probability density for a single variable)

Joint exceedence
The probability that two related variables will simultaneously exceed specified values, e.g. wave height greater than x at the same time as sea level greater than y

Joint probability
Referring to the distribution and extremes of two related variables

Joint probability analysis
A commonly used expression for the study of joint probability, usually implying assessment of dependence and prediction of extreme conditions

Joint Probability Method
Method used by POL for joint probability analysis of tide & surge (also Revised JPM and Spatial Revised JPM)

Lagged dependence
Dependence between two variables, involving a time lag, e.g. dependence between precipitation one day and surge the following day

Lower dependence sector (in context)
Referring to the sector in which dependence is expected to be lowest when data sets are separated by wave direction sector (actual direction depends on location)

Mapping (in context)
The collation of dependence results for a particular variable-pair on a map showing the locations of the data sets from which they were derived

Marginal
Referring to the distribution and extremes of a single variable in discussion which might otherwise be thought to refer to two or more variables

Monte Carlo simulation (in context)
Random simulation of hundreds of years of records of related variables whose distributions, dependences and extremes are known

Normal
A symmetrical probability distribution, specified by mean and standard deviation

Normalise (in context)
Numerical procedure where actual data are transformed (in magnitude) to fit a given probability model, whilst maintaining their original ranking (magnitude ordering)

North Atlantic Oscillation Index (NAOI)
Numerical indicator of the (oscillating) pressure difference between the Azores and south-west Iceland, used in long-term climate prediction

Pareto
The Generalised Pareto Distribution is used by JOIN-SEA for prediction of marginal extremes

Peaks Over Threshold (POT)
A method of preparing data for extremes analysis, in which independent maxima above a threshold are identified and extracted

Plotting position
Referring to the exceedence probability assigned to discrete measured values, e.g. whether 0.01, 0.02 or something in between for the highest of fifty recorded values

r-largest
The average number of records per year chosen to set the threshold needed for Peaks Over Threshold data preparation (r = 1 would be called Annual Maxima)

Rank (in context)
The position of a particular record within a data set when the records have been ordered by magnitude (largest would be Rank 1, second largest would be Rank 2 etc)

Record (in context)
Record of one or more variables at a particular time and place (unprocessed records are usually made at a specified time interval) regardless of the values of the variables

Records per year
The average number of records per year, needed to assign exceedence probabilities to high values

Return period
The average period of time between successive exceedences of a given threshold, e.g. wave height of x or flooding at y

Rho (ρ) (in context)
A dependence measure applied in this report to wave height & sea level, to wave height & surge and to wind-sea & swell

Sea level
Still water level of the sea in the absence of wave effects, as would be recorded by a tide gauge

Simplified method (in context)
A method of joint probability analysis given in Section 3.5 of this report, suitable for non-specialist use

Spatial dependence
Referring to the dependence between variables measured some distance apart from each other

Statistical models
Referring to standard probability distributions, defined by mathematical expressions, with parameter values determined from data sets to which they are fitted

Surge
Sea level minus predicted tide, indicating the component of sea level (positive or negative) due to non-astronomical causes

Swell
Referring to longer period wave conditions, usually occurring in open water as shorter period wave conditions decay after a storm

Tail (in context)
The extreme upper end of a probability distribution, from which extremes would be predicted

Threshold (in context)
A particular value (sometimes specified by exceedence probability) of a variable, above which it is of greater interest and/or will be analysed differently

Tide
The astronomical component of sea level, predictable irrespective of weather conditions

Tide-surge interaction parameter (a(X : d))
Dependence measure used in the tide-surge interaction analysis (and unlike χ, ρ and CF, a higher value implies lower flood risk than a lower value)

Trend
A gradual but consistent change in the mean value of a variable over a long period of time, e.g. mean sea level responding to climate change

Uniform (in context)
A rectangular probability distribution, specified only by the ranks of the records within it

Variable-pair (in context)
Referring in general to pairs of variables to be subjected to joint probability analysis, and in particular to the six pairs relevant to flood risk chosen for use in this project

Water level
Referring to water level within an estuary or river due to the combined effect of all environmental variables

1. INTRODUCTION

1.1 Background

Several different environmental variables are potentially important in design and assessment of flood and coastal defences, for examples waves, tides, surges, river flows, rainfall, swell and wind. For several years Defra has been funding work on joint probability, looking at the dependence between the variables and how best to quantify their combined impact on defences. Work focused primarily on its applications to waves and sea levels (at HR Wallingford) and to tides and surges (at POL). Joint probability methods have also been applied to rainfall, surge and river flow on the UK east coast (at CEH Wallingford) and to wind-sea and swell as part of a programme of research on swell and bi-modal sea conditions (at HRW). Demonstration calculations have also been made (at HRW) with wind speeds and sea levels, and with waves and currents.

Methods have been developed, tested and applied in consultancy studies by the researchers involved, and benefits demonstrated, but take-up within the industry has been patchy. There are two main themes to the reasons given by users and potential users for their reluctance to embrace joint probability methods. One relates to the difficulty in understanding and applying the methods, coupled with a lack of 'official' guidance for use of the methods in a prescribed way. The other relates to the lack of published information on the dependence between variables, of key importance for appropriate use of joint probability methods.

Similar methods have been demonstrated to be applicable to three partially dependent variables, namely sea levels, waves and flows, during recent research on extreme total water levels in estuaries (at HRW). This research is complete and a draft report will be issued shortly. The completed study of dependence between surge, river flow and precipitation on the British east coast, carried out by CEH Wallingford as part of the same project, is extended here to encompass the remaining coasts of England, Wales and Scotland.

Specialist joint probability analysis software named JOIN-SEA was developed during the Defra-funded programme of research. It has been in use in consultancy studies at HR Wallingford for about five years, and has been taken up by a small number of UK consultants. POL's published predictions (Dixon and Tawn, 1997) of UK extreme sea levels are widely used in the industry. However, the subtleties of application of JOIN-SEA and Dixon and Tawn (1997) have not always been appreciated outside the originating organisations, and in some instances they have not been applied to full advantage.

A programme of dissemination and evaluation within the UK coastal engineering community began with a successful Defra-sponsored specialist workshop held at Wallingford in December 1998, with presentations by Professor Jonathan Tawn, HR Wallingford and the Proudman Oceanographic Laboratory. Following this workshop, Defra funded a project entitled 'Joint probability: Dissemination, beta-testing and alternative applications' during 1999/2001. This included two further specialist workshops at HR Wallingford: one in February 2000 focussing on briefing on the use of JOIN-SEA, and one in March 2001 to collate feedback from industry users.

The present project continues the process of dissemination and appropriate take-up of joint probability methods in flood and coastal defence design and assessment. This is relevant to Defra and Environment Agency policies with regard to safe and effective design and construction of flood and coastal defences. It will directly support the assessment of flood and erosion risk by helping to refine estimates of extreme environmental loading, the source of risk in many cases (Defra / Environment Agency, 2002). It will bring the best available methods for analysis and application of joint probability into wider use by the river and coastal engineering community. It addresses three priority topics in the ROAME A statement for Theme 5 (Risk Evaluation and Understanding Uncertainty) of the Defra / Agency joint research programme. The topics are: the problem of joint probability (through an investigation of the dependence between flood-producing variables); the sensitivity of the estimate of dependence; and the impact of climate change on the dependence.

No fundamental developments were made during this project. Instead, existing methods, analyses and knowledge were brought together, extended where necessary to include England, Wales and Scotland, and made available, intelligible and relevant to a greater number of users in the UK.

1.2 Aims of the project

The formal objectives of the project are reproduced below.

1. To involve and consult the wider industry including relevant TAG leaders and framework consultants on their joint probability requirements.

2. To bring together recent joint probability work at HR Wallingford, CEH Wallingford and the Proudman Laboratory.

3. To extend it where necessary to the whole of England, Scotland and Wales.

4. To map dependence around and within England, Scotland and Wales for several variable-pairs relevant to flood and coastal defence.

5. To develop best practice guidelines for when and how joint probability methods and results should be used.

6. To draw up proposals for an open workshop and/or training seminars to explain methods and their appropriate use.

7. To assess research needs for development and take-up of joint probability methods and results.

The overall aim of the various objectives and approaches within this project is to increase appropriate take-up of joint probability methods in flood and coastal defence design and assessment, to be achieved through the following main points of information transfer between the project and the industry.

- The dependence maps and tables provide guidance on realistic levels of dependence between different variables, or whether independence would be a reasonable assumption, addressing the problem of lack of appropriate data for use in joint probability work.

- The accompanying best practice guidance will provide clear and relevant notes on when, where and how to apply joint probability methods and results, addressing the issue of reluctance to use methods at present limited to specialists.

- The assessment of industry needs increased the chance of appropriate take-up of methods and results. The assessment of research needs identified gaps needing to be filled by future research, including updates of existing analyses using longer data sets.

- The Defra / Agency sponsored open workshop will not only seek to publicise and disseminate methods, but will also clarify and demonstrate support for the use of such methods.

1.3 Outline of the project

1.3.1 Duration

The project began in December 2001, and runs until November 2004. However, it is intended that all research work will be completed by March 2004, with no more than final dissemination being carried into 2004/05.

The remainder of Section 1.3 is structured around the four main ‘approaches’ used in the contract to describe the scope of work for this project.

1.3.2 Wider industry needs

An outline of the best practice guide was prepared, based on the experience of the Project and Client Teams and comments previously received from about twenty external users, potential users and funders. An invitation to engage in consultation was issued to relevant TAG leaders, framework consultants, other consultants, Agency staff and Defra engineers, around sixty people in all. Probably as many again saw email copies of the consultation material.

About thirty people attended an open meeting at Wallingford on 30 May 2002. There were presentations by HR Wallingford, CEH Wallingford, the Agency and Defra, with a lively discussion. Several small changes were made to the project as a result of the discussion.

Indicative research projects for 2003/4 and beyond were prepared, to support appropriate take-up of joint probability methods and results over the next few years.

1.3.3 Dependence mapping

In this context, dependence indicates the likelihood of two (or more) variables taking high or extreme values at the same time. It is an essential part of joint probability analysis, but possibly the most difficult to quantify, at least for non-specialist users. The intention was to summarise dependence between key pairs of variables around England, Wales and Scotland in a form suitable for use in simplified joint exceedence analysis methods. In some cases, this involved only interpretation and plotting of existing results, but in most cases it involved extension of existing analyses, on a regional basis, to cover England, Wales and Scotland.

Where dependence (at high and extreme values) varies significantly with wind/wave direction, with season, or with time lag, this was investigated. As far as possible, consistent sources of data were used around England, Wales and Scotland, and the same data sets and locations were used for different variable-pairs.

Wave height & sea level; wave height & surge

New data sets of at least ten years duration were acquired for each of the three variables, namely wave height, sea level and surge, for all but one of the 24 tide gauge sites used in the present and previous CEH Wallingford surge/rainfall/flow analyses. For each location and each variable-pair, a site-specific dependence analysis was undertaken by HR Wallingford. Wave direction is often important, both in dependence mapping and in application in coastal engineering, and so for most locations two separate direction sectors were used in addition to ‘overall’. Results were collated and compared, and a small number of analyses were re-visited in an attempt to improve spatial consistency between neighbouring locations.

Tide & surge

The dependence plotting in the present report is based on re-interpretation of work done previously at the Proudman Oceanographic Laboratory (Dixon and Tawn, 1997) and no new analysis of source data was done within the present project.

River flow & surge, precipitation & surge

Water levels in the fluvial-tidal reach of rivers are influenced by both river flow and surge. Any dependence between the two therefore needs to be taken into account when estimating estuarine water levels using these two variables. Existing analyses by CEH Wallingford for the east coast were extended to the remaining coasts of Great Britain, using new data from 16 sea level stations, 72 river flow stations and 24 precipitation gauges, for the period 1963-2001. The river flow station network on the east coast was densified to comprise 58 gauges, and the flow-surge analysis was re-run. Physical explanations of why dependence between river flow and surge occurs in some places and not in others were sought. A precipitation-surge analysis was undertaken to facilitate these interpretations.

Wind-sea & swell

The dependence between wind-sea and swell was reported in the earlier HR Wallingford swell atlas for England and Wales. This analysis was based primarily on wave data from the UK Met Office European Wave Model, with validation against three sets of field wave data. A number of additional analyses were undertaken during the present project to extend the dependence mapping to Scotland. The data sets were longer than those used earlier for England and Wales, but otherwise the procedures were the same as before. Results were checked for spatial consistency and, unlike wave heights and sea levels, dependence between wind-sea and swell does not vary greatly around the country.

To avoid any distortion of the results, each institution reports its detailed results in terms of its preferred statistical dependence measure used for analysis of any particular variable-pair. However, the mapped results provide broadly consistent measures between different variable-pairs, to give a general impression of the variability of dependence between different variable-pairs and around the country. Advice on use of the dependence results in this and the best practice reports will accommodate the use of these different dependence measures.

1.3.4 Best practice guidelines

Apart from preparation of an outline of the best practice report in readiness for the industry consultation, the guidance report is not due to be produced until later in the project.

The guidelines will summarise best practice based on the experience of the Project Team and external consultees. They will pass on clear and relevant advice about how, when and why joint probability analyses and results should be used in project work. They will cover use of simplified methods, the dependence maps, JOIN-SEA, POL extreme sea level predictions and CEH Wallingford surge/flow results, demonstrated by example. They will cover data requirements, types of variable and application amenable to joint probability analysis, methods of checking results, how to incorporate assumptions about climate change, benefits and potential pitfalls.

The guidelines report will include an introductory user's guide, which could be extracted together with example dependence plots, to be published separately in the form of an Environment Agency introductory booklet. The guide will also contain a few case studies, and guidance for use by non-specialists in project work.

The project manager is considering the feasibility of developing a software tool, to be issued with the guidance report, to produce project-specific joint probability tables using the 'desk study approach'. The tool would have the advantages of being able to work with different dependence measures, any dependence value, any number of records per year and any joint exceedence return period, without the large number of alternative tables that would be needed in hard-copy format. It may also allow the extreme joint probability density to be tabulated, in addition to the joint exceedence extremes, for those who prefer to use it, but there are some practical difficulties still to be resolved. Input to the tool would comprise marginal extremes for each of two variables, the number of records per year and the required joint exceedence return period (all supplied by the user) and a dependence value (taken from the guidance report).

1.3.5 Dissemination of results

The main methods and events for dissemination comprise:

- industry consultation in Spring 2002 (see Section 1.4.1), including an open meeting at HR Wallingford on 30 May 2002;

- the present technical report in Spring 2003, containing details of the data sets, dependence analysis methodology and mapping of results; also a separate specialist report on the surge-flow dependence in Spring 2003;

- a best practice report, summarising the technical information on dependence and joint probability, and providing guidance on how to apply it;

- probably a software tool with the best practice report, to generate project-specific joint probability tables;

- an open meeting at HR Wallingford in 2004, at which the reports and methods will be described and offered.

The reports will include a simplified method for use of the dependence maps and tables in joint exceedence analysis, for a number of different return periods. The software tool, if developed, will be based on a slightly more flexible version of this simplified method, to be referred to as the ‘desk study approach’.

1.4 Related documents

1.4.1 The wider industry consultation

An outline of the best practice guide was prepared, based on the experience of the Project and Client Teams and comments previously received from about twenty external users, potential users and funders. The outline was prepared in two parts, separating the Executive Summary (probably to be re-named Introductory User’s Guide) and the main text.

An invitation to engage in consultation was prepared in two parts, comprising a letter and a fax-back form, offering one or more levels of engagement, namely written comments, mailing list membership, telephone interview and a one-day open meeting at Wallingford. The outline and invitation were posted for information and comment to relevant TAG leaders, framework consultants, other consultants, Agency staff and Defra engineers, around sixty people in all. Probably as many again saw email copies of the consultation material. Telephone contact was made with those requesting it during the subsequent two months.

About thirty people attended the open meeting at Wallingford on 30 May 2002. There were presentations by HR Wallingford, CEH Wallingford, the Agency and Defra, with a lively discussion. Several small changes were made to the project as a result of the discussion. Various documents circulated around the time of the open meeting, namely the agenda, attendance and respondent lists, summary table of responses and minutes of the meeting, are reproduced in Appendix 1.

Indicative research projects for 2003/04 and beyond were needed to support appropriate take-up of joint probability methods and results over the next few years. Ideas considered by the Project Team included new developments, new applications, refinements of existing methods, updating of previous published predictions of extreme sea levels and swell, new incentives to take-up (e.g. data access, Defra / Agency support), and methods to assess the quality and quantity of take-up. Titles of the projects for which proposals were prepared in Agency Short Form A format are:

- Collective risk of river flooding: A pilot study.
- Updated estimates of extreme sea levels at 'A' Class national tide gauge sites: Spatial analyses for the UK coast.
- Estimates of extreme sea levels in complex coastal regions.
- Incorporation of temporal dependence (sequencing) into JOIN-SEA long-term simulation.
- Update the 1995 swell atlas for England and Wales, extend to Scotland and develop a software tool for the main results.

The two proposals involving extreme sea levels were later assimilated into a larger proposal called:

- Environmental extremes: A managed programme.

1.4.2 The best practice guide and the report on surge-flow dependence

The guidance report will summarise best practice based on the experience of the Project Team and external consultees, including clear and relevant advice about how, when and why joint probability analyses and results should be used in project work. The report (Defra / Environment Agency, 2003a) will cover use of simplified methods, the dependence maps, JOIN-SEA, POL extreme sea level predictions and CEH Wallingford surge/flow results, demonstrated by example. It will cover data requirements, types of variable and application amenable to joint probability analysis, methods of checking results, how to incorporate assumptions about climate change, benefits and potential pitfalls. It may also include a software tool to simplify and extend the range of usage of the desk study approach.

A separate report (Defra / Environment Agency, 2003b) provides more detailed results from the investigation of dependence between surge and river flow for use by hydrologists than seemed appropriate to include in the present overall technical report on the dependence mapping. Defra / Environment Agency (2003b) includes full results from the time lagged and seasonal analyses. It also includes more interpretation of the possible meteorological and geographical reasons for dependence and its variation with location, season and time lag between surge and flow, and more discussion of the climate change issues. The appropriate level of information has been extracted and included in the present report, which is aimed at a more general readership.

The present report, the best practice report and the specialist report on surge-flow dependence comprise the written output of Project FD2308: Joint probability: dependence mapping and best practice.

1.4.3 Joint probability reports

Joint probability methods for use in flood and coastal defence work were developed from about 1980 onwards, and have been in fairly routine use since about 1990. Defra funded a series of joint probability projects at HR Wallingford, the Proudman Oceanographic Laboratory and CEH Wallingford. A few reports and papers from this programme of research and development are listed below.

HR Wallingford's JOINPROB analysis method for waves and sea levels was in use on consultancy studies for several years. It is described, and validated against field records of damage to coastal structures in HR Wallingford (1994).

HR Wallingford (1997) provides offshore swell conditions, with different frequencies of occurrence, for England and Wales, information on the joint probability of wind-sea and swell, and how to construct bi-modal wave spectra for input to coastal engineering design.

HR Wallingford's JOIN-SEA analysis method for waves and sea levels was developed and validated in the mid-1990s, and has been disseminated and used in consultancy studies since about 1997. HR Wallingford (2000a) details the theoretical developments and validation, HR Wallingford (2000b) is the accompanying user manual for the computer programs, Owen *et al* (1997) provides an introductory description of techniques, and Hawkes *et al* (2002) a slightly more technical description.

Defra / Environment Agency (2003c) describes an extension of the JOIN-SEA method to extreme water levels in rivers and estuaries, taking account of river flow, sea level and waves.

Jones (1998) describes the application of joint probability methods to the combined action of river flow and sea level, comparing the results of statistical analysis with continuous simulation modelling.

Reed and Dwyer (1996) discuss the objectives for research into the estimation of flood frequency at river confluences, and report some lessons learned from two case studies.

Svensson and Jones (2000) detail development of methods and analysis of the dependence between river flow, surge and rainfall using long-term measurements on the east coast of Britain. Svensson and Jones (2002) give a more concise description of the work for the Journal of Climatology.

Pugh and Vassie (1980), Tawn and Vassie (1989) and Tawn (1992) describe development of the joint probabilities method for tides and surges in prediction of extreme sea levels. Coles and Tawn (1990) describe further development of the approach and application to tide gauge sites around Britain.

During the 1990s, a major programme of sea level analysis at the Proudman Laboratory using joint probability methods produced Dixon and Tawn (1995) on extreme sea levels for UK A class tide gauge sites, and Dixon and Tawn (1997) extending the predictions to the whole of the UK coast.

1.4.4 Marginal extremes reports

Although marginal (single variable) extremes were predicted during derivation of the dependence results, they are not reported as part of the present project. The present analyses focused on dependence, and provide the most consistent published information on dependence between variable-pairs for use around England, Wales and Scotland. The marginal extremes, however, were derived only as a by-product of the dependence analyses, and in most cases more accurate and consistent predictions can be obtained from other published reports, some of which are outlined below.

A series of offshore design guidelines issued by the Health and Safety Executive, and before that the Department of Energy, provide contours of extreme wave height, extreme surge, extreme wind speed, tidal range etc for the British Isles. In the absence of site-specific modelling or investigation, offshore design guidelines, for example DoE (1977 and 1984) probably provide the best published source for extreme surges and extreme waves, covering the whole of the British Isles. Coles and Tawn (1990), Dixon and Tawn (1995) and Dixon and Tawn (1997) provide predictions of extreme sea levels for England, Wales and Scotland. HR Wallingford (1997) provides information on swell wave conditions for England and Wales.

Volume 2 of the Flood Estimation Handbook (Faulkner, 1999) shows UK maps of 1-hour and 1-day design rainfalls for different return periods, as well as maps of the median annual maximum rainfall (RMED) and rainfall growth rates used to derive them. Volume 3 of the Flood Estimation Handbook (Robson and Reed, 1999) provides a UK map of the index flood (the annual maximum flood, QMED) and describes different methods for deriving the growth curve. The design flood can then be estimated using the QMED and the growth curve.

Please note that none of these sources should be used indiscriminately, without an understanding of their potential limitations and intended usage.

2. THE PROJECT DATA SETS

2.1 Sea level and surge

Hourly measured sea level and surge data from twenty-four measurement stations were made available by the Proudman Oceanographic Laboratory for use during this project. The locations shown in Figure 2.1 were chosen to provide broad coverage of the whole coast of England, Wales and Scotland. The tide gauge stations and years of data are listed in Table 2.1. The records consist of levels in millimetres (relative to Chart Datum for sea level) at hourly intervals specified to GMT, missing data being flagged by error codes. Sea level is the recorded still (i.e. in the absence of waves) water level, and surge is the difference between sea level and predicted tide for that time and location.

Table 2.1 The sea level and surge data sets used in dependence analysis with waves

Tide gauge station	Sea level data	Surge data
Aberdeen	1964-1999	1964-1999
Avonmouth	1972-1998	1972-1998
Dover	1961-1999	1961-1999
Fishguard	1963-2001	1963-2001
Heysham	1964-2001	1964-2001
Holyhead	1964-2001	1964-2001
Ilfracombe	1968-1999	1968-1999
Immingham	1961-1999	1961-1999
Lerwick	1961-1999	1961-1999
Liverpool Gladstone Dock	1993-2001	1993-2001
Liverpool Princes Pier	1963-1986	1963-1986
Lowestoft	1964-1999	1964-1999
Milford Haven	1961-2001	1961-2001
Millport	1978-2001	1978-2001
Newhaven	1983-2001	1983-2001
Newlyn	1961-2001	1961-2001
North Shields	1968-1999	1968-1999
Portpatrick	1968-2001	1968-2001
Portsmouth	1991-2001	1991-2001
Sheerness	1965-1999	1965-1999
Tobermory	1990-2001	1990-2001
Ullapool	1966-2001	1966-2001
Weymouth	1991-2001	1991-2001
Wick	1965-1999	1965-1999

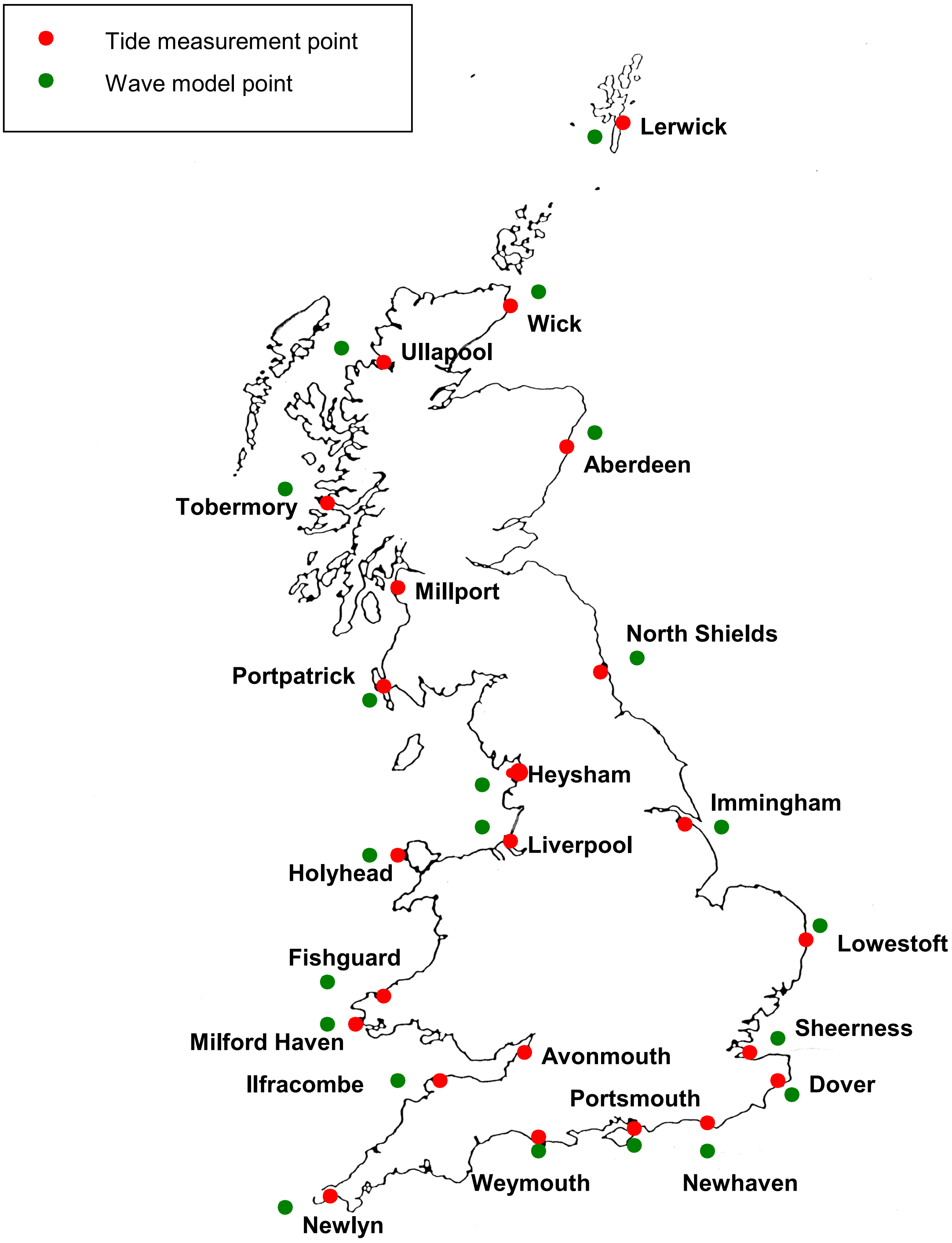


Figure 2.1 Tide measurement and wave model points

2.2 Waves and swell

In the absence of widespread long-term wave measurements, three-hourly wave data were extracted from the archive of the UK Met Office European wave model for use during this project. This model is run on a 25km grid primarily for forecasting purposes and, for particular individual locations, may not represent the best source of wave data, but it does offer consistent coverage over the whole study area. The particular locations used were chosen to correspond to the tide gauge stations listed in Table 2.1 (with the exception of Princes Pier for which there were no simultaneous data on waves and sea levels). The exact locations of the wave model grid points used are shown in Figure 2.1. In each case, data for the period April 1990 to March 2002 were used, with very few records missing. The records consist of significant wave height (in metres with centimetre resolution), mean wave period (in seconds, with tenths resolution) and mean wave direction (degrees North with one-degree resolution) for each of wind-sea, swell-sea and total-sea, at three-hourly intervals specified to GMT. (Records also include wind velocity but these were not used except in refinement of the swell records as described later.)

2.3 River flow

Daily mean river flows, generally for 9.00-9.00 GMT, from 130 stations in Great Britain, were extracted 1963-2001 (Figure 2.2, Table 2.2) from the National River Flow Archive at CEH Wallingford. The stations were chosen to be as far downstream as possible without being tidally influenced and to have as few missing data as possible. Stations further up in the catchment were used in some cases when the downstream records were short, or tidally influenced (only 54032, Severn at Saxon's Load), sometimes in addition to the downstream station. A few of the catchments are therefore nested. The network of 40 gauges draining to the east coast used in the previous study (Svensson and Jones, 2000) was densified to comprise 58 gauges in the present study, and longer, updated, records are now used.

2.4 Precipitation

Daily precipitation accumulations from 9.00-9.00 GMT were obtained from the UK Met Office. Precipitation data were extracted for the period 1963-2001 for 27 stations in catchments draining to the south and west coasts of Great Britain. Three of these were used also for the earlier east coast study (Svensson and Jones, 2000), which in total used 20 stations with data in the shorter period 1965-1997 (Figure 2.3, Table 2.3). The east coast surge-precipitation analysis was not repeated during the present study.

Table 2.2 General information about the 130 river flow stations with daily mean flow records in the period 1963-2001

Station	River	Location	Easting (km, GB national grid)	Northing	Altitude (m)	Catchment area (km^2)	Mean river flow (m^3/s)	Missing data (%)
2001	Helmsdale	Kilphedir	299.7	918.1	17.0	551	13.0	33.3
4001	Conon	Moy Bridge	248.2	854.7	10.0	962	52.0	16.0
7002	Findhorn	Forres	301.8	858.3	6.8	782	19.5	2.6
7004	Nairn	Firhall	288.2	855.1	7.2	313	5.6	43.6
8006	Spey	Boat o Brig	331.8	851.8	43.1	2861	64.6	2.6
9002	Deveron	Muiresk	370.5	849.8	25.3	955	16.4	5.1
10003	Ythan	Ellon	394.7	830.3	3.8	523	7.9	54.8
11001	Don	Parkhill	388.7	814.1	32.4	1273	20.3	20.3
12001	Dee	Woodend	363.5	795.6	70.5	1370	37.2	2.6
12002	Dee	Park	379.8	798.3	22.6	1844	46.9	27.8
13007	North Esk	Logie Mill	369.9	764.0	10.6	732	19.2	35.9
14001	Eden	Kemback	341.5	715.8	6.2	307	3.9	14.7
14002	Dighty Water	Balmossie Mill	347.7	732.4	16.1	127	1.5	19.9
15006	Tay	Ballathie	314.7	736.7	26.2	4587	169.8	0.0
15013	Almond	Almondbank	306.8	725.8	20.4	175	5.0	2.6
16004	Earn	Forteviot Bridge	304.4	718.3	8.0	782	28.8	27.6
17002	Leven	Leven	336.9	700.6	8.7	424	6.5	19.5
18002	Devon	Glenochil	285.8	696.0	5.5	181	4.5	2.6
18003	Teith	Bridge of Teith	272.5	701.1	14.8	518	23.9	0.2
18011	Forth	Craigforth	277.5	695.5	3.7	1036	48.7	49.8
19001	Almond	Craigiehall	316.5	675.2	22.8	369	6.1	2.6
19006	Water of Leith	Murrayfield	322.8	673.2	37.5	107	1.5	2.6
19007	Esk	Musselburgh	333.9	672.3	3.3	330	4.2	2.6
20001	Tyne	East Linton	359.1	676.8	16.5	307	2.8	2.9
21009	Tweed	Norham	389.8	647.7	4.3	4390	78.9	2.6
22001	Coquet	Morwick	423.4	604.4	5.2	570	8.6	2.2
22006	Blyth	Hartford Bridge	424.3	580.0	24.6	269	2.1	10.5
23001	Tyne	Bywell	403.8	561.7	14.0	2176	45.7	0.7
24009	Wear	Chester le Street	428.3	551.2	5.5	1008	14.7	37.8
25001	Tees	Broken Scar	425.9	513.7	37.2	818	16.8	0.0
26002	Hull	Hempholme Lock	508.0	449.8	2.8	378	3.4	17.4
27002	Wharfe	Flint Mill Weir	442.2	447.3	13.7	759	17.4	0.0
27003	Aire	Beal Weir	453.5	425.5	5.5	1932	36.2	2.3
27021	Don	Doncaster	457.0	404.0	4.4	1256	16.3	4.4
28009	Trent	Colwick	462.0	339.9	16.0	7486	85.3	0.0
28022	Trent	North Muskham	480.1	360.1	5.0	8231	90.7	14.7
29001	Waithe Beck	Brigsley	525.3	401.6	15.7	108	0.3	0.3
29002	Great Eau	Claythorpe Mill	541.6	379.3	6.6	77	0.7	1.0
31002	Glen	Kates Br/King St Br	510.6	314.9	6.1	342	1.2	0.6
32001	Nene	Orton	516.6	297.2	3.4	1634	10.1	25.7
33006	Wissey	Northwold	577.1	296.5	5.3	275	1.8	9.9
33007	Nar	Marham	572.3	311.9	4.6	153	1.1	0.0
33024	Cam	Dernford	546.6	250.6	14.7	198	1.0	0.5
33039	Bedford Ouse	Roxton	516.0	253.5	15.7	1660	11.6	25.2
34003	Bure	Ingworth	619.2	329.6	12.2	165	1.1	0.1
34006	Waveney	Needham Mill	622.9	281.1	16.5	370	1.8	2.5
34013	Waveney	Ellingham Mill	636.4	291.7	1.6	670	0.6	52.1
34019	Bure	Horstead Mill	626.7	319.4	1.3	313	2.1	32.2
35004	Ore	Beversham Bridge	635.9	258.3	2.4	55	0.3	7.2
35013	Blyth	Holton	640.6	276.9	12.3	93	0.4	19.4
36006	Stour	Langham	602.0	234.4	6.4	578	3.0	0.0
37001	Roding	Redbridge	541.5	188.4	5.7	303	2.0	0.0
37005	Colne	Lexden	596.2	226.1	8.2	238	1.0	0.1

Table 2.2 General information about the 130 river flow stations with daily mean flow records in the period 1963-2001 (continued)

37009	Brain	Guithavon Valley	581.8	214.7	16.2	61	0.4	0.1
37010	Blackwater	Appleford Bridge	584.5	215.8	14.6	247	1.3	0.1
39001	Thames	Kingston	517.7	169.8	4.7	9948	63.1	0.0
40011	Great Stour	Horton	611.6	155.4	12.5	345	3.2	4.7
40012	Darent	Hawley	555.1	171.8	11.2	191	0.6	2.4
40021	Hexden Channel	Hopemill Br Sandhurst	581.3	129.0	5.2	32	0.3	48.2
41004	Ouse	Barcombe Mills	543.3	114.8	5.2	396	3.6	19.7
41017	Combe Haven	Crowhurst	576.5	110.2	1.9	31	0.3	18.0
41023	Lavant	Graylingwell	487.1	106.4	20.7	87	0.3	21.1
42003	Lymington	Brockenhurst	431.8	101.9	6.1	99	1.0	0.8
42004	Test	Broadlands	435.4	118.9	10.1	1040	10.8	0.2
42006	Meon	Mislingford	458.9	114.1	29.3	73	1.0	0.0
43021	Avon	Knapp Mill	415.6	94.3	0.9	1706	19.9	33.1
44001	Frome	East Stoke Total	386.6	86.7	~9	414	6.4	12.0
45001	Exe	Thorverton	293.6	101.6	25.9	601	16.0	0.0
45005	Otter	Dotton	308.7	88.5	14.5	203	3.2	0.0
46002	Teign	Preston	285.6	74.6	3.8	381	9.2	2.6
46003	Dart	Austins Bridge	275.1	65.9	22.4	248	11.0	2.6
47001	Tamar	Gunnislake	242.6	72.5	8.2	917	22.4	0.0
47004	Lynher	Pillaton Mill	236.9	62.6	8.5	136	4.6	2.0
47007	Yealm	Puslinch	257.4	51.1	5.5	55	1.7	2.8
48007	Kennal	Ponsanooth	176.2	37.7	13.6	27	0.5	14.7
48011	Fowey	Restormel	209.8	62.4	9.2	169	4.9	0.0
49001	Camel	Denby	201.7	68.2	4.6	209	6.1	4.3
49002	Hayle	St Erth	154.9	34.1	7.0	48	1.0	13.2
50001	Taw	Umberleigh	260.8	123.7	14.1	826	18.5	0.0
50002	Torridge	Torrington	250.0	118.5	13.9	663	16.0	0.0
51003	Washford	Beggearn Huish	304.0	139.5	67.1	36	0.8	18.2
52009	Sheppey	Fenny Castle	349.8	143.9	5.8	60	1.1	3.8
53018	Avon	Bathford	378.5	167.0	18.0	1552	18.0	17.7
54001	Severn	Bewdley	378.2	276.2	17.0	4325	60.6	0.0
54032	Severn	Saxons Lode	386.3	239.0	7.5	6850	87.2	19.9
55023	Wye	Redbrook	352.8	211.0	9.2	4010	76.0	0.0
56001	Usk	Chain Bridge	334.5	205.6	22.6	912	27.9	0.0
56002	Ebbw	Rhiwderyn	325.9	188.9	30.6	217	7.6	7.4
57005	Taff	Pontypridd	307.9	189.7	45.1	455	19.7	20.5
58001	Ogmore	Bridgend	290.4	179.4	13.8	158	6.7	2.2
59001	Tawe	Ynystanglws	268.5	199.8	9.3	228	12.1	1.1
60003	Taf	Clog-y-Fran	223.8	216.0	7.0	217	7.5	7.3
60010	Tywi	Nantgaredig	248.5	220.6	7.8	1090	39.2	0.1
61002	Eastern Cleddau	Canaston Bridge	207.2	215.3	5.0	183	6.0	0.2
62001	Teifi	Glan Teifi	224.4	241.6	5.2	894	28.7	0.0
63001	Ystwyth	Pont Llolwyn	259.1	277.4	12.0	170	6.0	2.1
64006	Leri	Dolybont	263.5	288.2	14.6	47	1.3	0.0
65001	Glaslyn	Beddgelert	259.2	347.8	32.9	69	5.8	0.4
66001	Clwyd	Pont-y-Cambwll	306.9	370.9	15.3	404	6.3	0.0
67015	Dee	Manley Hall	334.8	341.5	25.4	1019	30.9	0.0
68020	Gowy	Bridge Trafford	344.8	371.1	4.1	156	1.1	42.7
69002	Irwell	Adelphi Weir	382.4	398.7	24.1	559	17.5	3.3
69007	Mersey	Ashton Weir	377.2	393.6	14.9	660	12.4	34.4
70004	Yarrow	Croston Mill	349.8	418.0	6.9	74	1.9	35.6
71001	Ribble	Samlesbury	358.7	431.4	6.0	1145	32.9	1.3
72004	Lune	Caton	352.9	465.3	10.7	983	35.3	5.1
72008	Wyre	Garstang	348.8	444.7	10.9	114	3.3	14.9
73002	Crake	Low Nibthwaite	329.4	488.2	38.6	73	4.0	2.0

Table 2.2 General information about the 130 river flow stations with daily mean flow records in the period 1963-2001 (continued)

73005	Kent	Sedgwick	350.9	487.4	18.9	209	8.9	15.0
74001	Duddon	Duddon Hall	319.6	489.6	14.8	86	4.8	17.9
74006	Calder	Calder Hall	303.5	504.5	26.4	45	1.8	9.3
75002	Derwent	Camerton	303.8	530.5	16.7	663	25.8	0.0
76007	Eden	Sheepmount	339.0	557.1	7.0	2287	51.9	12.2
77001	Esk	Netherby	339.0	571.8	14.3	842	26.1	7.5
78003	Annan	Brydekirk	319.1	570.4	10.0	925	29.5	12.2
79002	Nith	Friars Carse	292.3	585.1	19.8	799	27.5	0.0
79005	Cluden Water	Fiddlers Ford	292.8	579.5	22.9	238	7.9	1.9
81002	Cree	Newton Stewart	241.2	565.3	4.8	368	15.7	1.9
82001	Girvan	Robstone	221.7	599.7	9.1	246	6.6	2.0
83005	Irvine	Shewalton	234.5	636.9	4.8	381	9.6	23.3
84001	Kelvin	Killermont	255.8	670.5	27.0	335	8.6	0.1
84013	Clyde	Daldowie	267.2	661.6	7.5	1903	48.8	2.0
85001	Leven	Linnbrane	239.4	680.3	5.3	784	43.5	1.6
86001	Little Eachaig	Dalinlongart	214.3	682.1	10.1	31	1.8	15.0
93001	Carron	New Kelso	194.2	842.9	5.6	138	10.9	41.0
94001	Ewe	Poolewe	185.9	880.3	4.6	441	29.7	20.0
95001	Inver	Little Assynt	214.7	925.0	60.3	138	8.5	37.4
95002	Broom	Inverbroom	218.4	884.2	4.6	141	7.3	56.5
96001	Halladale	Halladale	289.1	956.1	23.2	205	5.1	33.3
97002	Thurso	Halkirk	313.1	959.5	30.2	413	8.8	23.1

Table 2.3 General information about the 44 rain gauges with daily data in the period 1963-2001

Gauge	Location	Easting (km, GB national grid)	Northing (km, GB national grid)	Altitude (m)	Mean annual precipitation (mm)	Missing data (%)
24724	Durham	426.7	541.5	102	650	15.4
43941	Dalton Holme	496.5	445.2	34	684	15.4
62254	Lower Barden Resr	403.5	456.3	227	1203	15.4
82583	Sheffield	433.9	387.3	131	830	15.4
152542	Seaton Mill	490.8	297.6	41	618	15.4
186331	Broom's Barn	575.3	265.6	75	592	15.4
222885	Belstead Hall	612.7	241.2	38	592	15.4
239172	Theydon Bois, Thrifts Hall Farm	545.7	198.7	75	596	15.4
264282	Wallingford	461.8	189.8	48	591	15.4
302179	Wye	605.8	146.9	56	744	0.0
320345	Bognor Regis	493.3	98.8	7	733	0.0
328989	Leckford	439.3	136.2	117	811	0.0
361850	Chudleigh	286.6	79.2	70	1012	0.0
381210	Penzance	146.8	30.2	19	1156	4.1
386255	Bude	220.8	106.3	15	906	0.0
404124	Ashcott, Bradley Cottage	343.9	136.5	35	733	5.8
412297	Lacock	392.1	170.2	49	716	0.0
444643	Kyre	363.8	262.0	99	736	8.3
489170	Neuadd Resr No.11A	303.3	218.4	463	2198	3.2
511627	Dale Fort	182.3	205.1	33	869	0.0
519357	Cwmystwyth	277.3	274.9	301	1805	1.3
541918	Llanuwchllyn	287.8	329.9	173	1676	0.8
547250	Loggerheads	320.0	362.2	215	931	0.0
565260	Knutsford	375.6	378.3	65	836	2.6
576634	Preston, Moor Park	353.7	431.1	33	997	2.8
588005	Coniston, Holywath	329.9	497.8	76	2473	0.0
610122	Eskdalemuir Observatory	323.5	602.6	242	1581	15.4
627478	Pullaugh Burn	254.4	574.1	183	2260	1.5
652672	Carnwath	297.4	646.4	208	849	0.0
666484	Younger Botanic Garden	214.1	685.7	12	2338	0.0
691637	Onich	202.8	763.3	15	2115	16.2
708615	Plockton	180.2	833.2	12	1430	0.5
717685	ULVA: Ulva House	144.2	739.1	15	1678	1.3
741962	Knockanrock	218.7	908.7	244	2047	27.4
757883	Hoy P.Sta.	313.7	960.7	23	961	0.0
763886	SHETLAND: Lerwick Observatory No.2	445.3	1139.7	82	1219	12.8
792393	Fairburn House	245.5	852.8	152	1009	0.9
812566	Elgin, Kirkhill	324.9	862.8	11	705	15.4
841537	Craibstone No.1	387.1	810.7	102	815	15.4
844215	Balmoral	326.0	794.6	283	833	15.4
870622	Faskally	291.8	759.9	94	879	15.4
894666	Glenquey Resr	298.2	702.9	277	1522	15.4
900662	Bush House	324.5	663.3	184	874	15.4
913320	Floors Castle	370.7	634.5	59	645	15.4

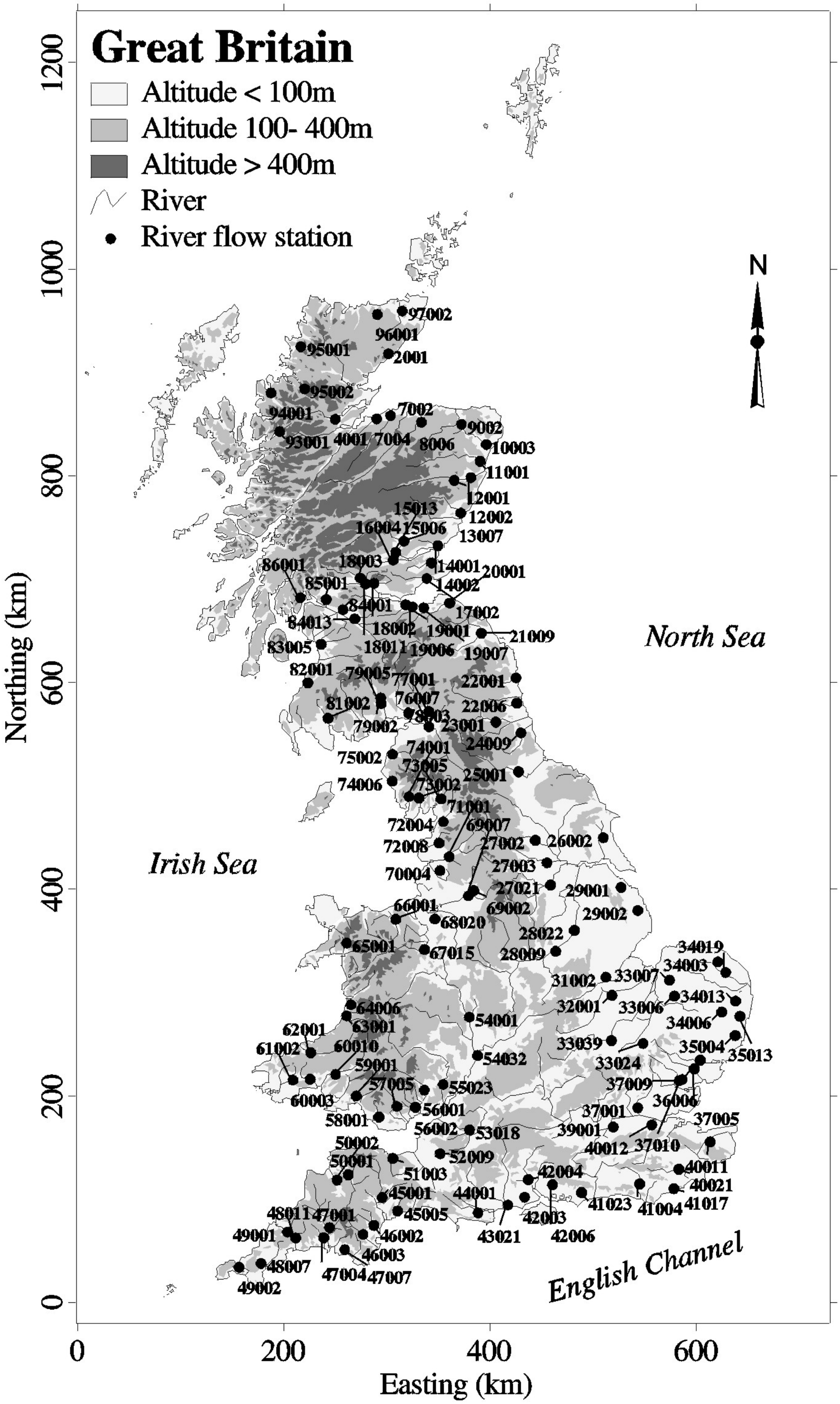


Figure 2.2 Locations of the 130 river flow stations in Great Britain

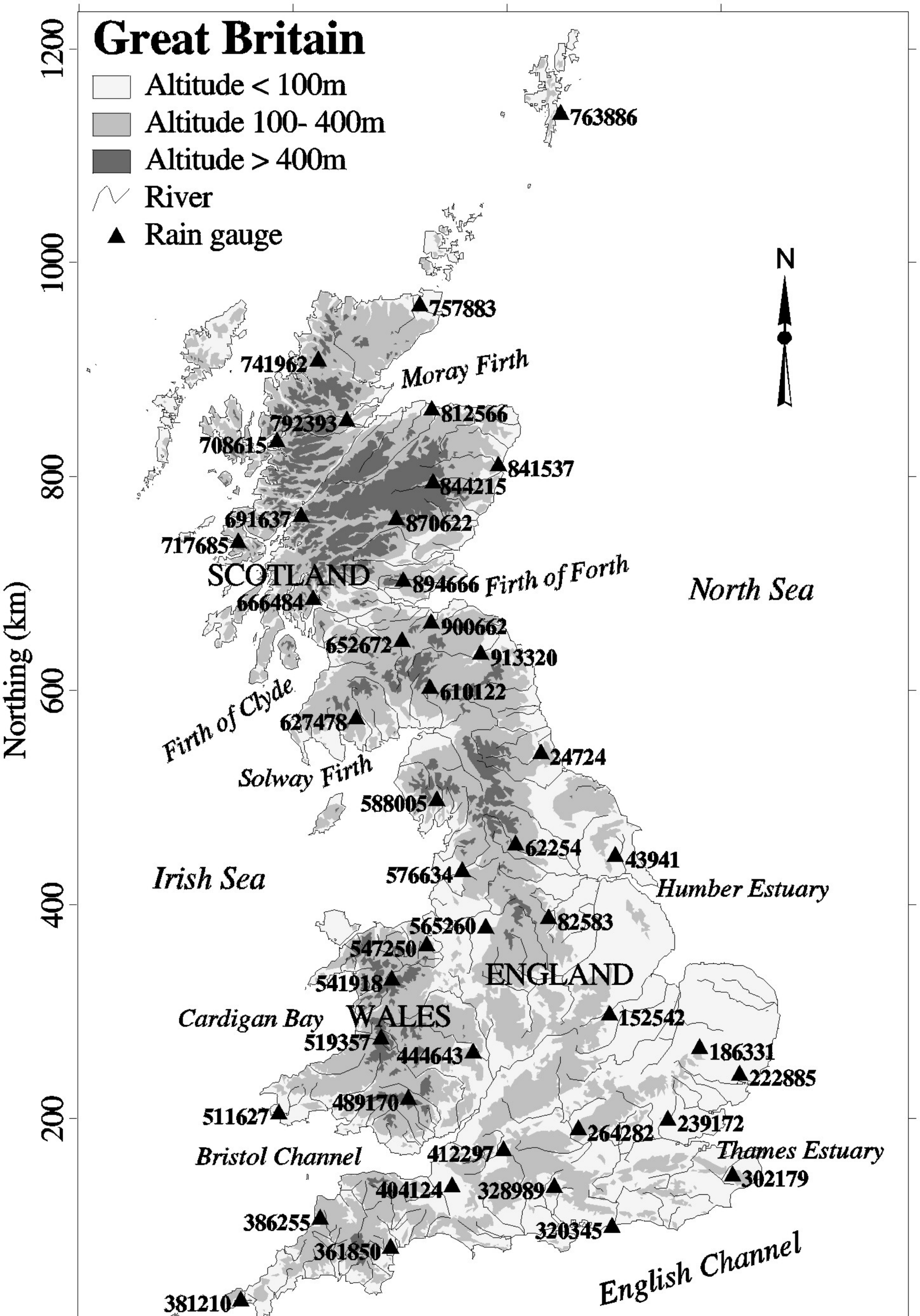


Figure 2.3 Locations of the 44 precipitation gauges in Great Britain

3. METHODS FOR DEPENDENCE ANALYSIS

3.1 Outline of methods and how used

Chapter 3 provides the main technical detail for the report, describing the analysis methods used, the dependence measures used, definitions, assumptions and data preparation. Each of the three institutions involved in this project worked in terms of its own preferred data preparation method, analysis method and dependence measure, appropriate to the variable-pairs involved. Section 3.2 describes the JOIN-SEA joint probability analysis method applied to wave height & sea level, wave height & surge, and wind-sea & swell, by HR Wallingford. Section 3.3 describes the method applied to river flow & surge, and precipitation & surge, by CEH Wallingford. Section 3.4 describes the method applied to tide & surge by the Proudman Oceanographic Laboratory.

Section 3.5 describes a simplified method for joint probability analysis, extending a method originally published in CIRIA (1996), based on a knowledge of dependence (from this report) and marginal extremes (from elsewhere). This method will be developed further into the ‘desk study approach’, forming part of the accompanying best practice report (Defra / Environment Agency, 2003a). Although the CIRIA (1996) method remains valid, it is recommended that new users adopt the desk study approach.

Section 3.6 contrasts the analytical approaches used by HR Wallingford and CEH Wallingford. The methods and dependence measures are not interchangeable, as fundamentally different statistical models are involved, but an approximate relationship is given between the ρ measure used by HRW and the χ measure used by CEH. Section 3.6 also explains how the ‘correlation factor’ needed for the CIRIA (1996) method can be estimated from either ρ or χ.

3.2 Method used for waves, swell, sea level and surge

3.2.1 The JOIN-SEA analysis method

Details of the theory, development, testing and validation of JOIN-SEA are given in HR Wallingford (2000a). The five main stages involved in running JOIN-SEA for each data set are briefly described below. In the context of mapping the dependence parameter ρ, only Steps 1 and 3 were needed, but Step 2 is an integral part of the analysis procedure, and Steps 4 and 5 were often run for checking purposes.

Step 1: Preparation of input data

Each record used as input to the joint probability analysis consists of a wave height, a wave period and a sea level (or alternatively a surge or a swell in the present study) using nearby measurement or prediction locations for both waves and sea levels. For analysis involving high sea levels, a convenient way of satisfying the requirement for the records to be both temporally independent, and relevant, is to use only those records representing conditions at the peak of each tidal cycle (i.e. one record every 12 or 13 hours).

Step 2: Fitting of marginal distributions
This stage involves the fitting of statistical models to wave heights, sea levels (or alternatively surges or swells in the present study) and wave steepnesses. Generalised Pareto Distributions are fitted to the top few percent of the marginal variables, i.e. wave heights and sea levels; the empirical distribution of wave steepnesses is modelled by a Normal regression on wave height.

Joint probability analysis is based on simultaneous information on the variables of interest. It is quite likely that there will be additional non-simultaneous data on at least one of the variables, with which to refine the extremes predictions for that one variable. JOIN-SEA incorporates any refinements by scaling during the long-term simulation of data, thus permanently building this information into the synthesised sea state data to be used in subsequent structural analysis.

Step 3: Fitting of statistical models for dependence
This involves conversion to Normal scales, and fitting of a dependence function to the bulk of the wave height and sea level (or alternatively surge or swell in the present study) data. Two alternative partial dependence statistical models are available to represent the dependence between wave heights and sea levels. These consist of a single Bi-Variate Normal (BVN) Distribution and a mixture of two BVNs. These models were chosen, since the dependence and extremes characteristics of the BVN are well understood.

The choice between one and two BVNs is usually determined by the relative goodness of fit to the data, but to maintain consistency of approach the single BVN was used throughout the dependence mapping project. In this and in the previous stages the user retains some control over the process, by being able to select both the thresholds above which the fitting will be applied, and the starting values for optimisation of the fits: this is assisted by reference to diagnostics to assess the fits.

Step 4 Long-term simulation
This stage involves simulation of a large sample of synthetic records of H_s, T_m and sea level (or alternatively surge or swell in the present study), based on the fitted distributions, and with the same statistical characteristics as the input data. This permits 1000's of years of sea conditions to be simulated with fitted distributions, extremes and dependences for wave height, sea level and wave period. This in turn provides for greater flexibility in the subsequent analysis of the synthesised sea state data.

Step 5: Analysis of joint exceedence extremes and structure functions
This stage involves analysis of the large simulated sample of data to produce extreme values for use in design and assessment of sea defences. These can take the form of extreme wave heights (and associated periods), extreme sea levels (or alternatively surges or swells in the present study), or extreme combinations of the two. In addition, any structure function (e.g. overtopping, run-up, force) which can be defined in terms of constants (e.g. wall slope, toe depth, crest elevation etc) and variables H_s, T_m and sea level, can be synthesised directly for every record in the simulated data sample. Direct analysis of the distribution and extremes of the structure variable is then relatively easy: extreme values can be estimated from the appropriate exceedence probability in the synthesised data.

3.2.2 Preparation of the data sets for analysis

For each data set analysed, simultaneous data on each of two variables (plus wave period) were required. Wave data from the UK Met Office European model were used in all of the HR Wallingford analyses, but a previous study (HR, 1997) had shown that data before April 1990 were unreliable for some UK coastal locations. For consistency, the period April 1990 to March 2002 was used for all wind-sea & swell analyses. The period of sea level or surge data available meant that a slightly shorter period of time (varying between tide gauge stations) was used for wave height & sea level and for wave height & surge.

<u>Wave height & sea level</u>

Sea level and wave data were matched up hour-by-hour over the approximately ten year data sets, interpolating wave height (and period) between the three-hourly source records. From these hourly joint data sets, one record was extracted at the peak of each tide, i.e. one every 12-13 hours, for use in subsequent joint probability analysis. Each data set then consisted of one record of wave height, wave period and sea level per tide over a period of about ten years, so around 7000 records in all, and 707 per year.

<u>Wave height & sea level, by direction sector</u>

In practice wave height and sea level is the most commonly used variable-pair, and it is common to divide the procedures and calculations into two or three wave direction sectors. These might loosely represent, say, distantly, regionally and locally generated waves, depending on the open water distances in each direction. In this study most wave and sea level data sets were sub-divided into two sectors, different in each case, to represent the direction bands where higher and lower dependence was expected. Each data set then consisted of a few thousand records of wave height, wave period and sea level, at a rate of a few hundreds per year over a period of about ten years. These 'high dependence' and 'low dependence' sub-sets were analysed in the same way as the overall data sets, and in some cases the directional bands were refined following initial results.

<u>Wave height & surge</u>

Surge and wave data were matched up hour-by-hour over the approximately ten year data sets, interpolating wave height (and period) between the three-hourly source records. A convenient way of extracting independent records was adopted, focusing on positive surges (the situation of interest for flood risk) and keeping consistency with the approach used for wave height & sea level. From these hourly joint data sets, one record was extracted at each positive maximum surge, subject to a minimum separation between maxima of 12 hours, i.e. no more than one per tide. The number of records extracted varied slightly from one tide gauge station to another, but was usually just under one per day. Each data set then consisted of a few thousand records of wave height, wave period and surge, at a rate of a few hundreds per year over a period of about ten years.

<u>Wind-sea & swell</u>

The source wave data records contain separate wind-sea and swell components at three-hourly intervals. The data preparation technique used here around the Scottish coast was developed during an earlier swell mapping project for England and Wales (HR Wallingford, 1997). The Met Office European wave model applies a rather arbitrary definition of swell, dividing the total wave energy into wind-sea and swell components for each record based only on the wind speed at the time and location of that record. The same filtering of swell records was applied here as was used in the previous study, with the general aim of retaining as swell only those records that would tend to be recognised as such by coastal engineers. Swells reported at wave periods below 8s, wave steepnesses above 0.02, and those at times of rapidly reducing wind speed, were re-designated as wind-sea. All three-hourly records with non-zero swell component were retained for use in subsequent joint probability analysis. The number of records varied between locations, depending upon exposure to swell, but was typically 50% of the total possible number of records. Each data set then consisted of around 15000 records of wind-sea wave height, wind-sea wave period and swell wave height, at a rate of one thousand or so per year over a period of about twelve years.

As there was no requirement for temporal continuity, potential records for which one or both variables were missing were simply excluded at data preparation stage. This would not bias the results or change the number of records per year, but would simply reduce the number of years of data in the sample.

3.2.3 Selection of JOIN-SEA parameter options

A few test analyses were run for each of the variable-pairs. The dependence parameter ρ was not particularly sensitive to threshold (representing the proportion of the top end of the joint distribution used) and so for consistency, all the main results are presented for a threshold of 0.9. That is to say that the analysis focused on the top ten percent of records of each variable, and in particular upon those records including values in the top ten percent of both variables.

In site-specific studies of large waves and high sea levels, HR Wallingford's normal practice has been to consider the use of up to three separate direction sectors (usually specified by wave direction). These are chosen to represent the different populations of data (e.g. south-westerly swell waves, south-easterly locally generated waves etc) that might be present in any data set, each of which might have a different level of dependence between the two variables. In the present analyses, direction sectors were used for analysis of wave height & sea level, but the sectors were set slightly differently for each location to reflect their differing exposure to different types of wave conditions.

For four locations (Avonmouth, Liverpool, Millport and Tobermory) the range of wave directions was relatively small, and so division into sectors would not have been helpful either for illustrative purposes or for later use of the dependence results. Analysis of wave height & surge and of wind-sea & swell was also not divided into direction sectors because the higher values of surge and swell tended to be limited to relatively narrow bands of wave direction at any particular location.

In a few instances, where dependence varied through a data sample, the alternative statistical dependence model available in JOIN-SEA, consisting of a mixture of two Bi-Variate Normal distributions, might have been preferred to the single BVN. However, to maintain continuity of approach throughout the analysis of around one hundred data sets, and continuity of results between neighbouring locations, the mixture model was not used during this project.

The choices explained in the preceding paragraphs were aimed at provision of dependence information in a consistent way for use in dependence mapping and simplified joint probability analysis. In a site-specific study using the analytical approach offered by JOIN-SEA, the mixture of two BVNs should be used if this gives a better representation of the statistical dependence. Similarly, alternative thresholds and/or direction sectors might be chosen to represent better the individual site characteristics.

3.2.4 Meaning of the ρ correlation parameter

Details of the Bi-Variate Normal and Bi-Variate Normal Threshold distributions are given in HR Wallingford (2000a) and are not repeated here.

- $0 < \rho < 1$ corresponds to positive dependence
- $\rho = 0$ corresponds to independence
- $-1 < \rho < 0$ corresponds to negative dependence

Typically, for the variable-pairs of interest in this study, above the 0.9 threshold (see Section 3.2.3) adopted for analysis in this project, ρ varies between about –0.1 (slight negative dependence) and 0.7 (strong positive dependence).

JOIN-SEA applies a normalisation process to each variable before dependence analysis (applied in reverse to return to actual values during the subsequent Monte Carlo simulation). Each distribution is transformed so that it has a Normal distribution with a mean of zero and a standard deviation of one. In this transformed state the joint distribution takes a shape similar to those illustrated in Figure 3.1 for different values of ρ. This transformation does not distort the dependence between the two variables, since the ranking of the individual records is unchanged. However, it would be misleading to draw extreme values directly from the diagrams in Figure 3.1 without reference to the reverse transformation function (which could be encapsulated in the form of the marginal extremes derived elsewhere).

3.2.5 Confidence intervals for ρ

There are three main sources of error and/or uncertainty.

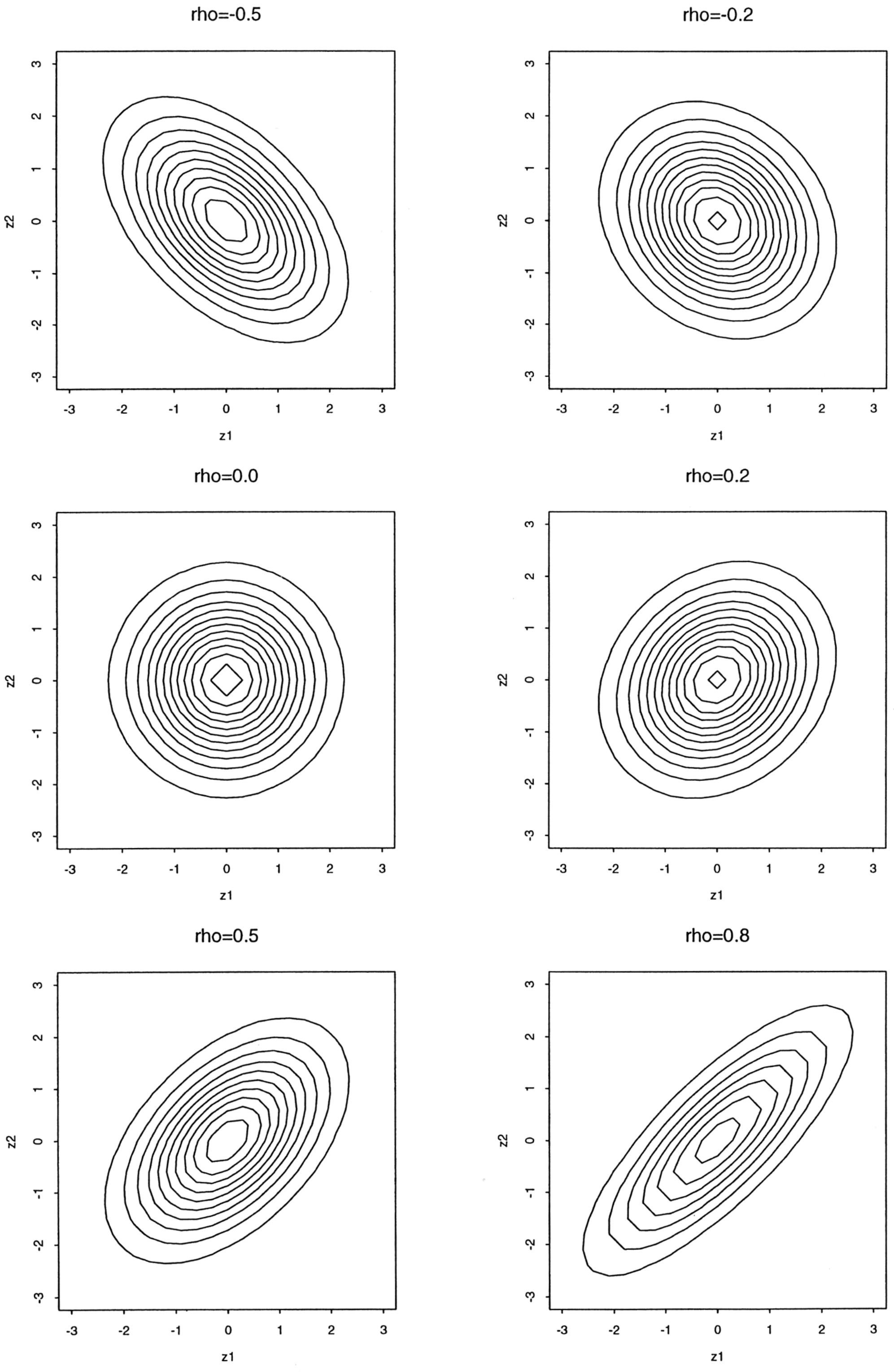
The data, in particular the wave data which are hindcast, may be inaccurate. There may be systematic error due, for example, to poor model calibration or poor instrument levelling. There will also be the general uncertainty involved in using a numerical model, with incomplete information on bathymetry and wind field, to predict wave conditions. Evaluation of these uncertainties is outside the scope of the present study but, in the context of mapping dependence, which does not depend critically on precise evaluation of the marginal variables, their impact will probably be small.

A poor choice of statistical model could introduce error, by forcing the data to follow an inappropriate statistical distribution, which may become even more inappropriate when extrapolated to extremes. Potential for this type of error was tested by means of a small number of sensitivity tests using alternative statistical models or thresholds, suggesting that the BVN statistical model with a fixed 0.90 threshold provides a robust estimate of ρ.

Estimation of the parameters of the chosen statistical model is subject to uncertainty, depending on the quantity and scatter of the data. This uncertainty can be evaluated in terms of the standard errors of the parameter estimates, including in this case the correlation coefficient ρ. The 95% confidence band for ρ, indicating limits above and below which there is only a 2.5% chance of the true value of ρ lying, is given by ±1.96 standard errors around the central estimate of ρ.

3.2.6 Checking of the results

Results were compared with those for neighbouring locations and with previous research and consultancy studies. This provided a general check on the suitability of the approach and its application, and helped to identify any occasional errors in the data preparation or analysis. In a few cases, the direction sectors initially chosen to separate the data into different types (e.g. south-westerly, northerly etc) were altered slightly in order to improve spatial consistency amongst the direction-dependent analyses. Apart from these minor refinements, the dependence results are presented as calculated, without any editorial adjustments.


rho=-0.5

rho=-0.2

rho=0.0

rho=0.2

rho=0.5

rho=0.8

Figure 3.1 Joint density contours for the bi-variate normal dependence structure when the marginal distributions are standard normal

3.3 Method used for river flow, precipitation and surge

A full description of the method used is given in Appendix 2, including the mathematics and justification for the approach. A shorter descriptive summary is given here.

3.3.1 The χ dependence measure

A dependence measure specially suited for estimating dependence as the variables reach their extremes was used. The measure, χ, has been described in detail by Buishand (1984) and Coles *et al.* (2000). Buishand employed it to assess the inter-station dependence in precipitation data, whereas Coles *et al.* applied it to several different variables, among them precipitation and surge data.

When used for bi-variate random variables with identical marginal distributions, the measure χ provides an estimate of the probability of one variable being extreme provided that the other one is extreme.

In this application the marginal distributions are not likely to be identical, and are therefore transformed to become so. Further, the marginals are unknown and must be estimated using their empirical distributions. Thus, one approach to obtaining an estimate of identical marginal distributions is to simply to rank each set of observations separately, and divide each rank with the total number of observations in each set. The data sets transformed in this way contain complete information about the joint distribution but no longer have information on the marginal distributions.

The value of χ can be interpreted as the risk (for a given threshold) that one variable is extreme, given that the other is. Suppose that one variable exceeds the threshold corresponding to a certain (small) exceedence probability. Then, if the dependence between the variables is estimated to be $\chi = 0.1$, it means that there is a 10% risk of the other variable exceeding the threshold corresponding to the same probability. Therefore once, on average, in ten successive periods of 10 years, one would expect the 10 year return period value of the first variable to be accompanied by the 10 year return period value of the second variable. In the same period of 100 years, one would expect ten occasions where 1 year return period values of both variables are exceeded during the same record, and a 10% chance of 100 year values of both variables. As the variables approach their extremes, $\chi = 1$ signifies total dependence and $\chi = 0$ signifies either independence or negative dependence.

3.3.2 Selection of threshold level

The data are transformed onto an annual maximum non-exceedence probability scale, enabling interpretation of the dependence between the variables in the familiar context of different return periods. The transformation is achieved through a peaks-over-threshold (POT) approach, which is considered to give a more accurate estimate of the probability distribution than using only the annual maximum series. The independence criterion used in this study was that two POTs must not occur on consecutive days, but be separated by at least three days.

The dependence measure χ can be estimated for any threshold. Initial trials showed a fairly constant, slightly decreasing, value of χ for annual maximum non-exceedence probabilities between about 0.1 and 0.5. For higher probabilities, χ tended to become zero as no observation-pair exceeded both thresholds (Appendix B of Svensson and Jones, 2000). The threshold selected was 0.1, corresponding to data values that about 2.3 events per year will exceed. The annual maximum will exceed this threshold in 9 out of 10 years. The use of a threshold in this sort of range is dictated by two requirements: to have enough data points above the threshold in order to be able to estimate dependence reliably, and for the threshold to be high enough to regard the data points as extreme.

Only observation-pairs where both observations in the pair were available were included in the analysis. A minimum of 1825 observation-pairs, equivalent to five complete years of simultaneous data, was set as a requirement for χ to be estimated reasonably reliably.

3.3.3 Significant dependence

The values of χ corresponding to the 5% significance level were estimated using a permutation method. This type of method is used to generate data sets in which independence would hold. A large number of data sets are generated and a test statistic, in this case χ, is calculated for each of these new data sets. This provides a sample of χ corresponding to independently occurring data. If the χ calculated for the original data set is rather different to most of the χ calculated from the generated values, then this suggests that the two original records are not independent.

In total, 199 permutations of the data were made for each station-pair and a new χ was calculated each time. The 199 χ values were subsequently ranked in descending order and the 10th largest value was accepted as corresponding to the 5% significance level, the actual value of which varied between about 0.02 and 0.09 in this study. If the χ calculated for the original series exceeded this value, then the data provide reasonably strong evidence that the dependence between the variables can be considered genuine.

3.3.4 Confidence intervals

Confidence intervals give an indication of the range of values within which the ‘true’ dependence χ can be expected to lie. A bootstrapping method, based on the generation of many new data set resamples, was used to estimate the confidence intervals. In contrast to when significance levels were estimated and independence between the two series was sought, each observation-pair is here kept intact and treated as one record.

The original sample of observation-pairs is used as the distribution from which the resamples are chosen randomly. A large number of data sets are generated and a test statistic, in this case χ, is calculated for each of these new data sets. This provides a sample of χ that would occur for a range of situations, as χ is calculated from some resamples including many data-pairs consistent with dependence, and from some resamples including many data-pairs consistent with independence.

In total, $B = 199$ bootstrap samples of the data were made for each station-pair and a new χ was calculated each time. The 199 χ values were subsequently ranked in descending order and the 10th and 190th largest values were accepted as delimiting the 90% confidence interval.

Because the computations were very computationally demanding, confidence intervals were estimated only for the primary variable-pair, surge & flow (precipitation being used only to aid in the interpretation of why dependence occurs).

3.3.5 Preparation of the data sets for analysis

Daily mean flows and daily precipitation accumulations were used without any further preparation. Daily mean river flows were used rather than flows of a higher resolution as the latter were not available in digitised form for many stations. Daily mean flows are indicative of the magnitude of the peak flow during the day, especially for more slowly responding catchments. However, it may not necessarily be the case that peak flows would be more appropriate to use, as water levels in the estuary are influenced by the possibly rather slow change in storage in the estuary.

Daily series of maximum surge, and of maximum surge occurring at high tide, were derived from the hourly records, for the water day, 9.00-9.00 GMT. The daily maximum surge was thought to have the strongest connection with precipitation and river flow because of the relatively short duration of very high surges. However, because of tide-surge interaction in shallow water areas, such as the south of the North Sea, daily maximum surge occurring at high tide was also extracted from the hourly data.

3.4 Method used for tide and surge

3.4.1 Discussion of the issues and development of the analysis method

Dependence in the time series of sea level, or its components the tide and the surge, can take several forms. These time series exhibit a self-dependence. That is, the time series are serially correlated through time such that each hourly observation is not statistically independent. This form of dependence is dealt with in the Joint Probability Method (JPM) and the Revised Joint Probability Method (RJPM) by the use of an Extremal Index.

Here the interest is with dependence between tide and surge. An example of this can be witnessed in the Thames estuary where the larger surges occur near the mid-tide level and are attenuated at high tidal levels. Extreme events can be caused by a combination of a large surge with a moderate tide. This is normally referred to as tide-surge interaction.

The methods used here are modifications of those developed by Pugh and Vassie (1980) and Tawn (1992). They were revised by Dixon and Tawn (1994, 1995 & 1997) to give the additional flexibility required for application to the wide range of known interaction characteristics and to improve numerical stability. The results in the latter documents have been slightly modified herein to generate a graphical representation of the degree of tide-surge interaction round the UK coastline and to give an indication of whether interaction is an important element in the estimation of extreme levels.

A key feature for successful application of the JPM and the RJPM is the accurate estimation of the tail of the surge probability distribution. If the tide and surge are known to be independent processes, the distribution of the surge conditional on tidal level is the same for all tidal levels. So standard estimation techniques for the extremes of a stationary sequence, such as the r-largest method, can be used. However, in shallow water areas, dynamic processes such as bottom friction cause the tidal and surge components to interact. Accounting for such interaction in the modelling of extreme surges is important, since ignoring this feature and proceeding as if the processes were independent is liable to result in a significant overestimation of return levels of the sea level.

Interaction characteristics vary from site to site due mainly to water depth and variations in topography. As interest here lies only in interaction between extreme surges and the associated tides, then results from dynamical studies, for example Prandle and Wolf (1978), cannot be applied. Consequently, the form of the interaction at any site must be estimated directly from the observed data. Using data from each site separately is problematic for sites with short record lengths. If very few large surges occur at high tidal levels, the cause could be either

1. the presence of interaction between the tide and surge, or

2. the tide and surge are independent, but by chance few occurred at high tidal levels.

Distinguishing between the two possibilities is critical in determining whether or not interaction exists for any given site. As longer records become available it will become feasible to make this distinction but, for the present, most problems are encountered using data from sites with short record lengths. However, as interaction varies smoothly over reasonable spatial scales, estimation of interaction should generally be improved by viewing the estimation as a spatial problem.

The standard oceanographic approach used to assess the level of tide-surge interaction from observations is to examine the standard deviation of the surge conditional on the tidal level (Prandle and Wolf, 1978; Pugh 1987; Walden *et al.*, 1982). Since interest here is in the extreme surges, a more relevant aspect of interaction is based on the extremes of the distribution of surges as a function of tidal level. Tests and models for interaction have been developed based on these characteristics. These tests for interaction between extreme surges and the associated tidal level are contained in Dixon and Tawn (1994, 1995 & 1997) and most of the present results were obtained from these reports.

3.4.2 Test for interaction of the extremes

By splitting the tidal range into equi-probable bands there are then an equal number of surge observations with associated tides in each band, and so a surge observation has an equal probability of falling in any one of the bands. If the surge and tide are independent processes, then the number of surges per tidal band expected to exceed a common level, u, would be the same. If the surge and tide interact then the largest surges would occur at mid-tidal bands and the smallest in the highest tidal band. Consequently the number of surges per tidal band which exceed a high level, u, would be expected to differ throughout the tidal range, with most occurring in the mid-tidal bands and least in the highest tidal band.

The fact that the largest surges occur near mid-tide on a rising tide is a feature of tide-surge interaction. Interaction between tide and surge is generally a characteristic of externally generated surges that propagate with the tide through a shallow water area. The dynamics of wave propagation in shallow areas prevents the maximum surge level from arriving coincident with high tide. It is possible to have a locally generated surge occur at high tide but it should be noted that occurrences of this type are accounted for naturally in the derivation of dependent probability distributions of tide and surge.

Taking u to be the 99.75% empirical quantile of the surge distribution, then under independence we expect $N_{obs} \times 0.0025/N_{bands} = e$ observations above u per tidal band, where N_{obs} is the number of hourly observations at the site. N_{bands} is the number of tidal bands chosen and usually lies between 5 and 10. Interaction can then be tested using one of the standard test statistics. This test statistic will be small when the observed number of large surges per tidal band is close to the number expected if the tide and surge are independent, but is large when there is interaction.

Dixon and Tawn applied this test to the UK east coast sites, resulting, as might be expected, in coherent estimates except where the data series were short. A similar but more powerful test followed the development of their Spatial Revised Joint Probability Method (SRJPM). In SRJPM the test statistic was estimated spatially using observed data from all available ports and also the output from a hydrodynamic tide-surge model of the European Shelf. The data were weighted according to the length of the observed time series, more weight being given to the longer series. In addition, the method allowed for trends in the surge parameters, such that its magnitude was allowed to vary with time.

3.4.3 Estimation of the interaction function a(X : d) and amplification factor $q_s(d)$

In Dixon and Tawn (1995) two separately estimated smoothed spatial estimates for two fundamental physical components of the conditional surge process were obtained. The surge amplification was represented by $q_s(d)$, where d is the position along the UK coast measured clockwise from Wick. Tide-surge interaction was represented by a(X : d) where X is the tidal level. The tide was transformed into a normalised distribution over the range [0 1] thereby providing an identical range for all sites so that the interaction function could be spatially interpreted.

The functions a(X : d) and $q_s(d)$ were obtained in a form that transformed the surge distribution into [0 1] space such that the transformed surge was stationary for all tidal levels. Consequently, if the function a(X : d), for any given position d, proved to be constant against tidal level then it could be concluded that the tide and surge were independent. Departure from a constant indicated tide-surge interaction. These parameters were found to change smoothly along the coast and with tidal state. Results are introduced and discussed in Section 4.4.

3.5 Simplified method for joint probability extremes

3.5.1 Introduction to the simplified method

The purpose of this section is to provide a simple method of constructing tables of joint exceedence extremes, using existing information on marginal (single variable) extremes and an estimate of the dependence between the two variables. These combined conditions are called joint exceedence extremes, and the associated joint exceedence return period refers to the average time between occasions when both variables simultaneously exceed their specified values. It is assumed that the distribution and extremes of both variables are known, for marginal return periods between about 0.1 and 100 years, and that a general level of dependence can be inferred from this report or from other previous studies.

The 'simplified method' is intended for use in situations where lack of time series data or low sensitivity to dependence does not justify the time and expense of a rigorous site-specific analysis. It is based on the method described in Section 3.5.3 of the CIRIA *Beach management manual* (CIRIA, 1996) for a joint exceedence return period of 100 years. It incorporates all of the CIRIA manual method, which remains valid, but extends it to one additional dependence band, and a number of additional joint exceedence return periods. This keeps faith with the CIRIA manual method for existing users who wish to continue using it, whilst giving it a slightly greater range of applicability. However, it is recommended that new users adopt the 'desk study approach' (which will incorporate and extend the 'simplified method') to be introduced in the accompanying best practice report (Defra / Environment Agency, 2003a).

3.5.2 The simplified method

The simplified method uses a 'correlation factor' (CF), not originally intended as the basis of a probability model, but as a descriptive representation of actual dependence relative to independence and full dependence. CF is the ratio of the actual frequency of occurrence of a particular joint exceedence event to its probability of occurrence if the two variables were independent. In the CIRIA manual, four example CF values are used from the possible range of 1-70700 (excluding negative dependence) for a 100 year joint exceedence return period for a data set based on 707 records per year (one record per tide). CF values of 2, 20, 100 and 500 represent levels of dependence 'none', 'modestly correlated', 'well correlated' and 'strongly correlated'. For these particular levels of dependence, results are given in terms of pre-computed combinations of two variables, expressed in terms of their marginal (single variable) return periods. There will be more than one such combination for any given joint exceedence return period, and in any particular structural calculation, it is necessary to test all such combinations in order to find the worst case for that particular structure. Although this method does not have the precision and flexibility of an analytical approach, it is very much quicker to apply.

Although the magnitude of CF depends upon both return period and number of records per year, the required levels of dependence (relative to independence and full dependence) implicit in the four verbal descriptions can be maintained in an objective way, as explained in Appendix 3. Tables 3.1-3.5 list joint exceedence extremes, expressed in terms of marginal return periods, for each of five joint exceedence return periods and for each of five levels of dependence, covering the range of possibilities found in UK studies. (A new top level of 'super correlation' has been added to cover the range of dependence observed in the surge results.) The colour coded dependence bands used in Figures 4.1-4.5 correspond approximately to the five specific levels of dependence used in Tables 3.1-3.5 and, in the absence of site-specific data, are intended to provide the dependence estimate needed for use of the simplified method.

Table 3.1 Combinations of two variables for a joint exceedence return period of 1 year

Variable 1 marginal return period (years)	Variable 2 marginal return period (years) for different levels of dependence				
	none	modestly correlated	well correlated	strongly correlated	super correlated
	CF = 2	CF = 7	CF = 18	CF = 44	CF = 82
					(surge only)
0.01	0.28	1.0	N/A	N/A	N/A
0.02	0.14	0.5	1.0	N/A	N/A
0.05	0.06	0.20	0.5	1.0	N/A
0.1	0.028	0.10	0.25	0.6	1.0
0.2	0.014	0.05	0.13	0.3	0.6
0.5	0.006	0.020	0.05	0.12	0.23
1	0.003	0.010	0.025	0.06	0.12

Table 3.2 Combinations of two variables for a joint exceedence return period of 5 years

Variable 1 marginal return period (years)	Variable 2 marginal return period (years) for different levels of dependence				
	none	modestly correlated	well correlated	strongly correlated	super correlated
	CF = 2	CF = 10	CF = 33	CF = 100	CF = 225
					(surge only)
0.01	1.4	5	N/A	N/A	N/A
0.02	0.7	3.5	N/A	N/A	N/A
0.05	0.3	1.4	5	N/A	N/A
0.1	0.14	0.7	2.3	5	N/A
0.2	0.07	0.35	1.2	3.5	5
0.5	0.03	0.14	0.5	1.4	3.2
1	0.014	0.07	0.23	0.7	1.6
2	0.007	0.035	0.12	0.35	0.8
5	0.003	0.014	0.05	0.14	0.32

Table 3.3 Combinations of two variables for a joint exceedence return period of 20 years

Variable 1 marginal return period (years)	Variable 2 marginal return period (years) for different levels of dependence				
	none	modestly correlated	well correlated	strongly correlated	super correlated
	CF = 2	CF = 14	CF = 55	CF = 215	CF = 542
					(surge only)
0.01	6	N/A	N/A	N/A	N/A
0.02	3	20	N/A	N/A	N/A
0.05	1.1	8	20	N/A	N/A
0.1	0.6	4	16	N/A	N/A
0.2	0.3	2	8	20	N/A
0.5	0.11	0.8	3	12	20
1	0.06	0.4	1.6	6	15
2	0.03	0.20	0.8	3	8
5	0.011	0.08	0.3	1.2	3
10	0.006	0.04	0.16	0.6	1.5
20	0.003	0.02	0.08	0.3	0.8

Table 3.4 Combinations of two variables for a joint exceedence return period of 100 years

Variable 1 marginal return period (years)	Variable 2 marginal return period (years) for different levels of dependence				
	none	modestly correlated	well correlated	strongly correlated	super correlated
	CF = 2	CF = 20	CF = 100	CF = 500	CF = 1500
					(surge only)
0.01	28	N/A	N/A	N/A	N/A
0.02	14	100	N/A	N/A	N/A
0.05	6	60	N/A	N/A	N/A
0.1	2.8	28	100	N/A	N/A
0.2	1.4	14	71	N/A	N/A
0.5	0.6	6	28	100	N/A
1	0.28	2.8	14	71	N/A
2	0.14	1.4	7	35	100
5	0.06	0.6	2.8	14	42
10	0.03	0.28	1.4	7	21
20	0.014	0.14	0.7	4	11
50	0.006	0.06	0.28	1.4	4
100	0.003	0.03	0.14	0.7	2.1

Table 3.5 Combinations of two variables for a joint exceedence return period of 500 years

Variable 1 marginal return period (years)	Variable 2 marginal return period (years) for different levels of dependence				
	none	modestly correlated	well correlated	strongly correlated	super correlated
	CF = 2	CF = 28	CF = 182	CF = 1170	CF = 4150
					(surge only)
0.01	140	N/A	N/A	N/A	N/A
0.02	70	500	N/A	N/A	N/A
0.05	28	400	N/A	N/A	N/A
0.1	14	200	N/A	N/A	N/A
0.2	7	100	500	N/A	N/A
0.5	2.8	40	260	N/A	N/A
1	1.4	20	130	500	N/A
2	0.7	10	60	400	N/A
5	0.28	4	26	170	500
10	0.14	2.0	13	80	300
20	0.07	1.0	6	40	150
50	0.028	0.4	2.6	17	60
100	0.014	0.20	1.3	8	30
200	0.007	0.10	0.6	4	15
500	0.003	0.05	0.26	1.7	6

Tables 3.1-3.5 were prepared on the assumption of one record per tide, effectively assuming an event duration of 12.4 hours. The CEH Wallingford dependence analysis is based on one record per day, effectively assuming an event duration of 24 hours. A numerical adjustment is necessary before Tables 3.1-3.5 can be applied to data based on one record per day, to allow for the slightly reduced difference in probability between the independent and dependent cases. The adjustment factors, to be applied to the numbers in the body of the tables in Columns 2-6 are given in Table 3.6.

Table 3.6 Adjustment factors needed to apply Tables 3.1-3.5 to one record per day data

To be applied as multipliers to marginal return periods listed in Columns 2-6 of Tables 3.1-3.5	none	modestly correlated	well correlated	strongly correlated	super correlated
	1.94	1.68	1.51	1.37	1.28

Use of Tables 3.1-3.6

1. For each joint return period selected, refer to the appropriate table from amongst Tables 3.1-3.5 (interpolating between tables for any additional return periods).

2. Note the contents of Column 1 (for Variable 1), together with one of Columns 2-6 (for Variable 2), corresponding to the assumed level of dependence.

3. If the record type is one-per-day, as opposed to one-per-tide, apply the appropriate factor from Table 3.6 to the numbers in the second noted column, subject to the condition that the marginal return period for Variable 2 should not exceed the joint exceedence return period.

4. Convert the noted figures from marginal return periods to actual values of Variables 1 and 2.

5. Use each combination of Variables 1 and 2 as input to relevant flood risk calculations. The worst case (from amongst the different combinations) will have a return period for that flood risk approximately equal to the joint exceedence return period. Please note that the worst case combination may vary from one flood risk to another.

Comment on flood risk probability

The extreme conditions derived using the simplified method are joint exceedence combinations of two primary input variables, for example large waves and high sea levels, usually assuming a nominal event duration equal to the interval between records. Assuming that peak values of the two variables exist throughout the duration of the event, the return period of any associated response, for example overtopping of a seawall, will typically be only half as much. (The difference between joint exceedence return period and the return period of any associated response is discussed in HR Wallingford, 2000a.) This approach would therefore appear to be unconservative for design by a factor of around two in terms of return period. However, the assumption that peak values of both variables will occur simultaneously, and not just within the same event (e.g. during the same tidal cycle or the same day) tends to be conservative by a factor of around two to four in terms of return period. Thus, joint exceedence return period is not the same as the return period of the response but, taking all the simplifying assumptions together, it is a reasonable approximation.

3.5.3 Example application of the simplified method

The task

Consider the joint probability of large waves and high sea levels to the west of the Shetland Islands. In Example 1, estimate the joint exceedence extremes for a return period of 100 years. In Example 2, estimate the joint exceedence extremes for waves from the west and north-west (225-360°N) for a return period of 50 years.

Information required from elsewhere

Marginal extremes for each of the two environmental variables are assumed to have been derived elsewhere. For illustrative purposes, assume the following extreme values, in each case for return periods of 0.1, 1, 10 and 100 years:

> from the overall distribution of wave heights 7.0, 10.0, 13.0 and 16.0m,
>
> from the, say, 55% of wave heights within direction sector 225-360°N, 6.0, 9.0, 12.0 and 15.0m,
>
> from the overall distribution of sea levels 1.25, 1.40, 1.55 and 1.70mOD,
>
> from the sub-set of sea levels given that wave direction is 225-360°N, 1.15, 1.30, 1.45 and 1.60mOD

Note 1: In practice it may be difficult to estimate extreme sea levels for a limited direction sector, in which case the overall extremes can be used.

Note 2: To interpolate between specified marginal return period values, imagine return period being plotted on a log scale against the actual value of the environmental variable on a natural scale.

Step 1
For the required return periods, refer to Table 3.4 for Example 1, and interpolate between Tables 3.3 and 3.4 for Example 2.

Step 2
For Example 1 (all directions combined, for Shetland) Figure 4.1 indicates that large wave heights and high sea levels are 'well correlated'. Therefore, note the contents of Columns 1 and 4 of Table 3.4 for use in production of joint exceedence extremes (results in Columns 1 and 2 of Table 3.7).

For Example 2 (west and north-west only) Figure 4.2 indicates that large wave heights and high sea levels are 'strongly correlated'. Therefore, note the contents of Columns 1 and 5 (interpolating between Tables 3.3 and 3.4) for use in production of joint exceedence extremes (results in Columns 3 and 4 of Table 3.7).

Step 3
For Example 2, apply the adjustment factor of 1.37 (Table 3.6) as the data set is roughly one record per day (results in Columns 5 and 6 of Table 3.7).

Step 4
Convert to actual wave heights and sea levels (results in Columns 7-10 of Table 3.7).

Step 5
Calculate any flood risk function(s) of interest. An example response function 'total water level', equal to wave height plus sea level, is used here for illustration purposes (results in Columns 11 and 12 of Table 3.7).

Table 3.7 Worked example of the simplified method

Return periods for H_s and sea level (years)						Environmental variables (m or mOD)					
Step 2				Step 3		Step 4				Step 5	
Example 1		Example 2		Example 2		Example 1		Example 2		Ex 1	Ex2
H_s	Level	H_s	level	H_s	level	H_s	level	H_s	level	T.W.L.	
0.1	100	0.5	50	0.7	50	7.0	1.70	8.5	1.55	8.7	10.1
0.2	71	1	25	1.4	25	8.0	1.67	9.5	1.51	9.7	11.0
0.5	28	2	13	2.7	13	9.0	1.62	10.3	1.47	10.6	11.8
1	14	5	5	5	7	10.0	1.57	11.0	1.42	11.6	12.4
2	7	10	2.5	10	3.4	11.0	1.52	12.0	1.37	12.5	13.4
5	2.8	20	1.3	20	1.8	12.0	1.47	13.0	1.33	13.5	14.3
10	1.4	50	0.5	50	0.7	13.0	1.42	14.0	1.28	14.4	15.3
20	0.7	------	------	------	------	14.0	1.37	------	------	15.4	------
50	0.28	------	------	------	------	15.0	1.32	------	------	16.3	------
100	0.14	------	------	------	------	16.0	1.27	------	------	17.3	------

3.6 Relationship between methods

The dependence measures ρ (used by HR Wallingford), χ (used by CEH Wallingford) and CF (used in the simplified method) are introduced earlier in this report. They are not interchangeable as they each assume a different dependence structure. The form of the dependence functions and the relationships between them, which depend on the number of records per year and on the return period, are discussed in Appendix 3.

In practice it is unlikely to be necessary to convert between different dependence parameters, but Table 3.8 provides illustrative values, for a return period of 100 years, for data sets based on one record per day (used by CEH) and on one record per tide (used in most HRW analyses).

Table 3.8 Approximate relationship between dependence measures for a return period of 100 years

ρ	365 records per year		706 records per year	
	χ	CF	χ	CF
0.0	0.000	1.0	0.000	1.0
0.1	0.007	2.2	0.004	2.4
0.2	0.018	5	0.012	6
0.3	0.037	12	0.026	14
0.4	0.064	26	0.048	34
0.5	0.10	60	0.082	85
0.6	0.16	150	0.13	210
0.7	0.24	350	0.21	600
0.8	0.35	1000	0.31	1700
0.9	0.52	3300	0.49	5900
1.0	1.00	36500	1.00	70600

In order to retain the precision of the dependence calculations undertaken during the present study, results are quoted in terms of the different dependence measures used for different variable-pairs. The different dependence measures will also be retained in the accompanying best practice report (Defra / Environment Agency, 2003a). This should not make them any more difficult to use than a single dependence measure, since the proposed software tool for generation of joint exceedence extremes will be able to take any of them as input.

The relationship between ρ and χ varies slightly with return period, particularly at relatively low dependence ($\rho < 0.25$), and the corresponding CF values depend strongly on return period. Negative values are not given in Table 3.8 since, although ρ can measure negative dependence, χ cannot take a negative value and CF is effectively undefined below a value of one.

4. RESULTS OF DEPENDENCE ANALYSIS

4.1 Format of results

The dependence results for each variable-pair are shown both as maps and as tables. The tables contain all results, expressed in terms of the dependence measure used for the analysis, including confidence limits.

The original intention had been to present the dependence results using common formats for all variable-pairs. Although it would have been easier to abide by this original intention, during the industry consultation, and as the project progressed, it became clear for a number of reasons that this approach was not the most helpful. Differences between the statistical models used by the three institutions involved in the analysis were significant and, although an approximate conversion could have been made, there would have been considerable loss of precision in so doing.

The original intention had been to limit the information presented in the maps to a common illustrative format, just using four or five different colours to indicate different ranges of dependence, corresponding to those used in the 'simplified method' (Section 3.5). However, the 'simplified method' (where a choice of one of the five colour-coded levels of dependence is made) will now be superseded by a new 'desk study approach' which is not limited either to a particular dependence measure or to pre-computed values of that measure. The loss of information and precision associated with simplifying the results down to a common colour-coded format is unnecessary, particularly for spatially separated variable-pairs which would be difficult to illustrate in this way. The mapping approach actually adopted incorporates the simplicity of the originally intended colour-coding approach for established users who prefer to continue using the 'simplified method'. It also provides more precise information needed as input to the alternative approaches to be described in the accompanying best practice report (Defra / Environment Agency, 2003a).

Similarly, it would have been possible to convert all ρ values to nearest equivalent χ values, or vice versa, but the different statistical models underlying the χ and ρ values would involve some distortion of the results in converting from one parameter to the other. As with the colour coding scheme, the convenience of converting all to a single dependence parameter would not justify the loss of precision in the results.

4.2 Results for waves, swell, sea level and surge

4.2.1 Dependence analysis results

These variables were analysed using the HR Wallingford JOIN-SEA method, and the results are expressed in terms of the correlation coefficient, ρ, above a chosen threshold applicable to all analyses. The main results, intended for routine use in subsequent studies, are best estimates of ρ for a threshold of 0.90, i.e. focussing on data records in which both variables are expected to be exceeded no more than 10% of the time. However, to give an idea of the uncertainty involved, some alternative results are also tabulated.

Correlation coefficients for thresholds of 0.90 and 0.95 are listed in Tables 4.1-4.5. In the tables, columns headed 'Best', 'Low' and 'High' represent the best estimates of ρ and alternative estimates set one standard error below and above those best estimates. 'Low' and 'High' correspond to the 69% confidence interval for ρ; 90% and 95% confidence intervals can be found as ±1.65 and ±1.96 standard errors around the best estimate. The highlighted column '**Best**' under 'Threshold: 0.90' is considered to be the most robust estimate of the ρ values, appropriate for routine use.

Tables 4.1-4.3 list correlation coefficients for wave height & sea level for each location analysed, for all directions combined, and for sectors in which dependence was expected to be higher than average and lower than average for any particular location. These sectors are defined by wave direction, as explained in Section 3.2.3, and vary from one location to another. The directional bounds (wave directions from) of the higher and lower dependence sectors are listed in Tables 4.2 and 4.3, respectively.

Table 4.4 gives the correlation coefficients between wave height & positive surge for all directions combined. Table 4.5 presents the correlation coefficients for wind-sea wave height & swell wave height in Scotland. (Analysis for England and Wales had been done previously (HR Wallingford, 1997) using a different method, not producing a precise value of ρ.)

The best estimates of correlation coefficients for a threshold of 0.90 around the UK coast are plotted in Figures 4.1-4.5, corresponding to the results listed in Tables 4.1-4.5. Figure 4.1 shows ρ for wave height & sea level for all directions combined. Figures 4.2 and 4.3 show ρ for wave height & sea level for the higher and lower dependence sectors, respectively, together with sectors and arrows denoting the mean direction of the sector. Figure 4.4 shows ρ for wave height & positive surge. Figure 4.5 shows ρ for wind-sea wave height & swell-sea wave height for Scotland, and inferred ρ-bands for England and Wales from earlier analysis (HR Wallingford, 1997). In the figures, the five colours used correspond approximately to the five example levels of dependence used in the simplified method (Section 3.5), and the actual values given for CF are applicable only to a 100 year joint return period:

- green is for $\rho \leq 0.11$ where the variables of interest are effectively independent of each other, corresponding to CF = 2 (representing the range 1.0-2.5, 'independent');

- blue is for $0.12 \leq \rho \leq 0.37$ where there is a modest dependence between the two variables, corresponding to CF = 20 (representing the range 2.5-25, 'modestly correlated');

- pink is for $0.38 \leq \rho \leq 0.53$ where the two variables are well related, corresponding to CF = 100 (representing the range 25-125, 'well correlated');

- red is for $0.54 \leq \rho \leq 0.70$ where there is a strong dependence between the two variables, corresponding to CF = 500 (representing the range 125-600, 'strongly correlated');

- purple is for $\rho > 0.70$ where there is a very strong dependence between the two variables, corresponding to CF = 1500 (representing 600+, 'super correlated').

Table 4.1 Correlation coefficient (ρ, wave height & sea level): all wave directions combined

Station	Threshold: 0.90			Threshold: 0.95		
	Best	Low	High	Best	Low	High
Lerwick	**0.502**	0.455	0.545	0.500	0.434	0.561
Wick	**0.302**	0.246	0.356	0.265	0.180	0.347
Aberdeen	**0.213**	0.153	0.273	0.239	0.152	0.322
North Shields	**0.172**	0.109	0.234	0.227	0.136	0.314
Immingham	**0.162**	0.100	0.223	0.222	0.133	0.307
Lowestoft	**0.420**	0.368	0.469	0.524	0.459	0.582
Sheerness	**0.086**	0.023	0.150	0.132	0.033	0.229
Dover	**0.079**	0.012	0.146	0.135	0.037	0.230
Newhaven	**0.167**	0.104	0.228	0.175	0.083	0.265
Portsmouth	**0.410**	0.359	0.458	0.463	0.394	0.527
Weymouth	**0.374**	0.323	0.423	0.391	0.319	0.459
Newlyn	**0.198**	0.140	0.254	0.188	0.103	0.271
Ilfracombe	**0.110**	0.041	0.179	0.151	0.050	0.250
Avonmouth	**0.169**	0.103	0.235	0.178	0.079	0.275
Milford Haven	**0.280**	0.227	0.332	0.305	0.228	0.378
Fishguard	**0.367**	0.318	0.414	0.394	0.326	0.458
Holyhead	**0.435**	0.377	0.489	0.411	0.324	0.490
Liverpool	**0.200**	0.136	0.263	0.302	0.216	0.385
Heysham	**0.289**	0.234	0.342	0.351	0.273	0.424
Portpatrick	**0.584**	0.546	0.619	0.632	0.583	0.676
Millport	**0.550**	0.509	0.588	0.603	0.550	0.652
Tobermory	**0.395**	0.343	0.443	0.387	0.312	0.457
Ullapool	**0.248**	0.189	0.305	0.259	0.175	0.340
Note 1. 'Low' and 'High' represent one 'standard error' below and above the best estimate. To obtain 90% and 95% confidence limits for ρ, take ±1.65 or ±1.96 standard errors around the best estimate. For example, the 95% confidence limits for Avonmouth would be 0.040 and 0.298.						

Table 4.2 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is higher

Station	Direction sector	Threshold: 0.90			Threshold: 0.95		
		Best	Low	High	Best	Low	High
Lerwick	220^0-45^0	**0.557**	0.506	0.605	0.555	0.480	0.621
Wick	45^0-180^0	**0.173**	0.076	0.267	0.202	0.062	0.334
Aberdeen	330^0-45^0	**0.222**	0.109	0.329	0.213	0.041	0.373
North Shields	330^0-45^0	**0.359**	0.268	0.443	0.425	0.303	0.534
Immingham	330^0-45^0	**0.353**	0.266	0.434	0.408	0.290	0.514
Lowestoft	330^0-45^0	**0.789**	0.746	0.824	0.831	0.782	0.869
Sheerness	330^0-45^0	**0.371**	0.228	0.498	0.435	0.237	0.599
Dover	330^0-45^0	**0.161**	0.034	0.285	0.227	0.047	0.395
Newhaven	210^0-280^0	**0.169**	0.082	0.254	0.179	0.049	0.304
Portsmouth	70^0-210^0	**0.547**	0.430	0.645	0.682	0.459	0.821
Weymouth	70^0-210^0	**0.482**	0.407	0.549	0.496	0.391	0.588
Newlyn	60^0-180^0	**0.424**	0.250	0.570	0.367	0.100	0.584
Ilfracombe	210^0-330^0	**0.127**	0.054	0.199	0.146	0.038	0.251
Milford Haven	180^0-270^0	**0.271**	0.207	0.332	0.310	0.221	0.395
Fishguard	180^0-270^0	**0.362**	0.305	0.417	0.391	0.310	0.465
Holyhead	180^0-270^0	**0.440**	0.367	0.508	0.429	0.321	0.527
Heysham	220^0-280^0	**0.320**	0.249	0.387	0.347	0.247	0.439
Portpatrick	150^0-270^0	**0.616**	0.567	0.660	0.656	0.592	0.711
Ullapool	210^0-290^0	**0.396**	0.290	0.492	0.412	0.260	0.543

Note 1. 'Low' and 'High' represent one 'standard error' below and above the best estimate. To obtain 90% and 95% confidence limits for ρ, take ±1.65 or ±1.96 standard errors around the best estimate.

Note 2. Avonmouth, Liverpool, Millport and Tobermory were not divided into sectors, as the range of wave directions was already limited.

Table 4.3 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is lower

Station	Direction sector	Threshold: 0.90			Threshold: 0.95		
		Best	Low	High	Best	Low	High
Lerwick	45^0-220^0	**0.361**	0.262	0.452	0.445	0.316	0.557
Wick	330^0-45^0	**0.111**	-0.002	0.222	0.140	-0.026	0.299
Aberdeen	45^0-180^0	**0.123**	0.024	0.219	0.077	-0.075	0.225
North Shields	45^0-180^0	**-0.010**	-0.135	0.116	0.065	-0.123	0.246
Immingham	45^0-180^0	**-0.079**	-0.207	0.051	-0.057	-0.257	0.143
Lowestoft	45^0-180^0	**0.035**	-0.078	0.149	0.053	-0.121	0.224
Sheerness	45^0-180^0	**0.068**	-0.025	0.161	0.149	0.014	0.279
Dover	210^0-280^0	**0.054**	-0.040	0.149	0.041	-0.111	0.189
Dover	70^0-210^0	**-0.049**	-0.280	0.192	0.145	-0.186	0.446
Newhaven	70^0-210^0	**0.155**	0.052	0.255	0.207	0.060	0.347
Portsmouth	210^0-280^0	**0.400**	0.332	0.464	0.444	0.351	0.529
Weymouth	210^0-280^0	**0.286**	0.209	0.359	0.331	0.223	0.431
Newlyn	180^0-360^0	**0.209**	0.149	0.268	0.204	0.114	0.291
Milford Haven	270^0-30^0	**-0.092**	-0.239	0.060	-0.149	-0.402	0.104
Fishguard	270^0-30^0	**-0.122**	-0.278	0.041	Insufficient data		
Holyhead	270^0-60^0	**0.034**	-0.113	0.181	0.151	-0.068	0.356
Heysham	280^0-330^0	**-0.082**	-0.273	0.117	-0.006	-0.303	0.283
Portpatrick	270^0-30^0	**0.110**	0.010	0.210	0.114	-0.039	0.262
Ullapool	290^0-45^0	**-0.036**	-0.136	0.064	-0.112	-0.278	0.054

Note 1. 'Low' and 'High' represent one 'standard error' below and above the best estimate. To obtain 90% and 95% confidence limits for ρ, take ±1.65 or ±1.96 standard errors around the best estimate.

Note 2. Avonmouth, Liverpool, Millport and Tobermory were not divided into sectors, as the range of wave directions was already limited.

Note 3. Dover has two low dependence sectors as its exposure to waves from widely different directions required the use of three sectors altogether.

Table 4.4 Correlation coefficient (ρ, wave height & surge) all wave directions combined

Station	Threshold: 0.90			Threshold: 0.95		
	Best	Low	High	Best	Low	High
Lerwick	**0.676**	0.620	0.723	0.671	0.589	0.738
Wick	**0.426**	0.344	0.500	0.335	0.205	0.453
Aberdeen	**0.403**	0.321	0.478	0.333	0.206	0.449
North Shields	**0.389**	0.307	0.464	0.386	0.266	0.493
Immingham	**0.567**	0.510	0.618	0.530	0.441	0.607
Lowestoft	**0.659**	0.604	0.706	0.655	0.574	0.722
Sheerness	**0.459**	0.395	0.518	0.471	0.380	0.552
Dover	**0.568**	0.510	0.620	0.514	0.423	0.595
Newhaven	**0.682**	0.638	0.722	0.712	0.654	0.761
Portsmouth	**0.754**	0.714	0.787	0.767	0.714	0.811
Weymouth	**0.730**	0.690	0.765	0.740	0.683	0.786
Newlyn	**0.619**	0.566	0.665	0.619	0.544	0.683
Ilfracombe	**0.737**	0.694	0.775	0.738	0.676	0.788
Avonmouth	**0.754**	0.720	0.784	0.795	0.754	0.829
Milford Haven	**0.814**	0.786	0.838	0.824	0.786	0.855
Fishguard	**0.839**	0.815	0.859	0.841	0.808	0.868
Holyhead	**0.834**	0.797	0.864	0.845	0.795	0.881
Liverpool	**0.787**	0.754	0.815	0.779	0.730	0.819
Heysham	**0.691**	0.645	0.731	0.629	0.550	0.695
Portpatrick	**0.833**	0.804	0.857	0.848	0.811	0.876
Millport	**0.704**	0.656	0.746	0.676	0.599	0.739
Tobermory	**0.646**	0.583	0.701	0.630	0.534	0.707
Ullapool	**0.688**	0.636	0.733	0.704	0.630	0.763
Note 1. 'Low' and 'High' represent one 'standard error' below and above the best estimate. To obtain 90% and 95% confidence limits for ρ, take ±1.65 or ±1.96 standard errors around the best estimate.						

Table 4.5 Correlation coefficient (ρ, wind-sea H_s & swell H_s) all wave directions combined

Station	Threshold: 0.90			Threshold: 0.95		
	Best	Low	High	Best	Low	High
Lerwick	**0.156**	0.121	0.191	0.148	0.096	0.200
Wick	**0.390**	0.358	0.421	0.414	0.370	0.457
Aberdeen	**0.424**	0.390	0.456	0.429	0.382	0.475
Portpatrick	**0.354**	0.316	0.391	0.307	0.248	0.363
Millport	**0.077**	0.041	0.113	0.041	-0.016	0.097
Tobermory	**0.450**	0.424	0.475	0.363	0.321	0.403
Ullapool	**-0.012**	-0.056	0.032	0.023	-0.044	0.091
Note 1. 'Low' and 'High' represent one 'standard error' below and above the best estimate. To obtain 90% and 95% confidence limits for ρ, take ±1.65 or ±1.96 standard errors around the best estimate.						

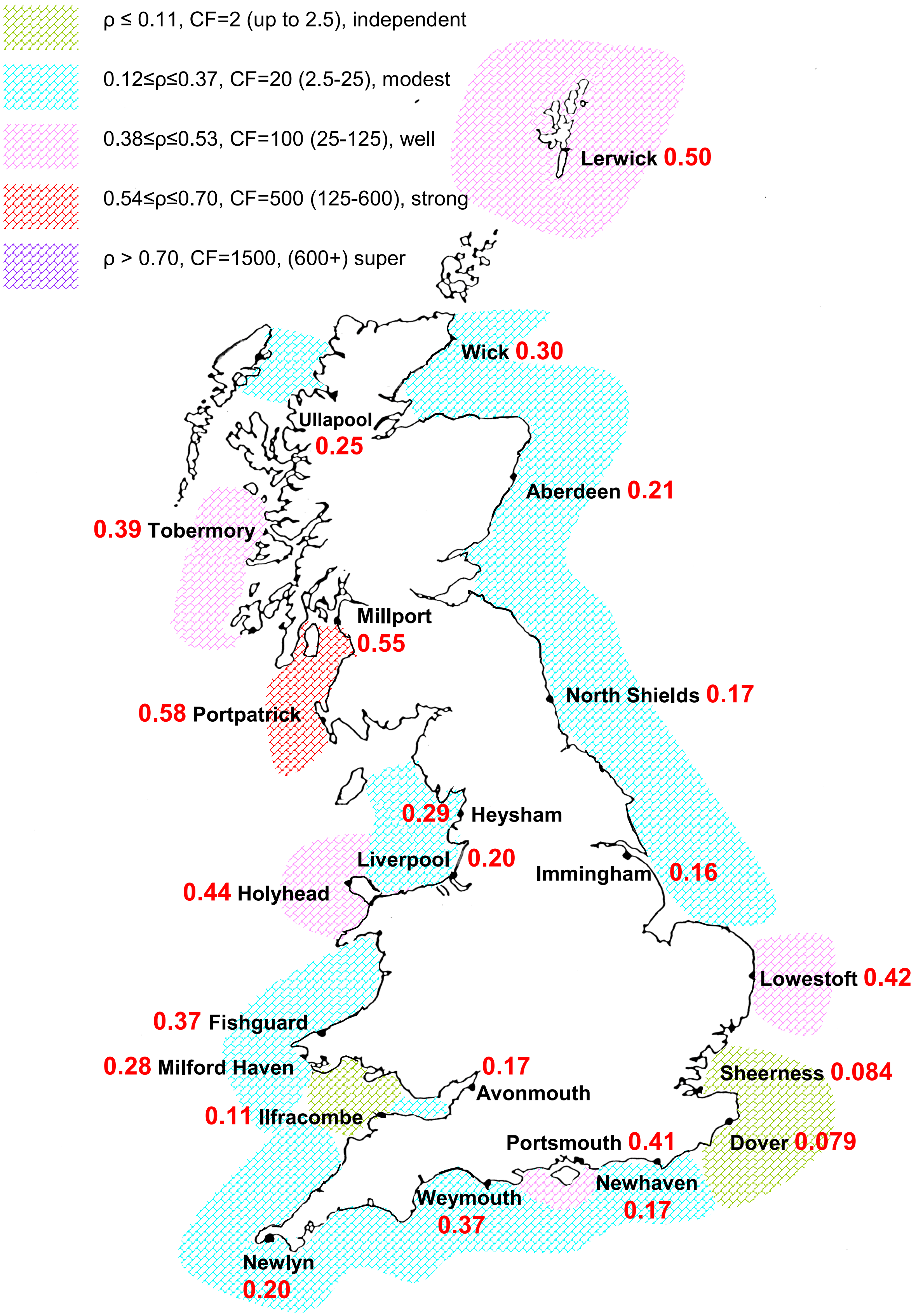


Figure 4.1 **Correlation coefficient (ρ, wave height & sea level): all wave directions combined**

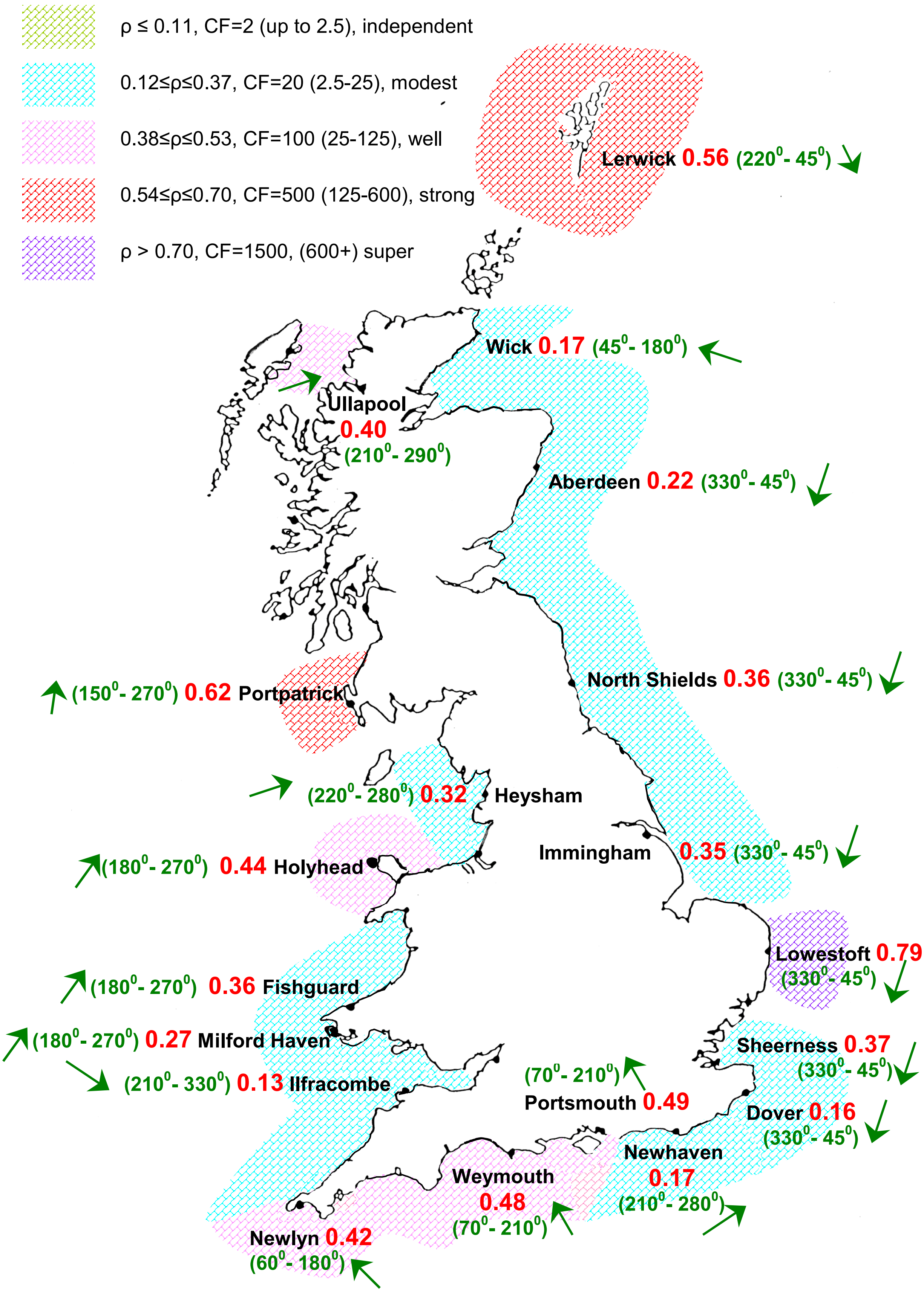


Figure 4.2 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is higher

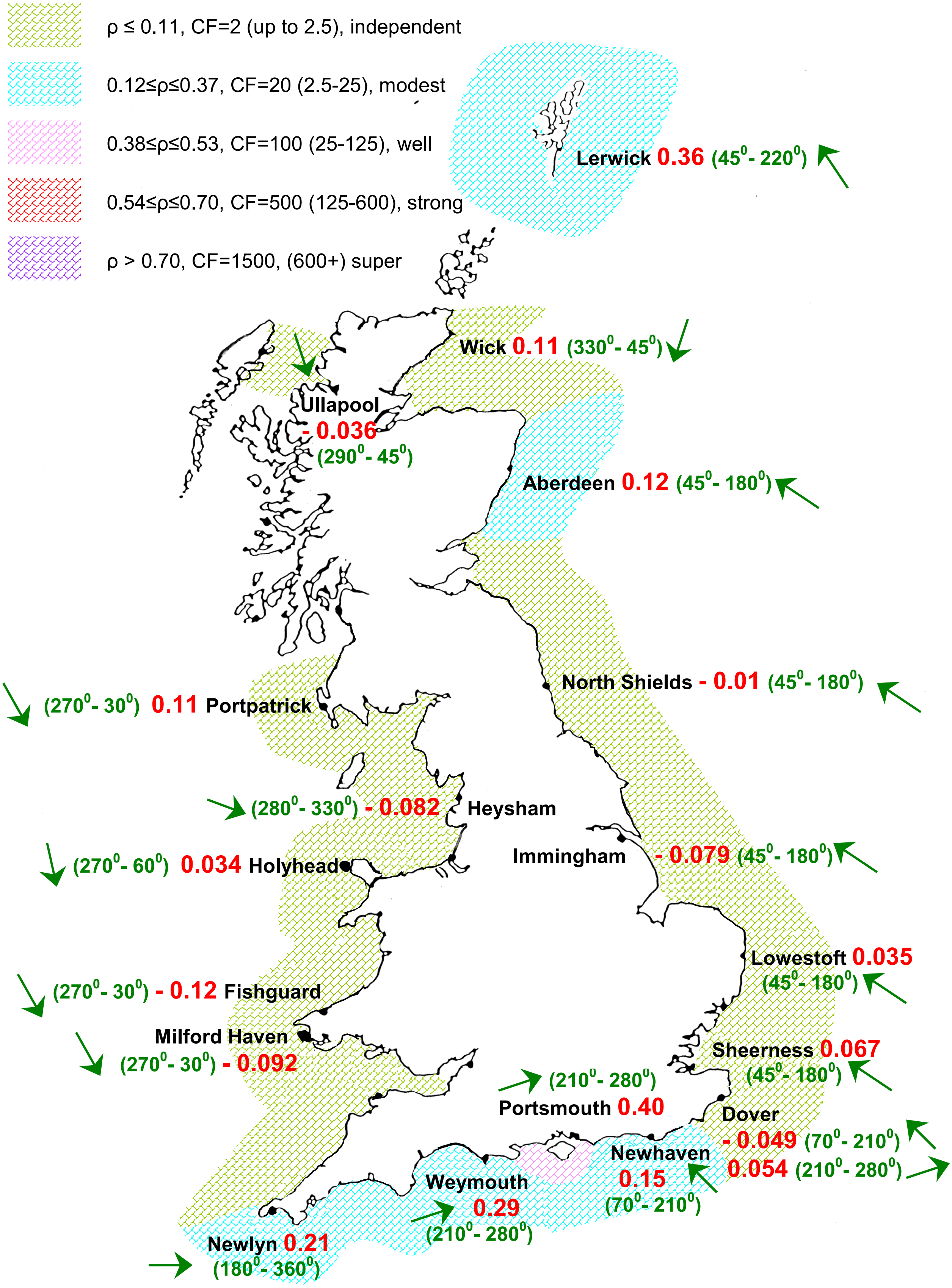


Figure 4.3 Correlation coefficient (ρ, wave height & sea level): wave direction sector in which dependence is lower

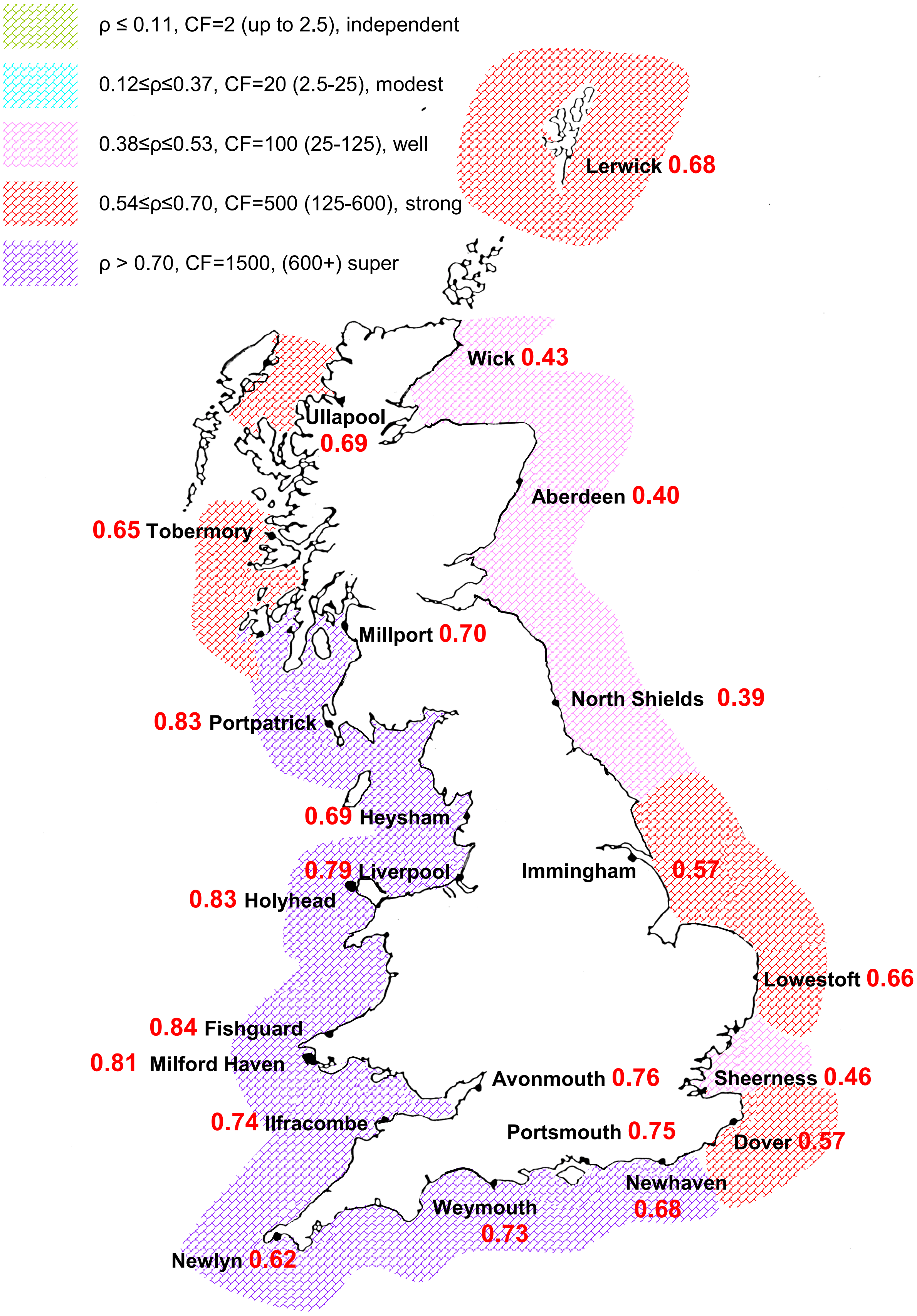


Figure 4.4 Correlation coefficient (ρ, wave height & surge): all wave directions combined

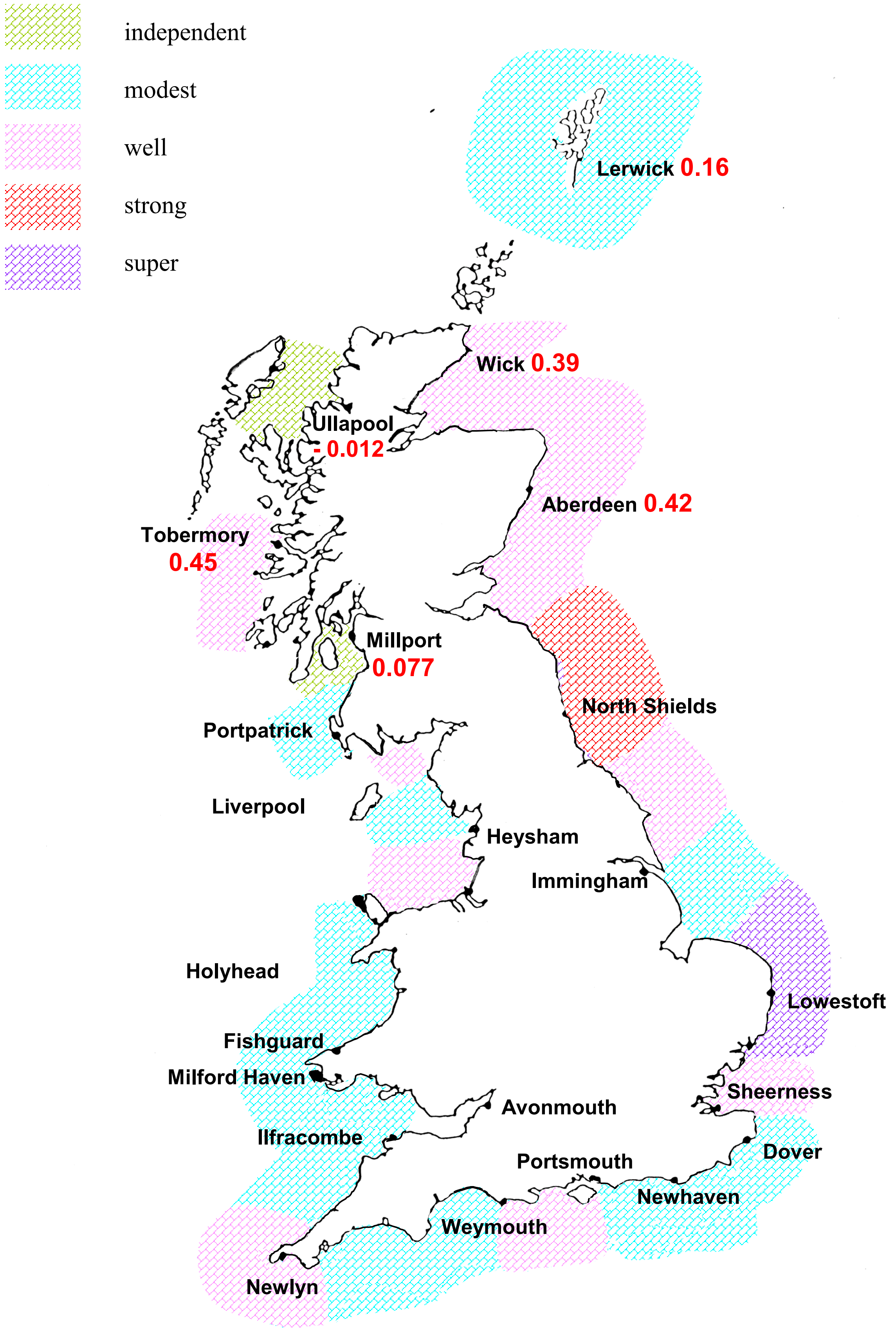


Figure 4.5 Correlation coefficient (ρ, $Hs_{\text{wind-sea}}$ and Hs_{swell}): all wave directions combined

4.2.2 Example tests of sensitivity to correlation coefficient ρ

Tables 4.1-4.5 provide best estimates of ρ for several different variable-pairs and for several different locations. The tables also provide alternative estimates of ρ which could be used to test the sensitivity of extremes predictions to uncertainties in ρ. Two locations were selected for sensitivity tests on wave height and sea level for all directions combined: Weymouth ($\rho = 0.374$, modestly correlated) and Portpatrick ($\rho = 0.584$, strongly correlated).

1000 year simulations were made for each site, for the best estimate of ρ, and also for the corresponding high and low values of ρ given in Table 4.1. The resulting joint exceedence curves for joint exceedence return periods of 1, 10 and 100 years are plotted in Figures 4.6 and 4.7.

As one would expect, for any particular group of three lines in Figures 4.6 and 4.7, the high value of ρ tends to give the highest joint exceedence values, and the low value of ρ the lowest joint exceedence results. However, the differences are quite small and, for the 100 year joint return period, are of the same order of magnitude as the general uncertainties associated with predicting a 100 year event from a 1000 year simulation (illustrated by the lack of smoothness in the plotted lines).

These example calculations suggest that sensitivity to ρ, given that a reliable best estimate of ρ is available from this report, is a fairly small part of the overall uncertainties associated with prediction of extreme sea states and their impacts. They also indicate that any serious attempt to evaluate that sensitivity should be based either on a simulation about one hundred times longer than the joint return period of interest, or on an alternative analytical approach.

4.2.3 Example comparisons between JOIN-SEA and the simplified method

The two examples used in Section 4.2.2 were also used to compare joint exceedence predictions produced by JOIN-SEA with those that would be produced by the simplified method described in Section 3.5. The joint exceedence curves from Figures 4.6 and 4.7, for waves and sea levels at Weymouth and Portpatrick corresponding to best estimates of ρ, are carried forward to Figures 4.8 and 4.9.

Figure 4.1 indicates 'modestly correlated' and 'strongly correlated' for waves and sea levels at Weymouth and Portpatrick, respectively. Application of these levels of dependence in Tables 3.1-3.4 (interpolating between Tables 3.2 and 3.3 for the 10 year joint return period) gives the (shorter) curves plotted in Figures 4.8 and 4.9 for comparison with equivalent JOIN-SEA predictions.

The comparison between JOIN-SEA and the simplified method is good for all three return periods at Weymouth, good at the highest return period for Portpatrick and fair for the other two return periods at Portpatrick. The observed differences are consistent with the intention that the simplified method should be conservative, and with the way that it uses only one of a small number of pre-computed joint exceedence curves.

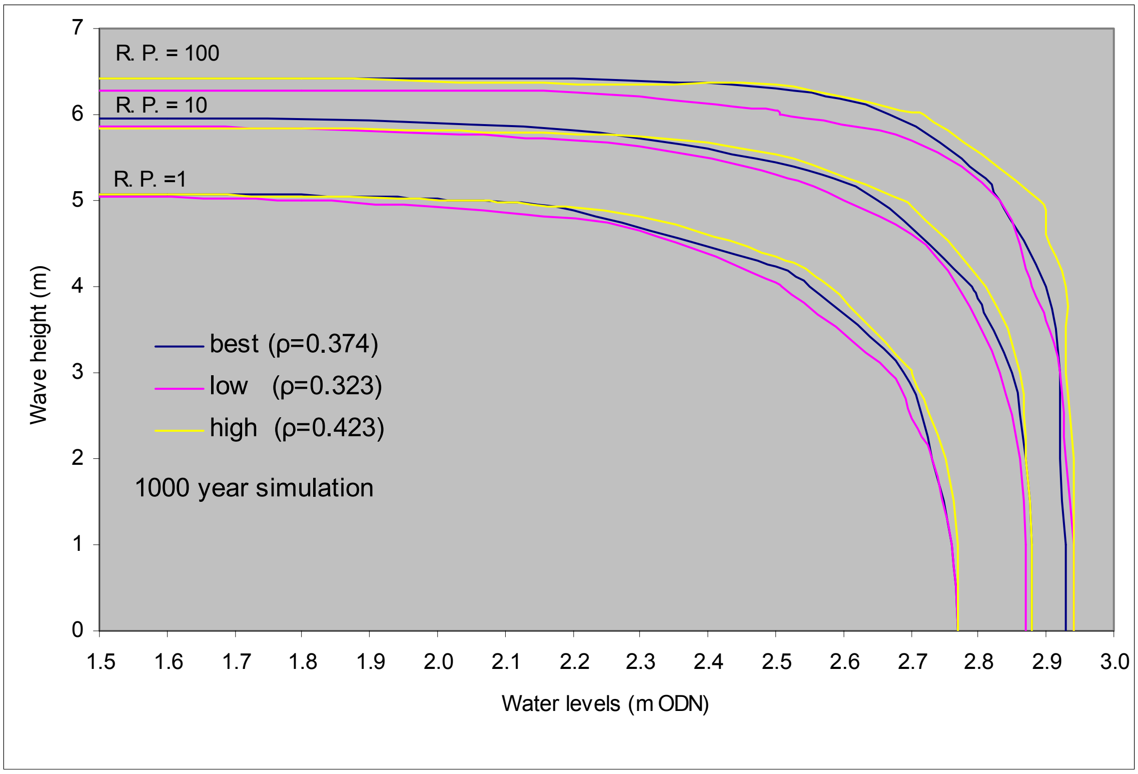


Figure 4.6 Weymouth joint exceedence curves

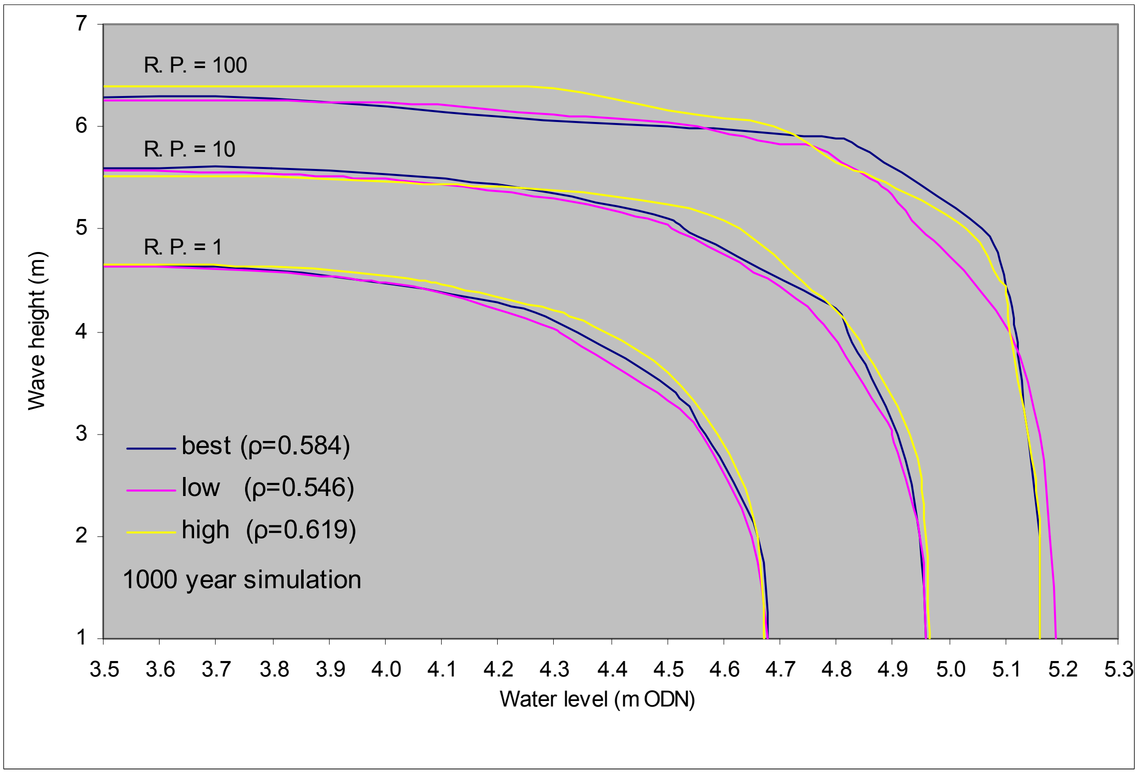


Figure 4.7 Portpatrick joint exceedence curves

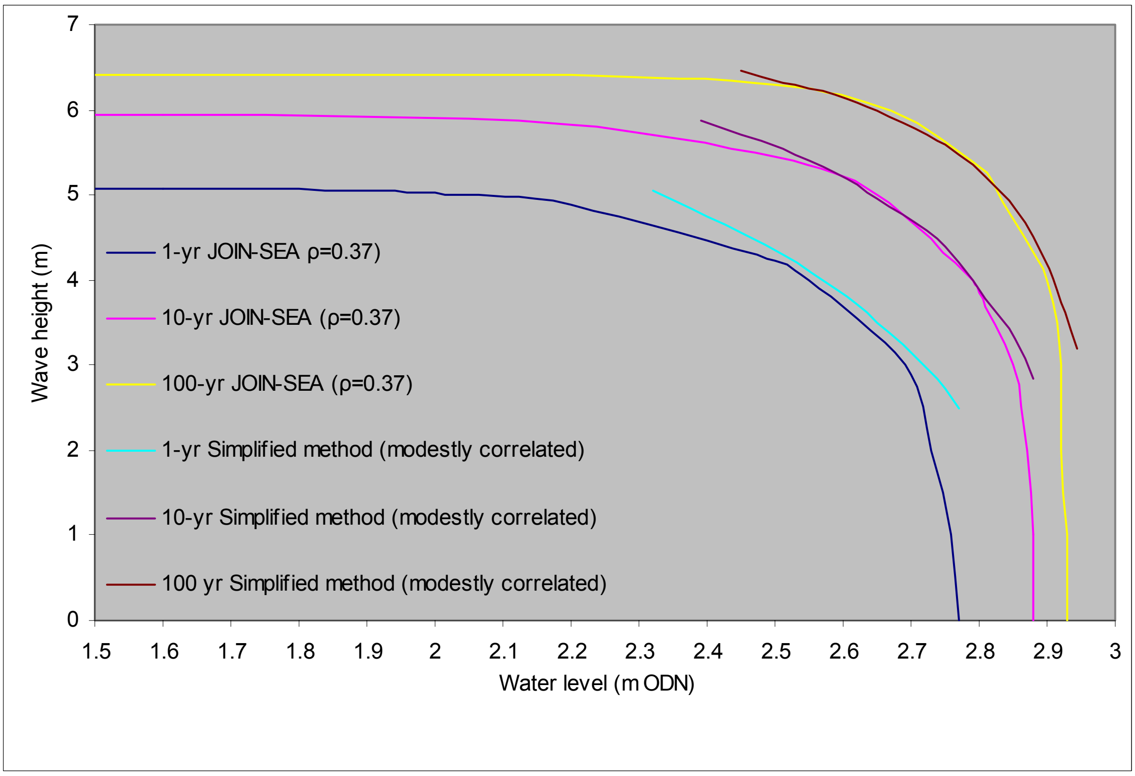


Figure 4.8 Weymouth JOIN-SEA method vs simplified method

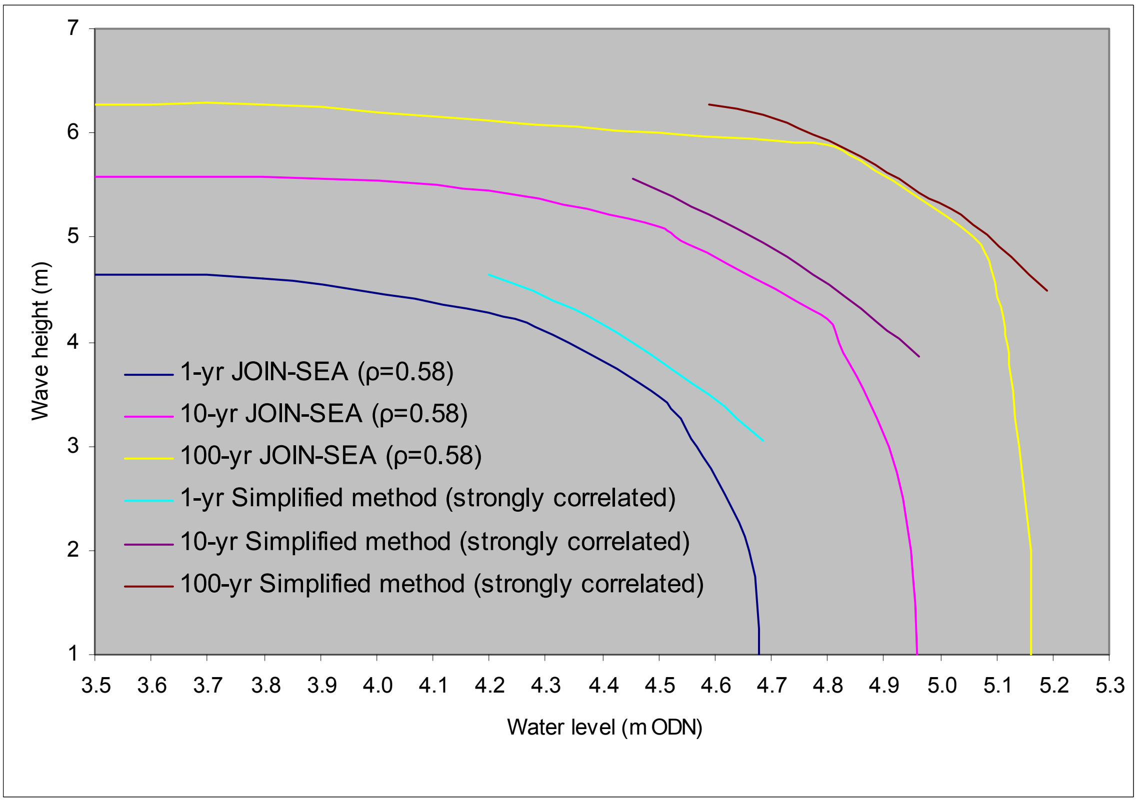


Figure 4.9 Portpatrick JOIN-SEA method vs simplified method

4.3 Results for river flow, precipitation and surge

The dependence measure χ (Section 3.3) was used to estimate the dependence between extreme river flow and surge and between extreme precipitation and surge. River flow and surge both influence the water levels in an estuary, whereas precipitation is used in the present study only to assist in the interpretation of why surge-flow dependence occurs in particular places and not in others.

Because the variables are used at a daily resolution, results are indicative only of where extreme river flow and surge *may* occur simultaneously. Modelling of how sea levels and river flow affect the water levels in the estuary needs to be done at a higher resolution to assess actual estuary water levels. However, if there were no dependence on a daily basis, such as presented in this study, it is highly unlikely that there would be dependence at a higher temporal resolution.

The same-day, all-year, analysis is discussed in detail, whereas lagged, seasonal and climate change analyses are discussed only briefly. See Defra / Environment Agency (2003b) for further information on these topics.

4.3.1 Dependence between river flow & surge

Results of the dependence analysis are presented in tables, and graphically on maps where pairs of stations with dependence exceeding a particular value are connected by lines (e.g. Figure 4.10). For the flow-surge variable-pair, dependence is estimated only for neighbouring stations. That is, one surge station on either side of the river estuary is paired with the river flow station, unless the surge station is located in, or very near, the estuary in which case only that surge station is used. Because of the short surge records for Portsmouth and Weymouth, the river flow stations between these surge stations have also been paired with the long surge record at Newlyn. The two river flow gauges on the north coast, 96001 and 97002, have been paired with surges at Ullapool, Lerwick and Wick. Note that Liverpool has two surge records, Princes Pier and Gladstone Dock. River flow stations between Heysham and Holyhead have therefore also been paired with (up to) three surge stations. One station-pair (Liverpool Princes Pier and 68020) has too few simultaneous observations for χ to have been estimated. However, the record for Liverpool Gladstone Dock sufficiently overlaps that of 68020. The station-pairs used are listed in Tables 4.6 and 4.7.

Dependence between river flow and surge can vary over short distances as each river responds differently, depending on its catchment characteristics such as area and geology. Small and impervious catchments generate faster runoff with a shorter time to peak flow than larger and more permeable catchments. Because there is less local variation in the sea, dependence in the surge variable is stronger over long distances than is dependence in the flow variable (not shown, see Svensson and Jones, 2000; Defra / Environment Agency, 2003b). The site-specific nature of the river flow characteristics means that a dense network of gauges is needed, and results may be difficult to generalise to a larger area. However, some regional patterns emerge.

Figure 4.10 shows dependence between river flow and daily maximum surge around the coast of Great Britain. Although dependence significant at the 5% level may be found at catchments spread along most of the coastline, higher dependence ($\chi > 0.1$) is generally found in catchments in hilly areas with a southerly to westerly aspect. Here, precipitation in south-westerly airflow, which is generally the quadrant of prevailing winds (e.g. Barrow and Hulme, 1997), will be orographically enhanced as the first higher ground is encountered. The sloping catchments may respond quickly to the abundant rainfall, and the flow peak may arrive in the estuary on the same day as a large surge occurs.

There are four regions where surge-flow dependence generally exceeds $\chi > 0.1$: the north side of the Firth of Forth (which although on the east coast, is the first hilly area encountered in south-westerly air flow), the western part of the English south coast, southern Wales, and around the Solway Firth. Table 4.6 shows the estimated dependence, χ, and the associated 5% significance level and limits of the 90% confidence interval.

The generally low dependence on the eastern part of the south coast of England may be related to these being generally permeable, predominantly chalk, catchments which respond slowly to rainfall. Runoff may therefore not form on the same day as the surge occurs.

In shallow water, wave characteristics such as speed and amplitude are influenced by water depth. When the increase in water depth due to tide and surge is not negligible compared to the total water depth, complex non-linear interaction between tide and surge will occur. In order to reduce the influence of this problem on the dependence analysis, the dependence between river flow and daily maximum surge occurring at high tide was estimated (Figure 4.11, Table 4.7). The general pattern of areas with higher dependence is similar to that using the daily maximum surge. However, dependence becomes significant in a few places where it previously was not, for example south of the Humber estuary and north of the Thames estuary.

4.3.2 Dependence between precipitation & surge

The dependence between precipitation and daily maximum surge was studied to assist in the interpretation of the causes of dependence in the flow-surge analysis. Dependence between precipitation and surge is much stronger on the south and west coasts than on the east (Figures 4.12 and 4.13). (These results are shown in two separate figures for clarity.)

On the east coast the strongest dependence occurs in the north, supporting the location of the strongest flow-surge dependence. On the south and west coasts dependence is widespread, and notably strong also for the eastern part of the south coast where flow-surge dependence is generally not significant. This suggests that dependence breaks down for some other reason, presumably because of the slowly responding chalk catchments in this area, as discussed above.

The dependence measure χ was not estimated for five station-pairs because of too few observations. The inter-station distance for all of these station-pairs is large, so this is of limited practical importance (for further details see Defra / Environment Agency, 2003b).

Table 4.6 Dependence measure, χ, between daily mean river flow and daily maximum surge: 5% significance level and 90% confidence intervals of χ

Flow station	Surge station	χ	5% signif. level	90% conf. interval, lower limit	90% conf. interval, upper limit	Flow station	Surge station	χ	5% signif. Level	90% conf. interval, lower limit	90% conf. Interval, upper limit
2001	Wick	0.03	0.04	-0.01	0.06	19007	Aberdeen	0.02	0.03	-0.01	0.06
2001	Aberdeen	0.02	0.03	-0.01	0.11	19007	North Shields	0.00	0.04	-0.01	0.04
4001	Wick	0.10	0.04	0.03	0.14	20001	Aberdeen	0.00	0.03	-0.01	0.04
4001	Aberdeen	0.11	0.04	0.06	0.17	20001	North Shields	0.00	0.02	-0.01	0.03
7002	Wick	0.03	0.02	0.01	0.08	21009	Aberdeen	0.12	0.03	0.06	0.21
7002	Aberdeen	0.03	0.02	-0.01	0.08	21009	North Shields	0.09	0.03	0.03	0.18
7004	Wick	0.10	0.05	0.04	0.15	22001	Aberdeen	0.00	0.02	-0.01	0.05
7004	Aberdeen	0.09	0.05	0.03	0.19	22001	North Shields	0.00	0.03	-0.01	0.04
8006	Wick	0.09	0.03	0.03	0.13	22006	Aberdeen	-0.01	0.03	-0.01	-0.01
8006	Aberdeen	0.11	0.03	0.04	0.16	22006	North Shields	-0.01	0.02	-0.01	-0.01
9002	Wick	0.00	0.02	-0.01	0.02	23001	North Shields	0.04	0.03	0.01	0.12
9002	Aberdeen	0.00	0.02	-0.01	0.03	24009	North Shields	0.04	0.03	-0.01	0.10
10003	Wick	0.01	0.04	-0.01	0.05	24009	Immingham	0.06	0.03	-0.01	0.12
10003	Aberdeen	0.01	0.04	-0.01	0.04	25001	North Shields	0.05	0.03	0.02	0.10
11001	Aberdeen	0.00	0.03	-0.01	0.02	25001	Immingham	0.03	0.02	0.00	0.08
12001	Aberdeen	0.10	0.03	0.05	0.17	26002	Immingham	0.01	0.03	-0.01	0.04
12002	Aberdeen	0.08	0.03	0.04	0.17	27002	Immingham	0.08	0.02	0.02	0.13
13007	Aberdeen	0.04	0.03	-0.01	0.12	27003	Immingham	0.05	0.03	0.01	0.07
13007	North Shields	0.03	0.02	-0.01	0.09	27021	Immingham	0.01	0.03	-0.01	0.03
14001	Aberdeen	0.12	0.04	0.05	0.20	28009	Immingham	0.00	0.04	-0.01	0.01
14001	North Shields	0.09	0.03	0.00	0.20	28022	Immingham	-0.01	0.03	-0.01	-0.01
14002	Aberdeen	0.06	0.04	-0.01	0.16	29001	Immingham	-0.01	0.03	-0.01	0.01
14002	North Shields	0.06	0.03	-0.01	0.16	29002	Immingham	0.00	0.04	-0.01	0.02
15006	Aberdeen	0.18	0.04	0.12	0.24	29002	Lowestoft	0.02	0.03	0.00	0.05
15006	North Shields	0.15	0.03	0.08	0.23	31002	Immingham	0.00	0.02	-0.01	0.02
15013	Aberdeen	0.14	0.03	0.06	0.23	31002	Lowestoft	0.01	0.03	-0.01	0.04
15013	North Shields	0.09	0.03	0.02	0.17	32001	Immingham	-0.01	0.03	-0.01	0.01
16004	Aberdeen	0.28	0.04	0.15	0.36	32001	Lowestoft	-0.01	0.03	-0.01	0.01
16004	North Shields	0.16	0.04	0.07	0.28	33006	Immingham	-0.01	0.03	-0.01	0.00
17002	Aberdeen	0.13	0.05	0.05	0.20	33006	Lowestoft	-0.01	0.03	-0.01	0.02
17002	North Shields	0.07	0.04	0.00	0.16	33007	Immingham	0.00	0.03	-0.01	0.01
18002	Aberdeen	0.20	0.04	0.11	0.26	33007	Lowestoft	0.01	0.03	-0.01	0.03
18002	North Shields	0.11	0.03	0.04	0.20	33024	Immingham	-0.01	0.03	-0.01	-0.01
18003	Aberdeen	0.20	0.04	0.11	0.28	33024	Lowestoft	0.01	0.03	-0.01	0.04
18003	North Shields	0.13	0.03	0.05	0.24	33039	Immingham	-0.01	0.03	-0.01	-0.01
18011	Aberdeen	0.22	0.06	0.09	0.30	33039	Lowestoft	-0.01	0.03	-0.01	-0.01
18011	North Shields	0.16	0.06	0.05	0.28	34003	Immingham	0.01	0.03	-0.01	0.03
19001	Aberdeen	0.06	0.03	0.02	0.10	34003	Lowestoft	0.01	0.03	-0.01	0.05
19001	North Shields	0.02	0.03	-0.01	0.05	34006	Immingham	0.00	0.03	-0.01	0.01
19006	Aberdeen	0.05	0.03	0.00	0.09	34006	Lowestoft	0.01	0.03	0.00	0.04
19006	North Shields	0.00	0.03	-0.01	0.04	34013	Immingham	0.01	0.08	-0.01	0.03

Table 4.6 Dependence measure, χ, between daily mean river flow and daily maximum surge: 5% significance level and 90% confidence intervals of χ (continued)

Flow station	Surge station	χ	5% signif. level	90% conf. inter-val, lower limit	90% conf. inter-val, upper limit	Flow station	Surge station	χ	5% signif. Level	90% conf. inter-val, lower limit	90% conf. Inter-val, upper limit
34013	Lowestoft	-0.01	0.05	-0.01	-0.01	44001	Newlyn	0.07	0.07	0.02	0.13
34019	Immingham	-0.01	0.03	-0.01	0.01	45001	Weymouth	0.13	0.06	-0.01	0.26
34019	Lowestoft	0.00	0.04	-0.01	0.03	45001	Newlyn	0.07	0.05	0.03	0.12
35004	Lowestoft	0.03	0.03	0.00	0.08	45005	Weymouth	0.24	0.05	0.07	0.38
35004	Sheerness	0.06	0.04	0.01	0.10	45005	Newlyn	0.19	0.04	0.12	0.25
35013	Lowestoft	0.01	0.03	-0.01	0.06	46002	Weymouth	0.25	0.07	0.14	0.39
35013	Sheerness	0.05	0.04	0.00	0.09	46002	Newlyn	0.23	0.05	0.17	0.28
36006	Lowestoft	0.00	0.03	-0.01	0.01	46003	Weymouth	0.16	0.06	-0.01	0.30
36006	Sheerness	0.00	0.04	-0.01	0.02	46003	Newlyn	0.14	0.05	0.09	0.19
37001	Sheerness	0.01	0.04	0.00	0.04	47001	Weymouth	0.20	0.05	0.07	0.32
37005	Lowestoft	0.02	0.03	0.00	0.04	47001	Newlyn	0.11	0.04	0.05	0.17
37005	Sheerness	0.03	0.04	0.00	0.07	47004	Weymouth	0.24	0.06	0.08	0.42
37009	Lowestoft	0.01	0.03	-0.01	0.04	47004	Newlyn	0.16	0.05	0.09	0.22
37009	Sheerness	0.02	0.03	0.00	0.06	47007	Weymouth	0.18	0.06	0.05	0.35
37010	Lowestoft	0.02	0.03	0.00	0.04	47007	Newlyn	0.08	0.04	0.05	0.14
37010	Sheerness	0.04	0.04	0.00	0.07	48007	Weymouth	0.20	0.08	0.07	0.28
39001	Sheerness	0.00	0.04	-0.01	0.02	48007	Newlyn	0.10	0.06	0.06	0.15
40011	Sheerness	0.00	0.03	-0.01	0.03	48011	Weymouth	0.25	0.05	0.10	0.35
40011	Dover	0.01	0.04	0.00	0.05	48011	Newlyn	0.17	0.05	0.09	0.24
40012	Sheerness	0.00	0.04	-0.01	0.02	49001	Newlyn	0.16	0.04	0.09	0.23
40021	Dover	-0.01	0.04	-0.01	0.03	49001	Ilfracombe	0.06	0.04	0.01	0.13
40021	Newhaven	0.15	0.07	0.05	0.29	49002	Newlyn	0.09	0.05	0.06	0.15
41004	Dover	0.06	0.05	0.00	0.10	49002	Ilfracombe	0.03	0.05	-0.01	0.08
41004	Newhaven	-0.01	0.05	-0.01	0.15	50001	Newlyn	0.07	0.05	0.04	0.11
41017	Dover	0.03	0.03	0.00	0.06	50001	Ilfracombe	0.05	0.03	0.01	0.12
41017	Newhaven	0.13	0.05	0.05	0.23	50002	Newlyn	0.06	0.05	0.01	0.10
41023	Newhaven	0.01	0.04	-0.01	0.05	50002	Ilfracombe	0.03	0.03	-0.01	0.09
41023	Portsmouth	0.04	0.06	0.01	0.10	51003	Ilfracombe	0.09	0.04	0.02	0.13
42003	Portsmouth	0.23	0.07	0.08	0.34	51003	Avonmouth	0.08	0.06	0.02	0.15
42003	Weymouth	0.14	0.07	-0.01	0.25	52009	Ilfracombe	0.04	0.03	-0.01	0.08
42003	Newlyn	0.06	0.03	0.04	0.13	52009	Avonmouth	0.01	0.04	-0.01	0.13
42004	Portsmouth	0.06	0.07	-0.01	0.15	53018	Avonmouth	0.11	0.03	0.04	0.18
42004	Weymouth	0.07	0.07	-0.01	0.15	54001	Avonmouth	0.05	0.04	0.02	0.11
42004	Newlyn	0.03	0.05	0.00	0.06	54032	Avonmouth	0.06	0.04	0.01	0.11
42006	Portsmouth	0.06	0.06	0.03	0.10	55023	Avonmouth	0.10	0.03	0.04	0.18
42006	Newlyn	0.03	0.06	0.00	0.05	56001	Avonmouth	0.21	0.04	0.13	0.30
43021	Portsmouth	0.03	0.06	-0.01	0.11	56002	Avonmouth	0.22	0.06	0.10	0.30
43021	Weymouth	0.07	0.08	-0.01	0.12	57005	Avonmouth	0.14	0.05	0.05	0.20
43021	Newlyn	0.01	0.07	-0.01	0.05	58001	Avonmouth	0.05	0.04	-0.01	0.11
44001	Portsmouth	0.13	0.10	0.03	0.24	58001	Milford Haven	0.06	0.04	0.00	0.09
44001	Weymouth	0.12	0.11	0.01	0.20	59001	Avonmouth	0.08	0.04	0.01	0.14

Table 4.6 Dependence measure, χ, between daily mean river flow and daily maximum surge: 5% significance level and 90% confidence intervals of χ (continued)

Flow station	Surge station	χ	5% signif. level	90% conf. inter-val, lower limit	90% conf. inter-val, upper limit
59001	Milford Haven	0.13	0.03	0.06	0.18
60003	Avonmouth	0.11	0.04	0.02	0.21
60003	Milford Haven	0.13	0.04	0.06	0.20
60010	Avonmouth	0.13	0.04	0.05	0.20
60010	Milford Haven	0.15	0.03	0.06	0.19
61002	Milford Haven	0.14	0.03	0.06	0.19
62001	Fishguard	0.14	0.03	0.08	0.20
62001	Holyhead	0.09	0.04	0.03	0.18
63001	Fishguard	0.04	0.03	0.01	0.07
63001	Holyhead	0.03	0.02	-0.01	0.07
64006	Fishguard	0.02	0.02	-0.01	0.06
64006	Holyhead	-0.01	0.02	-0.01	0.04
65001	Fishguard	0.04	0.03	0.01	0.08
65001	Holyhead	0.06	0.02	0.02	0.13
66001	Holyhead	0.10	0.05	0.04	0.17
66001	Liverpool P P	0.05	0.04	0.03	0.10
66001	Liverpool G D	0.07	0.06	0.02	0.14
67015	Holyhead	0.13	0.05	0.05	0.22
67015	Liverpool P P	0.08	0.04	0.04	0.14
67015	Liverpool G D	0.15	0.09	0.08	0.24
68020	Holyhead	-0.01	0.04	-0.01	0.03
68020	Liverpool G D	-0.01	0.08	-0.01	0.06
69002	Liverpool P P	0.10	0.04	0.06	0.17
69002	Liverpool G D	0.09	0.06	0.04	0.24
69007	Liverpool P P	-0.01	0.06	-0.01	-0.01
69007	Liverpool G D	0.06	0.07	-0.01	0.14
70004	Liverpool P P	0.06	0.04	-0.01	0.16
70004	Liverpool G D	0.04	0.07	-0.01	0.22
70004	Heysham	0.03	0.04	-0.01	0.07
71001	Liverpool P P	0.13	0.03	0.08	0.19
71001	Liverpool G D	0.10	0.04	-0.01	0.17
71001	Heysham	0.09	0.03	0.04	0.14
72004	Liverpool P P	0.16	0.04	0.08	0.27
72004	Liverpool G D	0.12	0.04	-0.01	0.23
72004	Heysham	0.11	0.03	0.04	0.16
72008	Liverpool P P	0.09	0.04	0.01	0.18
72008	Liverpool G D	-0.01	0.07	-0.01	0.08
72008	Heysham	0.03	0.02	0.00	0.07
73002	Heysham	0.05	0.03	0.03	0.10
73002	Portpatrick	0.05	0.04	0.03	0.10
73005	Heysham	0.11	0.03	0.05	0.21
73005	Portpatrick	0.08	0.04	0.04	0.14
74001	Heysham	0.09	0.02	0.03	0.13
74001	Portpatrick	0.08	0.03	0.04	0.14
74006	Heysham	0.03	0.02	0.00	0.07
74006	Portpatrick	0.03	0.02	-0.01	0.07
75002	Heysham	0.07	0.04	0.04	0.11
75002	Portpatrick	0.13	0.03	0.06	0.15
76007	Heysham	0.15	0.05	0.09	0.20
76007	Portpatrick	0.09	0.04	0.05	0.14
77001	Heysham	0.12	0.04	0.06	0.18
77001	Portpatrick	0.13	0.03	0.08	0.20
78003	Heysham	0.11	0.03	0.06	0.20
78003	Portpatrick	0.14	0.03	0.10	0.20
79002	Heysham	0.19	0.02	0.10	0.27
79002	Portpatrick	0.20	0.03	0.12	0.28
79005	Heysham	0.14	0.03	0.10	0.23
79005	Portpatrick	0.15	0.03	0.08	0.22
81002	Heysham	0.07	0.02	0.03	0.10
81002	Portpatrick	0.11	0.03	0.06	0.17
82001	Portpatrick	0.13	0.02	0.08	0.20
82001	Millport	0.10	0.03	0.05	0.18
83005	Portpatrick	0.02	0.04	-0.01	0.07
83005	Millport	0.01	0.05	-0.01	0.06
84001	Millport	0.04	0.04	-0.01	0.10
84013	Millport	0.10	0.05	0.01	0.14
85001	Millport	0.05	0.04	0.02	0.11
86001	Millport	0.03	0.03	-0.01	0.07
93001	Tobermory	0.02	0.07	-0.01	0.06
93001	Ullapool	0.07	0.07	0.01	0.12
94001	Tobermory	0.05	0.09	-0.01	0.11
94001	Ullapool	0.10	0.06	0.04	0.16
95001	Ullapool	0.05	0.06	0.00	0.14
95001	Wick	0.04	0.05	-0.01	0.09
95002	Ullapool	0.09	0.06	0.01	0.19
96001	Ullapool	0.02	0.04	-0.01	0.07
96001	Wick	0.03	0.03	-0.01	0.06
96001	Lerwick	0.03	0.04	-0.01	0.07
97002	Ullapool	0.01	0.03	-0.01	0.06
97002	Wick	0.00	0.03	-0.01	0.04
97002	Lerwick	0.05	0.04	-0.01	0.09

Table 4.7 Dependence measure, χ, between daily mean river flow and daily maximum surge occurring at high tide: 5% significance level and 90% confidence intervals of χ

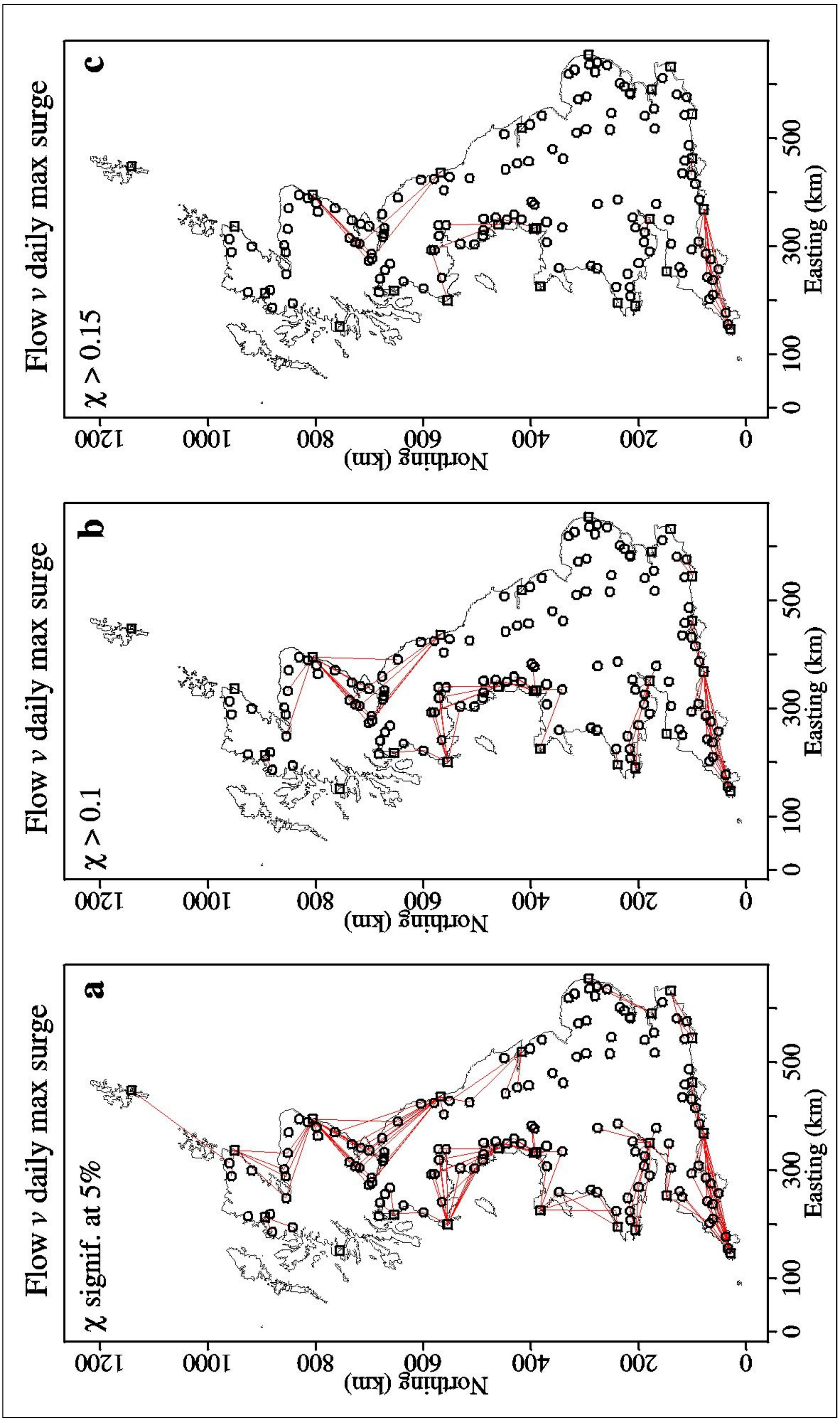
Flow station	Surge station	χ	5% signif. level	90% conf. interval, lower limit	90% conf. interval, upper limit	Flow station	Surge station	χ	5% signif. Level	90% conf. interval, lower limit	90% conf. interval, upper limit
2001	Wick	0.05	0.04	0.01	0.11	19007	Aberdeen	0.04	0.03	0.00	0.09
2001	Aberdeen	0.04	0.03	-0.01	0.09	19007	North Shields	0.03	0.03	0.00	0.08
4001	Wick	0.12	0.03	0.05	0.19	20001	Aberdeen	0.02	0.02	-0.01	0.04
4001	Aberdeen	0.10	0.04	0.04	0.16	20001	North Shields	-0.01	0.02	-0.01	-0.01
7002	Wick	0.04	0.02	0.01	0.10	21009	Aberdeen	0.11	0.03	0.07	0.19
7002	Aberdeen	0.06	0.03	0.02	0.10	21009	North Shields	0.08	0.03	0.03	0.13
7004	Wick	0.13	0.04	0.07	0.18	22001	Aberdeen	0.01	0.04	-0.01	0.04
7004	Aberdeen	0.10	0.04	0.03	0.15	22001	North Shields	-0.01	0.03	-0.01	0.01
8006	Wick	0.07	0.03	0.02	0.11	22006	Aberdeen	-0.01	0.04	-0.01	-0.01
8006	Aberdeen	0.08	0.03	0.04	0.13	22006	North Shields	-0.01	0.03	-0.01	-0.01
9002	Wick	0.00	0.02	-0.01	0.02	23001	North Shields	0.06	0.03	0.03	0.10
9002	Aberdeen	0.01	0.02	-0.01	0.03	24009	North Shields	0.04	0.03	-0.01	0.08
10003	Wick	0.01	0.06	-0.01	0.05	24009	Immingham	0.04	0.03	0.01	0.08
10003	Aberdeen	0.04	0.04	-0.01	0.07	25001	North Shields	0.04	0.03	0.02	0.09
11001	Aberdeen	0.03	0.03	-0.01	0.08	25001	Immingham	0.03	0.02	0.00	0.07
12001	Aberdeen	0.10	0.03	0.06	0.17	26002	Immingham	0.01	0.03	-0.01	0.03
12002	Aberdeen	0.07	0.03	0.04	0.13	27002	Immingham	0.06	0.02	0.03	0.11
13007	Aberdeen	0.02	0.02	-0.01	0.09	27003	Immingham	0.06	0.03	0.02	0.10
13007	North Shields	0.02	0.02	-0.01	0.08	27021	Immingham	0.03	0.03	0.00	0.05
14001	Aberdeen	0.06	0.04	0.02	0.14	28009	Immingham	0.00	0.03	-0.01	0.02
14001	North Shields	0.05	0.03	-0.01	0.14	28022	Immingham	0.00	0.03	-0.01	0.02
14002	Aberdeen	0.05	0.04	0.00	0.10	29001	Immingham	0.02	0.02	-0.01	0.04
14002	North Shields	0.05	0.03	-0.01	0.12	29002	Immingham	0.04	0.03	0.01	0.06
15006	Aberdeen	0.17	0.03	0.12	0.22	29002	Lowestoft	0.03	0.03	0.00	0.06
15006	North Shields	0.10	0.03	0.07	0.17	31002	Immingham	0.01	0.02	-0.01	0.04
15013	Aberdeen	0.11	0.03	0.06	0.19	31002	Lowestoft	0.01	0.02	-0.01	0.04
15013	North Shields	0.08	0.03	0.03	0.13	32001	Immingham	0.01	0.03	0.00	0.03
16004	Aberdeen	0.23	0.04	0.15	0.30	32001	Lowestoft	0.00	0.03	-0.01	0.01
16004	North Shields	0.16	0.04	0.07	0.24	33006	Immingham	0.00	0.03	-0.01	0.02
17002	Aberdeen	0.11	0.04	0.04	0.17	33006	Lowestoft	-0.01	0.03	-0.01	0.01
17002	North Shields	0.07	0.05	0.02	0.13	33007	Immingham	0.00	0.03	-0.01	0.02
18002	Aberdeen	0.16	0.04	0.10	0.23	33007	Lowestoft	0.00	0.03	-0.01	0.03
18002	North Shields	0.13	0.04	0.06	0.22	33024	Immingham	0.01	0.03	-0.01	0.04
18003	Aberdeen	0.19	0.04	0.15	0.28	33024	Lowestoft	0.01	0.03	0.00	0.04
18003	North Shields	0.13	0.04	0.07	0.20	33039	Immingham	-0.01	0.03	-0.01	0.01
18011	Aberdeen	0.18	0.04	0.10	0.24	33039	Lowestoft	-0.01	0.03	-0.01	-0.01
18011	North Shields	0.12	0.05	0.06	0.21	34003	Immingham	0.00	0.03	-0.01	0.03
19001	Aberdeen	0.03	0.03	0.00	0.11	34003	Lowestoft	0.02	0.02	-0.01	0.05
19001	North Shields	0.01	0.03	-0.01	0.04	34006	Immingham	0.00	0.04	-0.01	0.03
19006	Aberdeen	0.03	0.03	-0.01	0.07	34006	Lowestoft	0.01	0.03	0.00	0.04
19006	North Shields	0.00	0.03	-0.01	0.03	34013	Immingham	0.01	0.05	-0.01	0.03

Table 4.7 Dependence measure, χ, between daily mean river flow and daily maximum surge occurring at high tide: 5% significance level and 90% confidence intervals of χ (continued)

Flow station	Surge station	χ	5% signif. level	90% conf. inter-val, lower limit	90% conf. inter-val, upper limit
34013	Lowestoft	-0.01	0.05	-0.01	-0.01
34019	Immingham	-0.01	0.04	-0.01	0.00
34019	Lowestoft	0.00	0.03	-0.01	0.03
35004	Lowestoft	0.05	0.03	0.00	0.09
35004	Sheerness	0.08	0.03	0.03	0.11
35013	Lowestoft	0.04	0.03	0.00	0.08
35013	Sheerness	0.10	0.04	0.03	0.15
36006	Lowestoft	0.00	0.03	-0.01	0.01
36006	Sheerness	0.01	0.03	-0.01	0.05
37001	Sheerness	0.03	0.03	0.00	0.07
37005	Lowestoft	0.02	0.03	-0.01	0.04
37005	Sheerness	0.06	0.03	0.01	0.11
37009	Lowestoft	0.01	0.03	-0.01	0.03
37009	Sheerness	0.04	0.03	0.00	0.08
37010	Lowestoft	0.02	0.02	0.00	0.04
37010	Sheerness	0.05	0.04	0.01	0.08
39001	Sheerness	0.00	0.04	-0.01	0.01
40011	Sheerness	0.00	0.03	-0.01	0.02
40011	Dover	0.00	0.03	-0.01	0.02
40012	Sheerness	0.01	0.03	-0.01	0.04
40021	Dover	-0.01	0.04	-0.01	0.04
40021	Newhaven	0.02	0.08	-0.01	0.11
41004	Dover	0.01	0.04	-0.01	0.05
41004	Newhaven	0.04	0.05	-0.01	0.12
41017	Dover	0.01	0.04	-0.01	0.05
41017	Newhaven	0.04	0.05	-0.01	0.07
41023	Newhaven	0.02	0.06	-0.01	0.05
41023	Portsmouth	0.04	0.07	0.00	0.08
42003	Portsmouth	0.11	0.07	0.05	0.27
42003	Weymouth	0.17	0.06	0.06	0.34
42003	Newlyn	0.06	0.04	0.02	0.15
42004	Portsmouth	0.04	0.08	-0.01	0.18
42004	Weymouth	0.06	0.03	-0.01	0.13
42004	Newlyn	0.05	0.04	0.02	0.07
42006	Portsmouth	0.06	0.07	0.00	0.13
42006	Newlyn	0.05	0.05	0.03	0.07
43021	Portsmouth	0.06	0.08	0.00	0.14
43021	Weymouth	0.04	0.06	0.00	0.08
43021	Newlyn	0.05	0.06	0.02	0.08
44001	Portsmouth	0.11	0.10	0.01	0.27
44001	Weymouth	0.13	0.07	0.01	0.24
44001	Newlyn	0.09	0.06	0.04	0.15
45001	Weymouth	0.12	0.06	0.02	0.21
45001	Newlyn	0.08	0.05	0.03	0.14
45005	Weymouth	0.22	0.05	0.03	0.35
45005	Newlyn	0.18	0.03	0.11	0.24
46002	Weymouth	0.18	0.07	0.07	0.31
46002	Newlyn	0.20	0.05	0.13	0.25
46003	Weymouth	0.10	0.07	-0.01	0.21
46003	Newlyn	0.12	0.05	0.08	0.19
47001	Weymouth	0.18	0.05	0.05	0.28
47001	Newlyn	0.14	0.04	0.07	0.21
47004	Weymouth	0.25	0.09	0.10	0.39
47004	Newlyn	0.17	0.05	0.09	0.23
47007	Weymouth	0.17	0.09	0.03	0.40
47007	Newlyn	0.13	0.05	0.08	0.20
48007	Weymouth	0.19	0.09	0.09	0.30
48007	Newlyn	0.13	0.05	0.08	0.18
48011	Weymouth	0.26	0.06	0.11	0.38
48011	Newlyn	0.16	0.05	0.08	0.23
49001	Newlyn	0.16	0.03	0.10	0.23
49001	Ilfracombe	0.08	0.03	0.03	0.15
49002	Newlyn	0.11	0.05	0.07	0.17
49002	Ilfracombe	0.08	0.04	0.04	0.13
50001	Newlyn	0.07	0.04	0.04	0.12
50001	Ilfracombe	0.03	0.03	-0.01	0.12
50002	Newlyn	0.07	0.04	0.03	0.11
50002	Ilfracombe	0.05	0.04	0.01	0.16
51003	Ilfracombe	0.05	0.04	0.01	0.15
51003	Avonmouth	0.08	0.04	0.02	0.11
52009	Ilfracombe	0.02	0.02	-0.01	0.08
52009	Avonmouth	0.02	0.04	-0.01	0.09
53018	Avonmouth	0.10	0.03	0.01	0.12
54001	Avonmouth	0.08	0.04	0.02	0.11
54032	Avonmouth	0.06	0.05	0.01	0.10
55023	Avonmouth	0.06	0.04	0.01	0.10
56001	Avonmouth	0.14	0.04	0.05	0.23
56002	Avonmouth	0.13	0.08	0.04	0.20
57005	Avonmouth	0.10	0.05	0.04	0.24
58001	Avonmouth	0.05	0.04	-0.01	0.13
58001	Milford Haven	0.06	0.03	0.00	0.09
59001	Avonmouth	0.08	0.04	0.02	0.21

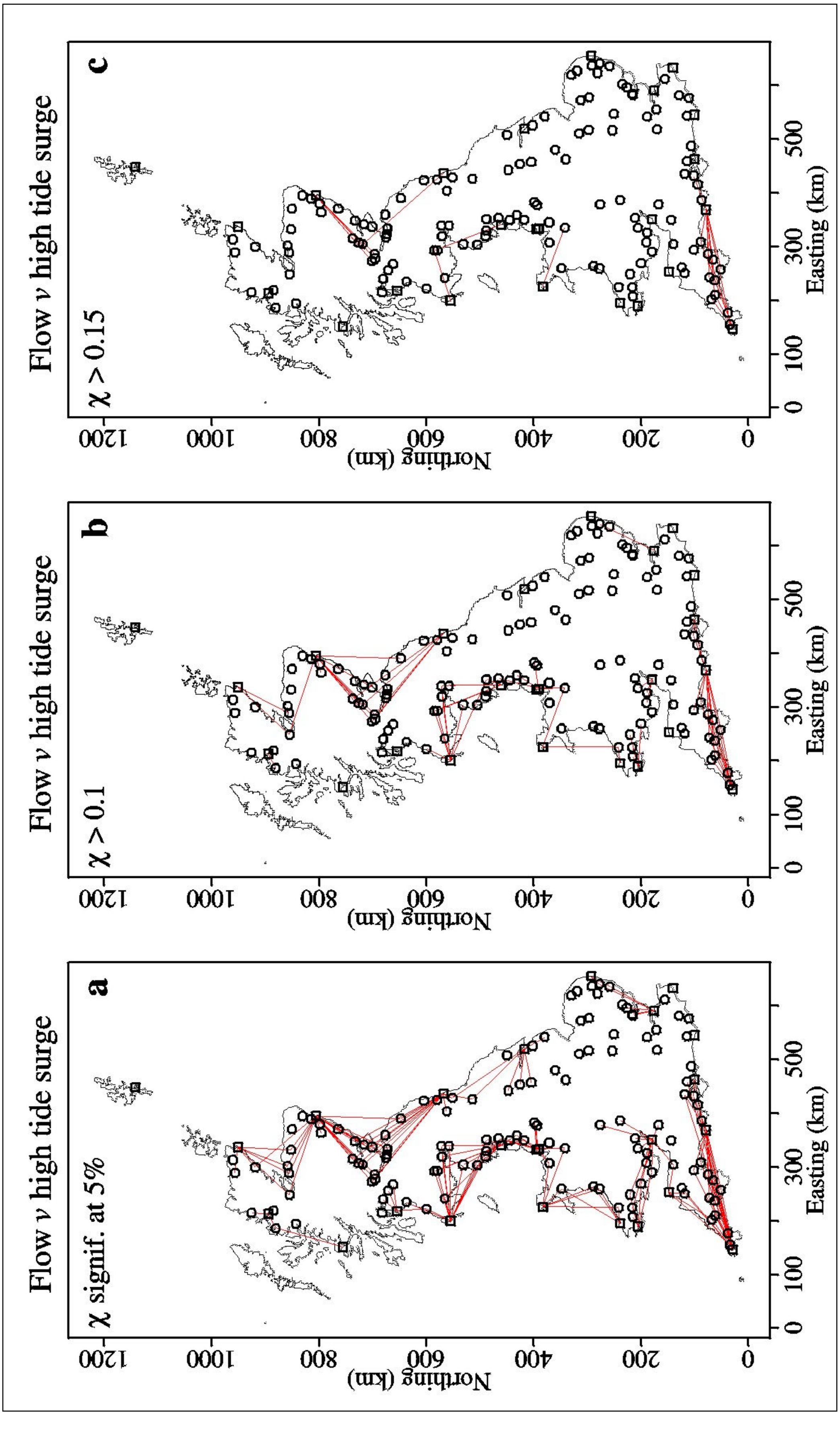
Table 4.7 Dependence measure, χ, between daily mean river flow and daily maximum surge occurring at high tide: 5% significance level and 90% confidence intervals of χ (continued)

Flow station	Surge station	χ	5% signif. level	90% conf. inter-val, lower limit	90% conf. inter-val, upper limit	Flow station	Surge station	χ	5% signif. Level	90% conf. inter-val, lower limit	90% conf. inter-val, upper limit
59001	Milford Haven	0.13	0.03	0.06	0.17	73005	Portpatrick	0.11	0.05	0.07	0.17
60003	Avonmouth	0.07	0.06	0.01	0.14	74001	Heysham	0.09	0.02	0.04	0.14
60003	Milford Haven	0.12	0.04	0.02	0.22	74001	Portpatrick	0.09	0.03	0.04	0.14
60010	Avonmouth	0.09	0.06	0.01	0.16	74006	Heysham	0.04	0.02	0.00	0.08
60010	Milford Haven	0.14	0.03	0.06	0.18	74006	Portpatrick	0.05	0.02	0.01	0.10
61002	Milford Haven	0.12	0.04	0.05	0.18	75002	Heysham	0.08	0.04	0.03	0.11
62001	Fishguard	0.13	0.04	0.07	0.19	75002	Portpatrick	0.12	0.03	0.06	0.15
62001	Holyhead	0.14	0.04	0.07	0.19	76007	Heysham	0.14	0.04	0.06	0.21
63001	Fishguard	0.04	0.03	0.01	0.09	76007	Portpatrick	0.08	0.04	0.05	0.13
63001	Holyhead	0.03	0.02	-0.01	0.06	77001	Heysham	0.11	0.03	0.06	0.17
64006	Fishguard	0.03	0.02	0.00	0.08	77001	Portpatrick	0.11	0.04	0.07	0.21
64006	Holyhead	0.02	0.01	-0.01	0.06	78003	Heysham	0.10	0.03	0.05	0.16
65001	Fishguard	0.06	0.03	0.03	0.11	78003	Portpatrick	0.13	0.03	0.08	0.19
65001	Holyhead	0.09	0.04	0.04	0.14	79002	Heysham	0.17	0.02	0.11	0.21
66001	Holyhead	0.08	0.03	0.01	0.16	79002	Portpatrick	0.19	0.03	0.10	0.25
66001	Liverpool P P	0.04	0.04	0.01	0.09	79005	Heysham	0.11	0.02	0.06	0.19
66001	Liverpool G D	0.09	0.06	-0.01	0.17	79005	Portpatrick	0.14	0.03	0.09	0.21
67015	Holyhead	0.16	0.04	0.08	0.24	81002	Heysham	0.05	0.03	0.02	0.10
67015	Liverpool P P	0.11	0.04	0.05	0.20	81002	Portpatrick	0.10	0.03	0.03	0.13
67015	Liverpool G D	0.15	0.09	0.03	0.26	82001	Portpatrick	0.12	0.03	0.06	0.18
68020	Holyhead	-0.01	0.03	-0.01	0.03	82001	Millport	0.09	0.03	0.01	0.15
68020	Liverpool G D	-0.01	0.05	-0.01	-0.01	83005	Portpatrick	0.05	0.03	0.01	0.09
69002	Liverpool P P	0.11	0.04	0.03	0.15	83005	Millport	0.01	0.03	-0.01	0.06
69002	Liverpool G D	0.09	0.07	-0.01	0.23	84001	Millport	0.06	0.05	0.01	0.10
69007	Liverpool P P	0.03	0.04	-0.01	0.08	84013	Millport	0.07	0.03	0.01	0.11
69007	Liverpool G D	0.10	0.04	-0.01	0.19	85001	Millport	0.06	0.04	0.03	0.11
70004	Liverpool P P	0.03	0.06	-0.01	0.20	86001	Millport	0.03	0.03	-0.01	0.10
70004	Liverpool G D	0.04	0.04	-0.01	0.19	93001	Tobermory	0.05	0.07	-0.01	0.09
70004	Heysham	0.04	0.03	0.01	0.08	93001	Ullapool	0.04	0.05	-0.01	0.09
71001	Liverpool P P	0.11	0.03	0.05	0.19	94001	Tobermory	0.09	0.09	0.01	0.18
71001	Liverpool G D	0.07	0.04	-0.01	0.20	94001	Ullapool	0.12	0.04	0.04	0.19
71001	Heysham	0.07	0.03	0.03	0.13	95001	Ullapool	0.05	0.05	-0.01	0.12
72004	Liverpool P P	0.13	0.03	0.03	0.21	95001	Wick	0.04	0.04	0.00	0.08
72004	Liverpool G D	0.07	0.04	-0.01	0.21	95002	Ullapool	0.07	0.07	0.02	0.15
72004	Heysham	0.08	0.04	0.03	0.16	96001	Ullapool	0.03	0.03	-0.01	0.06
72008	Liverpool P P	0.09	0.02	0.02	0.17	96001	Wick	0.05	0.03	-0.01	0.10
72008	Liverpool G D	-0.01	0.05	-0.01	0.05	96001	Lerwick	0.02	0.04	-0.01	0.07
72008	Heysham	0.05	0.03	0.03	0.11	97002	Ullapool	0.01	0.03	-0.01	0.05
73002	Heysham	0.07	0.03	0.04	0.10	97002	Wick	0.03	0.04	-0.01	0.07
73002	Portpatrick	0.05	0.04	0.03	0.10	97002	Lerwick	0.03	0.04	-0.01	0.07
73005	Heysham	0.10	0.04	0.05	0.20						



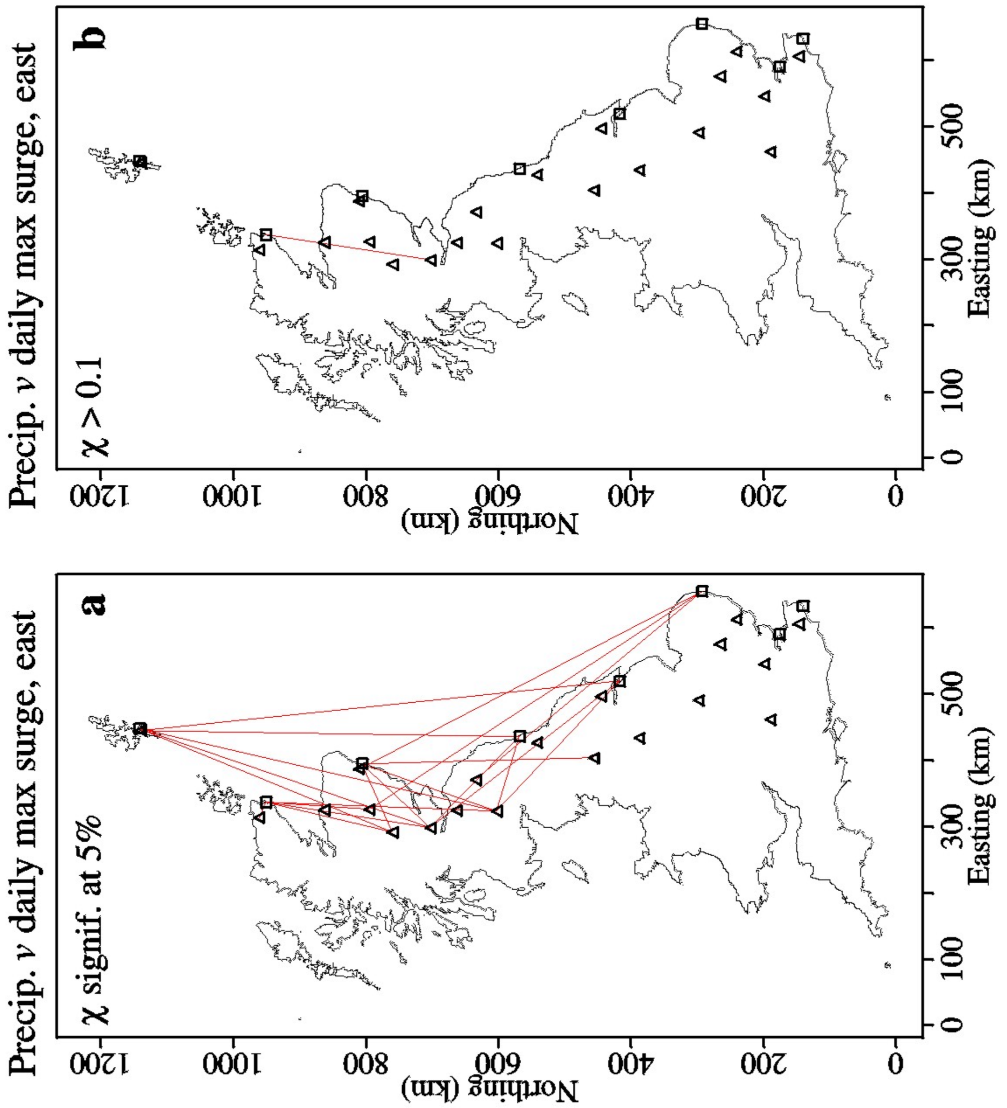
Lines connect station-pairs with χ exceeding a) the 95% point (significant dependence), b) 0.1, and c) 0.15. Dependence is only estimated for neighbouring stations, i.e. generally between each river flow station and one surge station on either side of the estuary (see text for further details).

Figure 4.10 Dependence between river flow and daily maximum surge



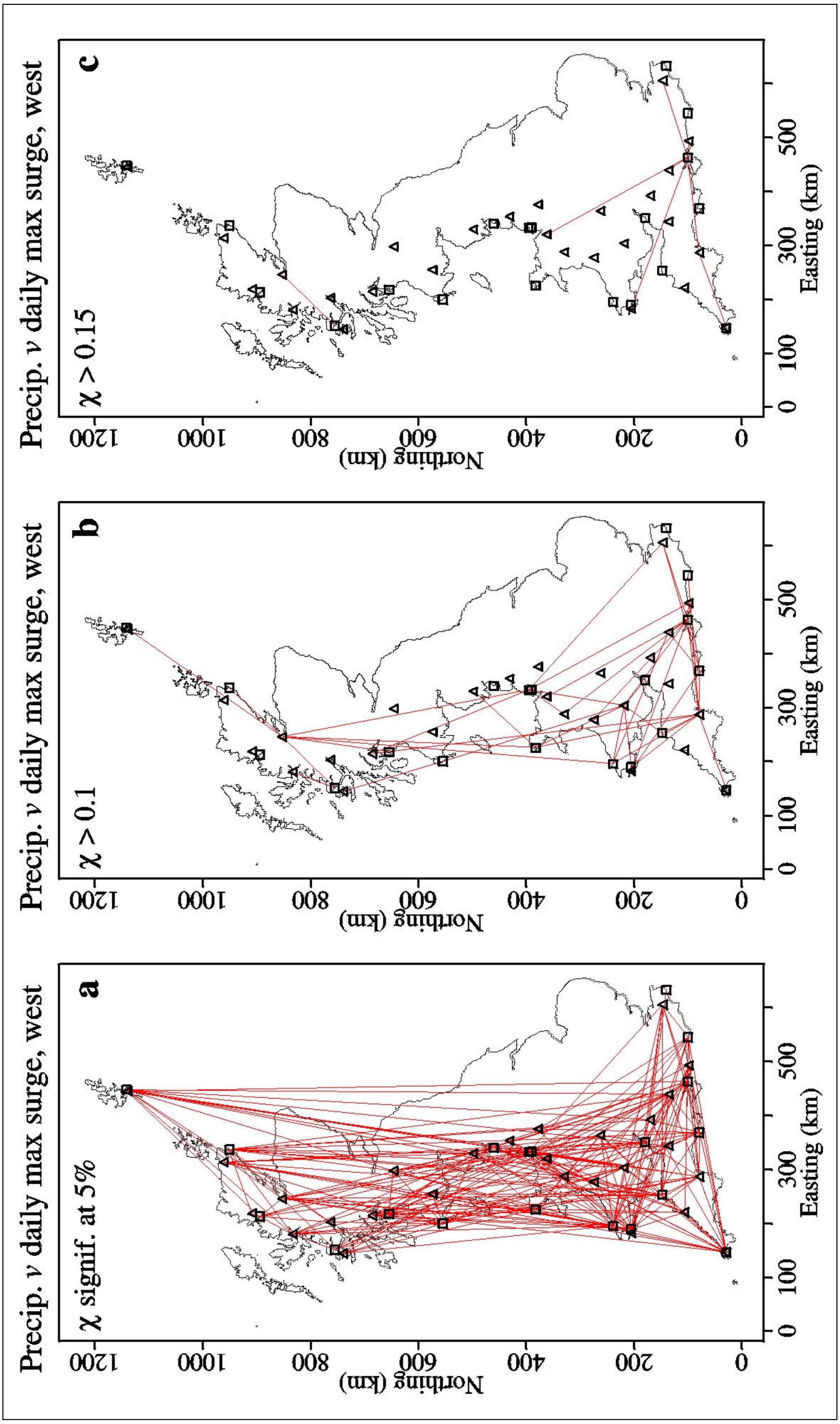
Lines connect station-pairs with χ exceeding a) the 95% point (significant dependence), b) 0.1, and c) 0.15. Dependence is only estimated for neighbouring stations, i.e. generally between each river flow station and one surge station on either side of the estuary (see text for further details).

Figure 4.11 Dependence between river flow and daily maximum surge occurring at high tide



Lines connect station-pairs with χ exceeding a) the 95% point (significant dependence) and b) 0.1. Dependence is estimated for all station-pairs.

Figure 4.12 Dependence between precipitation and daily maximum surge in catchments draining to the British east coast



Lines connect station-pairs with χ exceeding a) the 95% point (significant dependence), b) 0.1, and c) 0.15. Dependence is estimated for all station-pairs.

Figure 4.13 Dependence between precipitation and daily maximum surge in catchments draining to the British south and west coasts

4.3.3 Lagged analysis

The dependence between river flow and daily maximum surge is often strongest when surge and flow occur on the same day, for all coasts. In general, dependence on the east coast is strong also for flows lagged one day before or one day after the surge, and on the south and west coasts for flow lagged one day after the surge (Svensson and Jones, 2000; Defra / Environment Agency, 2003b). The longer temporal overlap for strong dependence on the east coast may be related to the longer duration of a positive surge event there. Surges on the east coast tend to last for about one day, whereas surges on the west coast are often shorter, lasting between about 9 and 15 hours (Heaps, 1967).

Slowly responding catchments may reach their peak dependence for larger lags. For example, the Severn (54001, 54032) reaches its peak when the flow is lagged two days behind the surge at Avonmouth (Appendix D of Svensson and Jones, 2000).

On the south and west coasts the dependence between precipitation and daily maximum surge is strongest on the same day, whereas on the east coast the dependence is strongest when precipitation precedes the surge by one day (Svensson and Jones, 2000; Defra / Environment Agency, 2003b). This allows more time for river flow to arrive in the estuary, and potentially for flow-surge dependence to occur also for more slowly responding catchments. The reason for the different behaviour in timing is probably related to the different processes generating the surge. On the west coast the surge is generally formed locally by the low atmospheric pressure and the southerly to westerly winds associated with a depression approaching Britain from the west, driving the water towards the coast (Lennon, 1963). On the east coast, however, the surge wave is often generated externally, to the north-west of Scotland, as an eastward-moving depression traverses the continental shelf. The surge wave then propagates southward along the British east coast, and may be further amplified by northerly winds behind the depression as this passes across the North Sea (Pugh, 1987). The formation of the surge after the depression and the associated fronts have passed Britain explains the one-day lag between precipitation and surge.

4.3.4 Seasonal analysis

Dependence between river flow and daily maximum surge in the northern part of the east coast is clearly stronger in winter than in summer, which is supported by the precipitation-surge analysis. The results for the south and west coasts are more ambiguous (especially for the precipitation-surge analysis), with both positive and negative differences between the seasons occurring in the same regions. However, the western part of the south coast shows more station-pairs with stronger flow-surge dependence in winter than in summer, whereas the west coasts of Wales and northward, mainly show higher flow-surge dependence in summer than in winter ($|\chi_{\text{diff}}| > 0.05$). The increase /decrease in dependence in different areas may be related to the seasonal variation in preferred storm tracks, and to higher soil moisture deficits in summer (especially in the south and east) disrupting the runoff process (Svensson and Jones, 2000; Defra / Environment Agency, 2003b).

4.3.5 Impact of climate change

It is difficult to assess the impact of climate change on the dependence between river flow and surge without embarking on a separate study involving global climate modelling. However, the flow-surge dependence appears to be largely influenced by the storm track of the depressions. It therefore seems reasonable to investigate the sensitivity of the dependence to shifts in preferred storm tracks. The North Atlantic Oscillation Index (NAOI) is a measure of the (oscillating) pressure difference between the Azores and south-west Iceland (Hurrell, 1995). When the NAOI is in its positive phase, storms tend to track in a north-easterly direction to the north of Scotland. However, when it is in its negative phase, storms tend to move eastwards along a more southerly track, at about 45°N (Rogers, 1990). Most global climate models suggest a shift towards the positive phase of the NAOI in the future (Gillett *et al.*, 2002).

The analysis here was restricted to October to March because the NAOI is most pronounced during the winter. Twelve winters each of high and low NAOI were selected. The differences in χ between high and low NAOI years are relatively modest, with 26 of 159 station-pairs having $|\chi_{diff}|$ exceeding 0.1. The results should be treated with caution as there are many station-pairs for which dependence was not estimated because of too few data observations, particularly on the south and west coasts. However, when looking at the geographical spread of $|\chi_{diff}| > 0.05$ some patterns emerge. On the east coast north of the Firth of Forth, the dependence between river flow and daily maximum surge tends to be higher in positive NAOI winters than in negative NAOI winters, whereas it is lower south of the Firth of Forth down to the Thames estuary. More station-pairs show a strengthening than a weakening of the dependence in high NAOI years on the west coast from Wales up to the Solway Firth area and possibly further north. See Defra / Environment Agency (2003b) for further details.

4.4 Results for tide and surge

The surge amplification factor, $q_s(d)$, where d is the position along the UK coast measured clockwise from Wick, and the tide-surge interaction function, $a(X : d)$, where X is the tidal level, were defined in Section 3.4.3. For convenience of analysis and spatial interpretation, tide and surge were both transformed into a normalised distribution over the range [0 1]. Consequently, if the function $a(X : d)$, for any given position, d, proved to be constant against tidal level then it could be concluded that the tide and surge were independent. Conversely, departure from a constant would indicate tide-surge interaction.

The above parameters were found to change smoothly along the coast and with tidal state. In Dixon and Tawn (1997) this procedure was applied to the full UK coastline resulting in the values shown in Figure 4.14. The top figure shows the interaction parameter $a(X : d)$ for the upper tidal band, the lower figure shows $a(X : d)$ for the mid-tide range, and the three-letter annotations indicate the tide gauge positions. Lack of tide-surge interaction at a site would result in an equal value of $a(X : d)$ from both curves. Conversely, interaction would result in a significant departure between the two.

The difference between the two curves has been used here to indicate the degree of interaction as a function of position on the UK coast. These are shown on the map in

Figure 4.15 in a digitised graphical form. The digitisation has been graded as follows and coloured accordingly on the map:

No Interaction	Green
Low Interaction	Blue
Medium Interaction	Purple
Strong Interaction	Orange

For Low and No Interaction the surge distribution can be considered as independent of the tide. For the Medium Interaction seen in the Irish Sea it is suggested that the tide would need to be separated into 5 equi-probable bands for use in extremes analysis and, for Strong Interaction, that 10 tidal bands are advisable. However, if a 10-banded tidal distribution is used for all cases, then no adverse effect will arise as the lower interaction cases will be accommodated correctly.

Positive interaction implies that surge is less likely to occur at high water than would be the case under the assumption of no interaction or independence. The relevant level of dependence is built in to extreme sea level analyses given in Dixon and Tawn (1997), but alternative methods that fail to take account of interaction, where it exists, would tend to over-estimate extreme sea levels.

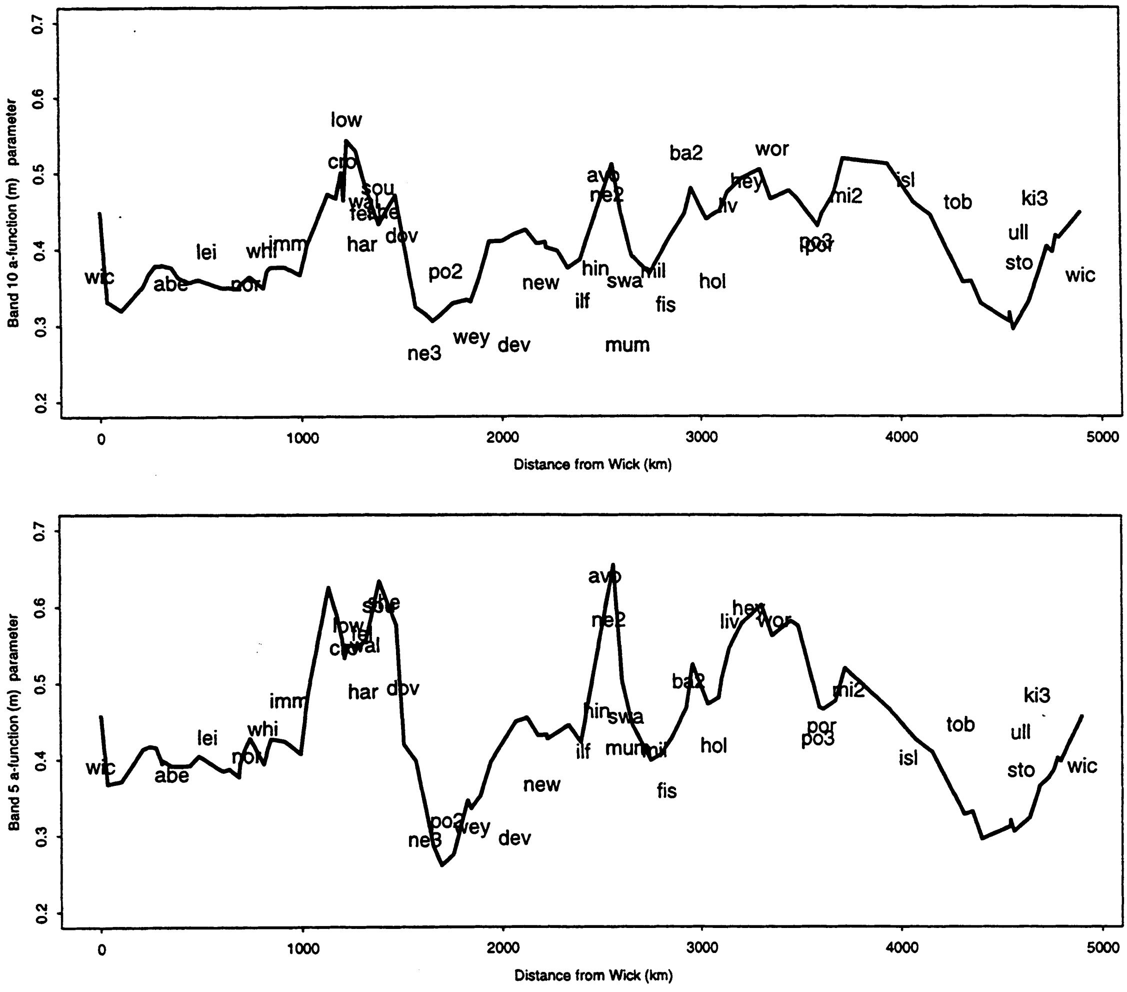


Figure 4.14 The tide-surge interaction parameter for the upper and mid-tide bands[1]

---

[1] The place names indicated in Figure 4.14 are (left to right) Wick, Aberdeen, Leith, North Shields, Whitby, Immingham, Cromer, Lowestoft, Felixstowe, Harwich, Walton, Southend, Sheerness, Dover, Newhaven, Portsmouth, Weymouth, Devonport, Newlyn, Ilfracombe, Hinkley, Avonmouth, Newport, Swansea, Mumbles, Milford Haven, Fishguard, Barmouth, Holyhead, Liverpool, Heysham, Workington, Port Erin, Portpatrick, Millport, Islay, Tobermory, Stornoway, Ullapool, Kinlochbervie and back to Wick.

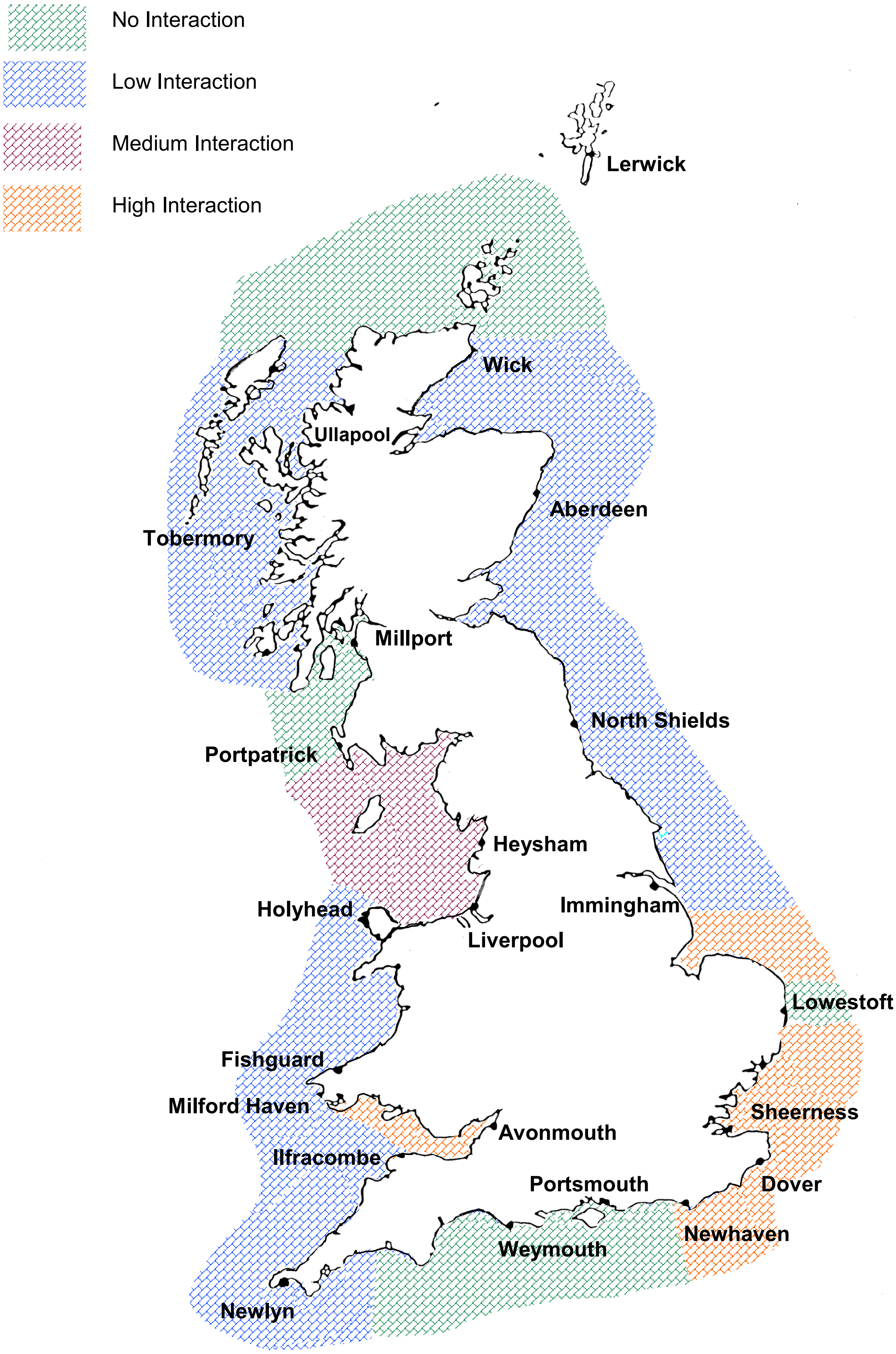


Figure 4.15 Relative levels of tide-surge interaction

5. CONCLUDING REMARKS ON USE OF THE DEPENDENCE RESULTS

5.1 Introduction

The main purpose of the present project was to facilitate understanding and take-up of joint probability approaches to flood risk and design of flood and coastal defences. This is achieved by descriptions of methods and their application, and provision of dependence information for several variable-pairs of interest at many locations around England, Wales and Scotland. The dependence maps are sufficient to give an impression of its spatial variability, and can be used to facilitate discussion of the physical reasons for dependence and its importance in flood risk estimation.

The remainder of this chapter summarises the use of the dependence information in subsequent joint probability analyses. More detailed guidance on procedures, checks and interpretation will be given in the accompanying best practice report (Defra / Environment Agency, 2003a).

5.2 Additional information required from the user

This project addresses dependence and *joint* probability, but not marginal (single variable) distributions and extremes. These need to be provided by the user, from existing reports or design guides (see list in Section 1.4.4) and/or site-specific analysis. The range of marginal return periods required from the user depends on the range of joint exceedence return periods to be considered, but will typically be about 0.1 year to 100 years. (If JOIN-SEA is to be used to synthesise a very large sample of joint probability data, then complete distributions are needed for the marginal variables.)

The user has some discretion over the number of records per year assumed for the variable-pair of interest, typically one per tide for marine records, and one per day or one per hydrograph for fluvial records. It is possible to select a smaller number of events per year, corresponding only to the high valued end of the distribution (although for multi-variable data it may not be obvious what constitutes a high value).

The user has some choice as to how results are presented or interpreted, whether in terms of extremes or joint probability density, and whether in terms of the multiple environmental variables or converted to one or more single structure functions (e.g. river level or overtopping rate).

5.3 Simplified method

Section 3.5.describes a simple method of producing joint exceedence combinations of two environmental variables, based on information on their marginal extreme values and the dependence between them. The method was originally developed only for the case of large waves and high sea levels, and only for the joint return period of 100 years, using a dependence measure called the correlation factor (CF). This report extends the original concept to a wider range of return periods and to other variable-pairs, stretching the CF concept well beyond its originally intended usage. Although new values of CF are given for these additional cases, it is probably easier to think in terms of the five verbal descriptions of dependence, which retain their originally intended meanings. In this simple approach, dependence has to be assumed to take one of five levels of dependence (independent, modestly correlated, well correlated, strongly correlated and super correlated) for which results have been pre-computed for a number of joint return periods.

The dependence maps involving waves are colour-coded to match these five levels of dependence, as it is intended that the dependence information can be taken from the maps for use in subsequent joint probability analysis. Alternatively, if dependence is estimated in terms of ρ or of χ, Table 3.7 provides a means of estimating CF and hence which of the five particular levels of dependence to use.

When using this approach it is important to remember that all two-variable combinations determined for any particular joint return period are equally likely to occur. Therefore every combination should be considered when assessing the derived or structure variable, and only the worst case results used in design. It is also worth remembering that the joint return period used in this way provides only an approximation to the return period of the response or risk calculated from the joint exceedence conditions.

5.4 Proposed software tool

A software tool proposed to be developed to correspond to the 'desk study approach' to be described in the best practice report will solve a number of potential problems in using dependence measures and joint probability analysis in a consistent way. The tool will be capable of accepting any of three alternative dependence measures as input, i.e. ρ, or χ or CF, sorting out the theoretical differences between the three and presenting results to the user in the familiar form of joint exceedence tables. To a limited extent, depending partly on the range of return periods available for the input marginal variables, the tool will also be able to estimate joint probability density derived from the joint exceedence values. Unlike the 'simplified method' which permits only discrete values of dependence and joint return period, the tool should be capable of accepting any values. Inputs to be supplied by the user to the 'desk study approach' tool will be:

- marginal extremes of Variable 1 and of Variable 2 (this is probably easier than distribution parameters but that may be an option);
- either ρ, or χ or CF (from the maps and/or tables in this report or elsewhere);
- number of records per year;
- joint exceedence return periods required.

5.5 Monte Carlo simulation method

The maps introduced in Chapter 4 assist understanding of the variations in dependence between key variable-pairs around England, Wales and Scotland, and the reasons for and implications of those variations. For most design applications, the simplified method and the proposed software tool will provide enough information on appropriate loadings, in terms of combinations of waves, sea levels, river flows etc.

In some applications it may be helpful to be able to combine the marginal distributions, the dependence function and the resulting joint extreme values in a more thorough way than is possible using the dependence maps and simplified joint probability analysis methods. These include risk analysis involving the distribution and extremes of derived or structure variables (e.g. river flooding, overtopping or breaching). The recommended approach is to use Monte Carlo simulation to generate a very large sample of data with the distributions of each of the marginal variables, the extremes of the marginal variables, and the dependence between the marginal variables. This would provide the flexibility to convert to equivalent long-term distributions of flood risk variables, and to test sensitivity to uncertainties, in addition to providing the usual design loading conditions.

The most practical way of generating the long-term simulation is to use JOIN-SEA or equivalent programs, as described in Section 3.2. The programs take as input, the distributions and extremes of the marginal variables, improved predictions of extremes if available from other sources, and a dependence function specified in terms of one or two correlation coefficients, ρ. Although developed primarily for large waves and high sea levels, JOIN-SEA can be applied equally well to other variable-pairs where dependence can be specified in terms of ρ. Guidance on the use of JOIN-SEA will be given in the accompanying best practice report (Defra / Environment Agency, 2003a).

6. REFERENCES

E Barrow and M Hulme (1997). Describing the surface climate of the British Isles. In M Hulme and E Barrow (eds.) Climates of the British Isles - past, present and future, Routledge, London, pp 33-62.

T A Buishand (1984). Bivariate extreme-value data and the station-year method. J. Hydrol., 69, 77-95.

CIRIA (1996). Beach management manual. CIRIA Report 153.

S Coles, J Heffernan and J Tawn (2000). Dependence measures for extreme value analyses. Extremes, 2, 339-365.

S G Coles and J A Tawn (1990). Statistics of coastal flood protection. Philosophical Transactions of the Royal Society London Series A, Vol 332, pp 457-476, 1990.

Defra / Environment Agency (2002). Risk, performance and uncertainty in flood and coastal defence: A review. Defra / Agency Report No FD2302/TR1 (also referenced as HR Wallingford Report SR 587).

Defra / Environment Agency (2003a). Use of joint probability methods for flood and coastal defence in England and Wales: A guide to best practice. Defra / Environment Agency R&D Technical Report FD2308/TR2 (also referenced as HR Wallingford SR Report), in preparation.

Defra / Environment Agency (2003b). Dependence between extreme sea surge, river flow and precipitation: a study in south and west Britain. Defra / Environment Agency R&D Technical Report FD2308/TR3 (also referenced as C Svensson and D A Jones (2003), Dependence between extreme sea surge, river flow and precipitation: a study in south and west Britain, CEH Wallingford Report).

Defra / Environment Agency (2003c). Extreme water levels in estuaries and rivers: The combined influence of tides, river flows and waves. Defra / Environment Agency R&D Technical Report FD0206/TR1 (also referenced as HR Wallingford SR Report), in preparation.

Department of Energy (1977). Guidance on the design and construction of offshore structures: Environmental considerations. Department of Energy design guide.

Department of Energy (1984). Environmental parameters on the United Kingdom Continental Shelf. Department of Energy Offshore Technology Report OTH 84 201.

M J Dixon and J A Tawn (1994). Extreme sea-levels at the UK A-Class sites: site-by-site analyses. Proudman Oceanographic Laboratory Internal Document No. 65.

M J Dixon and J A Tawn (1995). Extreme sea-levels at the UK A-class sites: optimal site-by-site analyses and spatial analyses for the east coast. Proudman Oceanographic Laboratory Internal Document No 72.

M J Dixon and J A Tawn (1997). Estimates of extreme sea conditions: Spatial analyses for the UK coast. Proudman Oceanographic Laboratory Internal Document No 112.

D Faulkner (1999). Rainfall frequency estimation. Volume 2 of the Flood Estimation Handbook, Institute of Hydrology, Wallingford.

N P Gillett, H F Graff and T J Osborn (2002). Climate change and the North Atlantic Oscillation. In J W Hurrell, Y Kushnir, G Otterson and M Visbeck (eds.), The North Atlantic Oscillation - Climatic significance and environmental impact, AGU Monograph Series, AGU.

P J Hawkes, B P Gouldby, J A Tawn and M W Owen. The joint probability of waves and water levels in coastal defence design. Journal of Hydraulic Research, 2002.

N S Heaps (1967). Storm surges. Oceanogr. Mar. Biol. Ann. Rev., 5, 11-47.

HR Wallingford (1994). Validation of joint probability methods for large waves and high water levels. HR Report SR 347.

HR Wallingford (1997). Swell and bi-modal wave climate around the coast of England and Wales. HR Report SR 409, November 1997.

HR Wallingford (2000a) with Lancaster University. The joint probability of waves and water levels: JOIN-SEA: A rigorous but practical new approach. HR Report SR 537, originally dated November 1998, re-issued with minor amendments in final form in May 2000.

HR Wallingford (2000b) with Lancaster University. The joint probability of waves and water levels: JOIN-SEA Version 1.0: User manual. HR Report TR 71, originally dated November 1998, re-issued with minor amendments in final form in May 2000.

I W Hurrell (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. Science, 269, 676-679.

JOIN-SEA Version 1.0 program and data file package JOIN-SEA_uncertainty Version 1.0 program and data file package JOIN-SEA uncertainty programs: user notes	Computer programs and user notes for JOIN-SEA, in principle freely available but in practice limited to specialist users at present

D A Jones (1998). Joint probability fluvial-tidal analyses: Structure functions and historical emulation. Institute of Hydrology, March 1998.

G W Lennon (1963). The identification of weather conditions associated with the generation of major storm surges along the west coast of the British Isles. Q. J. Roy Meteor. Soc., 89, 381-394.

M W Owen, P J Hawkes, J A Tawn and P Bortot (1997). The joint probability of waves and water levels: A rigorous but practical new approach. MAFF (now Defra) Conference of River and Coastal Engineers, Keele.

D Prandle and J Wolf (1978). The interaction of surge and tide in the North Sea and River Thames. Geophysical Journal of the Royal Astronomical Society, **55**, 203-216.
D T Pugh (1987). Tides, surges and mean sea-level. Wiley, Chichester.

D T Pugh and J M Vassie (1980). Applications of the joint probability method for extreme sea-level computations. Proceedings of the Institution of Civil Engineers, Part 2, **69**, 959-975.

D W Reed and I J Dwyer (1996). Flood frequency estimation of confluences: ideals and trials. Proceedings from the MAFF (now Defra) Conference of River and Coastal Engineers, Keele. MAFF, London, pp. 3.2.1-3.2.10.

A Robson and D W Reed (1999). Statistical procedures for flood frequency estimation. Volume 3 of the Flood Estimation Handbook, Institute of Hydrology, Wallingford.

J C Rogers (1990). Patterns of low-frequency monthly sea level pressure variability (1899-1986) and associated wave cyclone frequencies. J. Climate, 3, 1364-1379.

C Svensson and D A Jones (2000). Dependence between extreme sea surge, river flow and precipitation: a study in eastern Britain. Report to MAFF, CEH Wallingford, Wallingford, October 2000.

C Svensson and D A Jones (2002). Dependence between extreme sea surge, river flow and precipitation in eastern Britain. International Journal of Climatology, 22 (10), 1149-1168, 2002.

J A Tawn (1992). Estimating probabilities of extreme sea-levels. Applied Statistics, **41**, 77-93.

J A Tawn and J M Vassie (1989). Extreme sea levels: the joint probabilities method revisited and revised. Proceedings of the Institution of Civil Engineers, Part 2, Vol. 87, pp 429-442, 1989.

A T Walden, P Prescott and N B Webber (1982). The examination of surge-tide interaction at two ports on the central south coast of England. Coastal Engineering, **6**, 59-70.

APPENDICES

APPENDIX 1

Record of the industry consultation meeting at HR Wallingford
on 30 May 2002

Prior to the meeting
1 List of respondents to the consultation invitation
2 Summary of respondents' replies

The meeting on 30 May 2002
3 Programme for the meeting
4 Attendance list for the meeting
5 Notes from discussion during the meeting

Following the meeting
6 Titles and dates of research outlines for 2003/04/05

Appendix 1 Record of the industry consultation meeting at HR Wallingford on 30 May 2002

List of respondents to the consultation invitation, in the order received, with allegiance in Spring 2002

Marcus Francis, Halliburton
Colin Green, Middlesex University
Jonathan Tawn, Lancaster University
Simon Bingley, Environment Agency
Frank Law, Independent
Ping Dong, Dundee University
Keming Hu, Posford Haskoning
Dominic Hames, East London University
Max Beran, Independent
Alan Allison, Independent
John Horne, Defra
David Ayers, Defra
Rahman Khatibi, Environment Agency
Peter Allen-Williams, Defra
Edward Evans, Halcrow
Suresh Surendran, Environment Agency
David Blackman, Proudman Oceanographic Laboratory
Ann Calver, CEH Wallingford
Will McBain / James Lancaster, Arup Water
Jerzy Graff, BMT
Dominic Reeve, Nottingham University
Tony O'Hagan, Sheffield University
Pieter van Gelder, Delft University
John Pos, Mouchel
Robert Willows, Environment Agency
J Hutchison / Andy Parsons, Defra
David Jones, CEH Wallingford
Ian Meadowcroft, Environment Agency
Agnete Berger / Michael Drayton, RMS
Duncan Reed, Independent
Malcolm Brian, Kirk McClure Morton
Graham Siggers, ABPMER
Andrew Bradbury, New Forest District Council
Jonathan Cooper, WS Atkins
John Goudie, Defra
Matilda Kitou, High Point Rendel
Richard Sproson, Fugro Geos
Richenda Connell, UKCIP
Michael Owen, Independent
Richard Horrocks, Environment Agency

(Also acknowledgements, not included in the analysis of respondents replies, from Peter von Lany, Halcrow, Stuart Bull, WS Atkins, Z Gralewski RMC, David Harvey, Bristol University)

Summary of respondents' replies to the questions asked in the consultation invitation

Accept this invitation to participate in this consultation	Yes 37	No 3			
Involvement with joint probability is	Passing interest 5	Defence design 14	Research 20	Risk analysis 23	Regulatory 4
Knowledge of joint probability is	Newcomer 1	Passing familiarity 8	Occasional user 17	Experienced user 9	
Welcome arranged telephone discussion	Yes 21	No 19			
Interest in attending 30 May 2002 meeting	Yes 29	No 11			
Interest in occasional mailings	Yes 37	No 3			
Invitation to end-of-project open meeting	Yes 34	No 6			

JOINT PROBABILITY: DEPENDENCE MAPPING AND BEST PRACTICE
(for use in river and coastal defence work)

OPEN MEETING AT HR WALLINGFORD ON 30 MAY 2002

10.00-12.45 THE PROJECT

Introduction	(10.00-10.25)	
Chairman		Michael Owen, Independent
The Risk and Uncertainty Research Theme		Ian Meadowcroft, Agency
The Project		Suresh Surendran, Agency
Brief review of concepts and methods	(10.25-10.45)	Ben Gouldby, HRW
Review of the present project and data sets	(10.45-11.05)	Peter Hawkes, HRW
20 minute break – tea and coffee		
Detail of the desk study method to be assisted by the dependence maps	(11.25-11.50)	Peter Hawkes, HRW
The analytical method used by CEH Wallingford	(11.50-12.05)	David Jones, CEHW
The analytical method used by HR Wallingford	(12.05-12.35)	Ben Gouldby, HRW
Outline of the best practice guide for use of joint probability methods in flood and coastal defence	(12.35-12.45)	John Goudie, DEFRA
55 minute break - lunch		

1.40-3.00 CONSULTATION ON THE PROJECT

Workshop session on the best practice outline	(1.40-2.20)	Groups
Users' forum: review of outline guide and project	(2.20-3.00)	All
15 minute break – tea and coffee		

3.15-4.30 FUTURE WORK AND OTHER APPLICATIONS

Potential for use of joint probability in flood forecasting	(3.15-3.25)	Rhaman Khatibi, Agency
Application to complex structure functions	(3.25-3.40)	Ben Gouldby, HRW
Visitors' contributions and presentations on joint probability		Visitors
Completion dates for project and deliverables	(5 minutes)	Peter Hawkes, HRW
Open discussion until departure at 4.30		All

4.30-5.30 SPECIAL INTERESTS (optional smaller special interest group discussion on:)

CEH analysis of river flow, surge and rainfall	David Jones, CEHW
JOIN-SEA analysis	Ben Gouldby, HRW
Additional applications outside flood and coastal defence	Peter Hawkes, HRW
Further work and collaboration	Suresh Surendran, Agency

JOINT PROBABILITY: DEPENDENCE MAPPING AND BEST PRACTICE
(for use in river and coastal defence work)

OPEN MEETING AT HR WALLINGFORD ON 30 MAY 2002

Attendance at the meeting

Alan Allison	Independent	Group 2
Max Beran	Independent	Group 3
Agnete Berger	Risk Management Solutions	Group 2
David Blackman	Proudman Laboratory	Group 4
Ann Calver	CEH Wallingford	Group 3
Damien Crawford	WS Atkins Consultants	Group 4
Marcus Francis	Halliburton KBR	Group 1
John Goudie	DEFRA	Group 3
Ben Gouldby	HR Wallingford	Group 5
Dominic Hames	East London University	Group 5
David Harvey	Bristol University	Group 3
Peter Hawkes	HR Wallingford	Group 4
David Jones	CEH Wallingford	Group 2
Rahman Khatibi	Environment Agency	Group 2
Ian Meadowcroft	Environment Agency	Group 2
Cliff Ohl	HR Wallingford	Group 4
Michael Owen	Independent	Group 3
Dominic Reeve	Nottingham University	Group 1
Carrina Rosu	HR Wallingford	Group 5
Graham Siggers	ABP Marine Environmental Research Ltd	Group 5
Richard Sproson	Fugro GEOS	Group 5
Sun Weidong	HR Wallingford	Group 5
Suresh Surendran	Environment Agency	Group 1

Afternoon only

John Pos	Mouchel	Group 1
Paul Sayers	HR Wallingford	

Morning only

Robert Abernethy	HR Wallingford
Nigel Bunn	HR Wallingford
Kate Day	HR Wallingford
Jonathan Simm	HR Wallingford
Bridget Woods-Ballard	HR Wallingford

JOINT PROBABILITY: DEPENDENCE MAPPING AND BEST PRACTICE

(for use in river and coastal defence work)

OPEN MEETING AT HR WALLINGFORD ON 30 MAY 2002

Notes from discussion during the meeting

Discussion point	Project Team response
Morning discussion	
Will there be any attempt to validate methods or quantify uncertainties ?	Not part of *this* project, but the reports will summarise what was done in earlier joint probability projects using idealised data sets and field data on occurrences of damage. The level of accuracy we are looking for is to be within about a factor of three on joint return period (in addition to any uncertainty in the marginal variable extremes).
The discrepancy between joint exceedence return period and response return period was raised.	Discuss the potential errors and uncertainties in various analysis steps, whether they can be quantified and whether the conservative and unconservative assumptions balance out. In practice the discrepancy is quantifiable and is offset by conservative assumptions which can be incorporated elsewhere in the desk study approach. By far the largest potential for error comes from poor estimation of dependence.
An alternative approach was suggested in which the marginal distributions are extrapolated analytically, presumably building in dependence by an adjustment factor.	This, and continuous simulation, will be described but not 'promoted' in the reports.
Event and record definition was raised as an issue several times during the day.	Discuss whether event and record are effectively the same, and how they might be defined for different variable-pairs.
Can everything be expressed in terms of the same dependence measure ?	Probably yes and that is the intention for the main report.
Work Group 1: Comments on *Executive summary*	
Prefer a shorter (one-page) summary saying why we have to do joint probability analysis. The abstract sells the content to users and gives an overview. Theory in appendices; focus on methods. Think about the target audience, e.g. junior, senior and principal engineers. Discuss applicability of different project stages (pre-feasibility, feasibility) and scales. Consider whether the analysis can be done ? What data are available or required ?	The report format as described at the meeting is written into the contract for the project, although 'Executive summary' is perhaps the wrong terminology. The main volume is to include the 10-page and 50-page versions of the guide, sharing a common set of maps and tables. If we re-name the present 'Executive summary' to 'Introductory user's guide' and add a normal one-page summary, that seems to meet the requirements. The other points can also be addressed.

Work Group 2: Comments on *When to use joint probability analysis*	
Say when joint probability needs variable-pairs or when multiple variables may be involved (**). Outline a strategy for use of joint probability analysis, perhaps using a decision tree (**). Describe use of the procedures for other variables, e.g. groundwater and antecedent conditions. Provide a justification for the desk study approach and when it is useful (**). Discuss the limitations, suitability and variables (**). Consider application to risk management as well as defence design (*). **General interest points** Performance measures > feedback > validation of solutions. Case studies (*). Identify troubleshooters (named individuals) and issue feedback forms. Integrate into decision-making process. Seamless coast, rivers, inland.	Good points, and where within the general scope of the project (** probably, * possibly) will be addressed.
Work Group 3: Comments on *Dependence mapping*	
3.1 Introduction Why do it ? Better use and understanding of data > steers towards probabilistic approach. **3.2 Methods and definitions** Different definitions OK provided properly described, include temporal and inter-variable dependence. Try to standardise methods and terminology where possible. Recognise the possibility of changing dependence and its impact on the purpose of using data. **3.3 Data sources** Are the data up to the job ? Are there enough data ? Consider event and record definition. Consider seasonality. **3.4 Results** Consider impacts of long-term change, e.g. climate, bathymetry etc, and whether data may need ‘correction’. Warn about site-specific use and interpretation. 3.4.1 Discuss impact of wave period.	These points are accepted and will be addressed in the reports.
Work Group 4: Comments on *The desk study approach*	
Describe preliminary checks on appreciation of the local issues and which variables really matter. Mention that there will be some no-go areas for mapping dependence for some variable-pairs, where values cannot be inferred.	These points are accepted and will be addressed in the project reports.

<table>
<tr><td>Discuss the use of spatially and/or temporally lagged dependence:
- try all potential lag times, effectively as different variable pairs, e.g. surge lagging rainfall by zero, one or two days;
- the issue of concurrent high flows in neighbouring catchments will be addressed to some extent by the CEH analyses.

Comments on the general approach:
- discuss sensitivity testing on the dependence assumption;
- tend to be conservative > should we build in conservatism > but be aware of how much ?

Dependency, including spatially and temporally lagged > there is a limit to the area over which it can be applied without dependence varying.</td><td></td></tr>
<tr><th colspan="2">Work Group 5: Comments on *The analytical approach*</th></tr>
<tr><td>Rather than specify an arbitrary 3-year minimum data sample, guidelines should provide information linking uncertainty in extremes and dependence to sample size.

Provide notes on uncertainty types, namely data, statistical models and statistical inference.

Perhaps include example calculations, illustrating where uncertainty arises.

Comment on de-trending issues > are the data de-trended > is it important ?

Discuss event definition and declustering methods.</td><td>These points are accepted and will be addressed in the project reports, but perhaps in discussion rather than numerical terms.</td></tr>
<tr><th colspan="2">Other afternoon discussion</th></tr>
<tr><td>There was discussion of whether either the desk study or the analytical approach required or generated time series data (the presentations had been unclear on this point).</td><td>Time series data are not needed or generated by either method (the desk study approach can be used without any data). In practice, time series are needed for construction of simultaneous records for use in JOIN-SEA, but the analysis treats data as a sample rather than as a series.</td></tr>
<tr><td>Planning the analysis and use of results, and preparation of the input data (including event definition) are important stages.</td><td>Agreed, and data preparation often takes longer than joint probability analysis. Sometimes a variable of potential interest can be shown to be unimportant and can therefore be eliminated or reduced to ‘secondary’ status.</td></tr>
<tr><td>Can trend and climate change be included ?</td><td>Yes, fairly easy to include in all approaches if planned in advance (but whether we believe the changes is another matter!).</td></tr>
<tr><td colspan="2">Rahman Khatibi’s presentation on flood forecasting sparked a lively discussion outside the scope of the present project. He welcomes further discussion.</td></tr>
<tr><td colspan="2">David Harvey said that the EPSRC FloodRiskNet information exchange and discussion forum will soon be available at floodrisknet.org.uk and that the first meeting will be on 20/09/02. He welcomes new members.</td></tr>
<tr><td colspan="2">Peter Hawkes said that one of the first deliverables would be an interim report in Autumn 2002 on work to date and outlines for new work in this subject area for consideration for starts in 2002/03 and 2003/04. Ideas welcome up to July 2002, either to Peter or to Suresh Surendran.</td></tr>
</table>

Peter Hawkes
HR Wallingford
11/06/02

Titles and dates for research outlines for 2003/04/05

No.	Proposed project title	Author	Budget	Proposed duration
1	Collective risk of river flooding: A pilot study	CEHW	~£56k	All of these projects would be spread over two financial years 2003/04 and 2004/05.
2	Updated estimates of extreme still water levels at ‘A’ Class national tide gauge sites: Spatial analyses for the UK coast	POL	~£50k	
3	Estimates of extreme still water levels in complex coastal regions	POL	~£20k	
4	Incorporation of temporal dependence (sequencing) into JOIN-SEA long-term simulation	HRW	~£65k	
5	Update the 1995 swell atlas for England and Wales, extend to Scotland and develop a software tool for the main results	HRW	~£90k	
6	Proposals 2 and 3 were also included as tasks to begin in 2004/05 as part of an alternative larger research programme called Environmental extremes: A managed programme, with a proposed budget of £100-150k per year, beginning in 2003/04.			

APPENDIX 2

Dependence measure used for river flow, precipitation and surge

Appendix 2 Dependence measure used for river flow, precipitation and surge

1 The χ dependence measure

A dependence measure specially suited for estimating dependence as the variables reach their extremes was used for analysis of river flow, precipitation and surge. The measure, χ, has been described in detail by Buishand (1984) and Coles *et al.* (2000). Buishand employed it to assess the inter-station dependence in precipitation data, whereas Coles *et al.* applied it to several different variables, among them precipitation and surge data. The following description of the method is based on Coles *et al.* (2000).

When used for bi-variate random variables (X, Y) with identical marginal distributions, the measure χ provides an estimate of the probability of one variable being extreme provided that the other one is extreme:

$$\chi = \lim_{z \to z^*} \Pr(Y > z \mid X > z), \qquad (1)$$

where z^* is the upper limit of the observations of the common marginal distribution.

In this application the marginal distributions are not likely to be identical, and are therefore transformed to become so. Further, the marginals are unknown and must be estimated using their empirical distributions. Thus, one approach to obtaining an estimate of identical marginal distributions is to simply rank each set of observations separately, and divide each rank with the total number of observations in each set. This corresponds to a transformation of the data to Uniform [0, 1] margins.

Rather than estimating χ as the limit in Equation 1, it is convenient to approach the problem in a different way. Consider the bi-variate cumulative distribution function $F(x, y) = \Pr(X \leq x, Y \leq y)$. It describes the dependence between X and Y completely. The influence of different marginal distributions can be removed by observing that there is a function C in the domain $[0,1] \times [0,1]$ such that

$$F(x, y) = C\{F_X(x), F_Y(y)\},$$

where F_X and F_Y are (any) marginal distributions. The function C is called the copula, and contains complete information about the joint distribution of X and Y, apart from the marginal distribution. This means that C is invariant to marginal transformation. The copula can be described as the joint distribution function of X and Y after transformation to variables U and V with Uniform [0, 1] margins, via $(U, V) = \{F_X(X), F_Y(Y)\}$.

The dependence measure $\chi(u)$ is defined for a given threshold u as

$$\chi(u) = 2 - \frac{\ln \Pr(U \leq u, V \leq u)}{\ln \Pr(U \leq u)} \quad \text{for } 0 \leq \text{u} \leq 1. \qquad (2)$$

This is related to χ of Equation 1 by

$$\chi = \lim_{u \to 1} \chi(u) = \lim_{u \to 1} \Pr(V > u \mid U > u).$$

The choice of the particular form in Equation 2 is justified by Coles *et al.* (2000), for $u \to 1$, using the relation

$$\Pr(V > u \mid U > u) = \frac{\Pr(U > u, V > u)}{\Pr(U > u)} = \frac{1 - 2u + C(u,u)}{1-u} = 2 - \frac{1 - C(u,u)}{1-u}$$
$$\approx 2 - \frac{\ln C(u,u)}{\ln u}.$$

As the variables approach their extremes, $\chi = 1$ signifies total dependence and $\chi = 0$ signifies independence or negative dependence. The value of χ can be interpreted as the risk that one variable is extreme, given that the other is. Suppose that one variable exceeds the threshold corresponding to a certain (small) exceedence probability. Then, if the dependence between the variables is estimated to be $\chi = 0.1$, it means that there is a 10% risk of the other variable exceeding the threshold corresponding to the same probability. Therefore once, on average, in ten successive periods of 10 years, one would expect the 10 year return period value of the first variable to be accompanied by the 10 year return period value of the second variable. In the same period of 100 years, one would expect ten occasions where 1 year return period values of both variables are exceeded during the same record, and a 10% chance of 100 year values of both variables.

Equation 2 is the measure of dependence used for river flow & surge and for precipitation & surge in the present study. It can be evaluated at different quantile levels *u*. This will be discussed further below. For the moment, suppose that a particular level *u* is selected, which corresponds to threshold levels (x^*, y^*) for the observed series. In practice, Equation 2 is applied by counting the number of observation-pairs, (X, Y), so that

$$\Pr(U \le u, V \le u) = \frac{\text{Number of } (X,Y) \text{ such that } X \le x^* \text{ and } Y \le y^*}{\text{Total number of } (X,Y)} \tag{3}$$

and

$$\ln \Pr(U \le u) = \frac{1}{2} \ln\left(\frac{\text{Number of } X \le x^*}{\text{Total number of } X} \cdot \frac{\text{Number of } Y \le y^*}{\text{Total number of } Y}\right). \tag{4}$$

For much of the rest of this report, χ will be used as a short-hand symbol for χ(*u*) for a given way of choosing *u*, rather than denoting the limit as expressed by Equation 1.

2 Selection of threshold level

Rather than using the Uniform distribution for the margins, the data are transformed onto an annual maximum non-exceedence probability scale. This affects only the selection of the thresholds, and χ is calculated as outlined in Equations 2 to 4. The transformation enables interpretation of the dependence between the variables in a familiar context: that of different return periods. The annual maximum non-exceedence probability, a, is

$$a = \Pr(Annual\ maximum \leq x)\,, \tag{5}$$

where x is the magnitude of the variable. It relates to the return period, T_a, as $T_a = 1/(1\text{-}a)$. The transformation is achieved through a peaks-over-threshold (POT) approach, which is considered to give a more accurate estimate of the probability distribution than using only the annual maximum series (e.g. Stedinger *et al.*, 1993). The non-exceedence probability, p, of the POT series with a rate of λ events per year, is related to that of the annual maximum as

$$a = \exp(-\lambda(1-p))\,, \tag{6}$$

where 1-p is the exceedence probability of the POT series which can be estimated using a plotting position. Hazen's plotting position is a traditional choice, and leads to the estimate

$$\lambda(1-p) = \frac{N_e}{N}\cdot\frac{i-0.5}{N_e} = \frac{i-0.5}{N}\,, \tag{7}$$

where i is the rank of the independent POT events, N_e is the total number of POT events, and N is the number of years of observations. The highest observation is given rank 1, the second highest rank 2, etc. The independence criterion used in this study was that two POTs must not occur on consecutive days, but be separated by at least three days. Thus, substituting Equation 7 into Equation 6 results in the following transformation to the annual maximum scale:

$$a = \exp(-\frac{i-0.5}{N})\,. \tag{8}$$

The magnitude of x in Equation 5 corresponds to the magnitude of the POT with rank i in Equation 8 for the same annual maximum non-exceedence probability, a.

The dependence measure χ can be estimated for any threshold. Initial trials showed a fairly constant, slightly decreasing, value of χ for annual maximum non-exceedence probabilities between about 0.1 and 0.5. For higher probabilities, χ tended to become 0 as no observation-pairs exceeded both thresholds (Appendix B of Svensson and Jones, 2000). The threshold was selected to be $a = 0.1$. This corresponds to selecting a threshold for the data values that about 2.3 events per year will exceed. The annual maximum will exceed this threshold in 9 out of 10 years. The use of a threshold in this sort of range is dictated by two requirements: to have enough data points above the threshold in order to be able to estimate dependence reliably, and for the threshold to be high enough to regard the data points as extreme.

3 Missing data

Only observation-pairs where both observations in the pair were available were included in the count in Equations 2 to 4. A minimum of 1825 observation-pairs, equivalent to five complete years of simultaneous data, was set as a requirement for χ to be estimated reasonably reliably.

However, when estimating the threshold levels (x^*, y^*) for the margins, each margin was treated separately so that as much information as possible was used. The number of years, N, in Equations 7 and 8 was thus calculated for each series as

$$N = \frac{N_c}{N_t} N_{orig}$$

where N_c is the number of days with complete observations, N_t is the total number of days, and N_{orig} is the total number of years in the study period. Note that N is treated as a non-integer.

4 Significant dependence

The values of χ corresponding to the 5% significance level were estimated using a permutation method (e.g. Good, 1994). This type of method is used to generate data sets in which independence would hold. A large number of data sets are generated and a test statistic, in this case χ, is calculated for each of these new data sets. This provides a sample of χ corresponding to independently occurring data. If the χ calculated for the original data set is rather different to most of the χ calculated from the generated values, then this suggests that the two original records are not independent. Dependence occurring because both records show similar seasonal characteristics can be accounted for by generating data that show the same seasonal characteristics.

Two slightly different permutation methods were used for the east and for the west and south coasts, prompted by the larger amount of missing sea level data for the latter coasts. In the east coast study, which was carried out a few years prior to the present work, one of the records for each station-pair was permuted while the other was kept unchanged. The permutation of the data was performed by randomly reshuffling intact blocks of one year, in order to preserve the seasonality. Using all the years in the series works well for almost complete data records. However, for the west and south coasts, only years with observations were used for the reshuffling. Thus, a random resample of years (with observations) was drawn from each of the two series, so that the number of years in each resample equalled the number of years with any concurrent data in the original two series. Each year could be represented only once in each resample, to resemble a true permutation. For both methods no year was allowed to be paired up with itself, and leap years were permuted separately to non-leap years.

In total, 199 permutations of the data were made for each station-pair and a new χ was calculated each time. The 199 χ values were subsequently ranked in descending order and the 10th largest value was accepted as corresponding to the 5% significance level, or the 95% point of the null distribution (the distribution of values that would occur if data-pairs were independent). The actual value of this 5% significance level varied between about 0.02 and 0.09 in this study. If the χ calculated for the original series exceeded this value, then the data provide reasonably strong evidence that the dependence between the variables can be considered genuine. A stronger significance level was not used because this would have required a greater number of permutations. These were very time-consuming, and, considering the high number of station combinations, it was not deemed practical.

5 Confidence intervals

Confidence intervals give an indication of the range of values within which the 'true' dependence χ can be expected to lie. In the absence of infinitely long records, this true value is unknown. A bootstrapping method (e.g. Efron, 1979) was used to estimate the confidence intervals. Similar to the permutation method used for estimating significance, bootstrapping can be used where the underlying statistical population is unknown or where an analytical solution is impractical.

Bootstrapping is based on the generation of many new data set resamples. In contrast to when significance levels were estimated and independence between the two series was sought, each observation-pair is here kept intact and treated as one record.

The original sample of observation-pairs is used as the distribution from which the resamples are chosen randomly with replacement, i.e. with each observation-pair being returned to the original sample after it has been chosen, so that it may be chosen again. A large number of data sets are generated and a test statistic, in this case χ, is calculated for each of these new data sets. This provides a sample of χ that would occur for a range of situations, as χ is calculated from some resamples including many data-pairs consistent with dependence, and from some resamples including many data-pairs consistent with independence. Seasonality is kept intact by sampling in blocks of one year, rather than using individual observation-pairs.

In this study balanced resampling (e.g. Fisher, 1993) was used, which is a more efficient method. It ensures that each year occurs equally often overall among the total number of bootstrap samples. This is implemented by creating a vector of length BN consisting of the N years of record repeated B times. This array is then randomly reshuffled, and divided into slices of length N, to obtain B bootstrap samples.

In total, $B = 199$ bootstrap samples of the data were made for each station-pair and a new χ was calculated each time. The 199 χ values were subsequently ranked in descending order and the 10th and 190th largest values were accepted as delimiting the 90% confidence interval.

Because the computations were very computationally demanding, confidence intervals were estimated only for the primary variable-pair, surge & flow (precipitation being used only to aid in the interpretation of why dependence occurs).

6 References

T A Buishand (1984). Bivariate extreme-value data and the station-year method. J. Hydrol., 69, 77-95.

S Coles, J Heffernan and J Tawn (2000). Dependence measures for extreme value analyses. Extremes, 2, 339-365.

B Efron (1979). Bootstrap methods: another look at the jack-knife. Annals of Statistics, 7, 1-26.

N I Fisher (1993). Some modern statistical techniques for testing and estimation. Chapter 8 in Statistical analysis of circular data, Cambridge Univ. Press, pp 199-218.

P Good (1994). Permutation tests. Springer, New York.

J R Stedinger, R M Vogel and E Foufoula-Georgiou (1993). Frequency analysis of extreme events. In D R Maidment (ed.), Handbook of Hydrology, McGraw-Hill, London, pp 18.1-18.66.

C Svensson and D A Jones (2000). Dependence between extreme sea surge, river flow and precipitation: a study in eastern Britain. Report to MAFF, CEH Wallingford, Wallingford, October 2000.

APPENDIX 3

Comparison of dependence measures

Appendix 3 Comparison of dependence measures

1 Introduction to the correlation factor and χ models

The ‘simplified method’ described in Section 3.5 uses various levels of dependence, described in terms of their relationships to independence and full dependence, implicitly defining two factors for measuring the level of dependence. Here using c for the correlation factor described in Section 3.5, the relationships to independence and full dependence can be written,

$$c = \frac{p_b}{p_i} \quad \text{or} \quad p_b = cp_i$$

and

$$d = \frac{p_d}{p_b}$$

where

p_b = probability of joint exceedence taking account dependence,

p_i = probability of joint exceedence assuming independence,

p_d = probability of joint exceedence assuming full dependence.

On combining these,

$$cp_i = p_b = \frac{p_d}{d}$$

and, since both c and d are typically greater than one

$$p_i \le cp_i = p_b = \frac{p_d}{d} \le p_d \,.$$

If both variables have the same marginal probability p then

$$p^2 \le cp^2 = p_b = \frac{p}{d} \le p \,.$$

From this viewpoint the correlation factor, or any measure of dependence, characterises how far the joint probability of exceedence is between p^2 and p. However, a complete description requires consideration of non-equal marginal probabilities.

Unfortunately, the simple argument based on the correlation factor approach does not yield a valid probability model. For example, if the proposed model were

$$p_b = cp_1p_2 \,,$$

(where p_1, p_2 are the marginal probabilities) this would yield $p_b = cp_1$ when $p_2 = 1$, whereas it should give $p_b = p_1$ because this is required by the rules for marginal probabilities. It turns out that the above formula is not actually used in practice, since c is implicitly allowed to vary with p_b (this is outlined below). This modification does not overcome the underlying problem, but does yield a slightly different formula from that above. A similar defect arises for the simplistic formulae so far used for the "χ" model. Both formulae can be revised in similar ways to yield valid probability models while retaining the spirit of the original intentions. For example, the revised correlation factor model requires, in its initial form,

$$p_b = \{p_1 p_2\}^a$$

for some power a. The re-revised form of this is able to give a valid probability law while having the property that the above formula holds exactly for $p_1 = p_2$.

In the following the "correlation factor" approach is modified several times before the final proposed model is reached. Here K is used to denote the number of items being considered per year, and return periods are measured in years. For example, if T_b is the return period of both thresholds being exceeded, then

$$p_b = (KT_b)^{-1}.$$

2 The revised correlation factor model

The basic form of the model: i.e. –

$$c = \frac{p_b}{p_i} \quad \text{or} \quad p_b = cp_i \tag{1}$$

is actually applied in such a way that the "effective correlation factor", c, used in Equation (1) increases with the return period of interest T_b. For a correlation factor γ selected as appropriate for a 100 year return period, the correlation factor for a return period of T_b years is given by

$$\log\left(\frac{c}{2}\right) = \frac{\log\left(\frac{KT_b}{2}\right)\log\left(\frac{\gamma}{2}\right)}{\log\left(\frac{K \times 100}{2}\right)}$$

or

$$c = 2\left(\frac{KT_b}{2}\right)^{\delta} \qquad \text{where } \delta = \frac{\log\left(\frac{\gamma}{2}\right)}{\log\left(\frac{K\times 100}{2}\right)}.$$

Note that values of γ from 2 to $(K\times 100)$, which is the range "no dependence" to "complete dependence", gives values of δ from 0 to 1. Note also that with $\gamma = 20$, $\gamma = 100$ or $\gamma = 500$ for $K = 706$ (described as modestly, well or strongly correlated respectively), $\delta = 0.2819$, $\delta = 0.4789$ or $\delta = 0.6759$. This suggests that δ may be a reasonably "stable" measure of dependence. However, in this formulation the structure does not work correctly, since $\delta = 0$ yields $c = 2$ which is not the correct "correlation factor" for independence ($c = 1$ is the required value for the theory to work).

Given the above it is suggested that the conversion of "correlation factor" for differing return periods should be restructured as follows, leaving the imposition of a minimum value of $c = 2$ to a later stage of any practical implementation.

$$\log(c) = \frac{\log(KT_b)\log(\gamma)}{\log(K\times 100)}$$

or

$$c = (KT_b)^{\delta} \qquad \text{where } \delta = \frac{\log(\gamma)}{\log(K\times 100)} \qquad (2)$$

Note that values of γ from 1 to $(K\times 100)$, which is (now) the range "no dependence" to "complete dependence", gives values of δ from 0 to 1. Note also that with $\gamma = 20$, $\gamma = 100$ or $\gamma = 500$ for $K = 706$ (described as modestly, well or strongly correlated respectively), $\delta = 0.2683$, $\delta = 0.4125$ or $\delta = 0.5566$. Now $\delta = 0$ yields $c = 1$.

The restriction that c be at least 2 might be imposed as

$$c^* = \max\left\{2, (KT_b)^{\delta}\right\}.$$

With the improved formulation,

$$c = (KT_b)^{\delta} = p_b^{-\delta}$$

and the underlying model for joint probabilities can be expressed as

$$p_b = p_b^{-\delta} p_i,$$

giving

$$p_b = \{p_1 p_2\}^{\frac{1}{1+\delta}}.$$

This appears to give reasonable answers even for $\delta > 1$. However, the formulation is still flawed (for all values of δ) because it fails to give the correct results that

$$p_b = p_1 \text{ when } p_2 = 1$$
$$p_b = p_2 \text{ when } p_1 = 1.$$

Note that this same problem arises for the "χ" model that we have used so far for the case where the two variates do not have the same return period.

3 The "χ" model

To compare the "correlation factor" model with the "χ" model, note that the latter is defined in terms of non-exceedence probabilities

$$f_1 = 1 - p_1, \; f_2 = 1 - p_2$$

$$f_b = 1 - p_1 - p_2 + p_b$$
$$p_b = 1 - f_1 - f_2 + f_b .$$

The model previously outlined as the "χ" model is then

$$f_b = (f_1 f_2)^{1-\frac{1}{2}\chi} . \tag{3}$$

For this to be a properly defined model for the joint distribution function, a requirement is that when $f_1 = 1$, $f_b = f_2$ which this model fails.

To compare with the correlation factor model, the above gives

$$\begin{aligned} p_b &= 1 - f_1 - f_2 + f_b \\ &= 1 - f_1 - f_2 + (f_1 f_2)^{1-\frac{1}{2}\chi} \\ &= 1 - (1 - p_1) - (1 - p_2) + \{(1 - p_1)(1 - p_2)\}^{1-\frac{1}{2}\chi} \\ &= p_1 + p_2 - 1 + \{(1 - p_1)(1 - p_2)\}^{1-\frac{1}{2}\chi} \end{aligned}$$

which is then directly comparable with (1).

4 The revised "χ" model

The problem with the existing "χ" model can be overcome by replacing Equation (2) with something which has the right properties. One example is

$$f_b = \exp\left(-\left\{\left(-\log f_1\right)^{1/\alpha} + \left(-\log f_2\right)^{1/\alpha}\right\}^{\alpha}\right),$$

in which the parameter α is directly related to χ as

$$\chi = 2 - 2^{\alpha}, \qquad \text{or} \qquad \alpha = \frac{\log(2-\chi)}{\log 2}.$$

The above is based on the “logistic” bi-variate extreme-value model which is one which has the required property on the diagonal:

$$f_b = \left(f_1 f_2\right)^h \text{ when } f_1 = f_2 \text{ for some constant } h.$$

Here $h = 2^{\alpha}$. Thus the revised “χ” model would have

$$p_b = p_1 + p_2 - 1 + \exp\left(-\left\{\left(-\log(1-p_1)\right)^{1/\alpha} + \left(-\log(1-p_2)\right)^{1/\alpha}\right\}^{\alpha}\right).$$

When $p_1 = p_2 = p$,

$$\begin{aligned} p_b &= 2p - 1 + (1-p)^{2^{\alpha}}, \\ &= 2p - 1 + (1-p)^{2-\chi}. \end{aligned} \tag{4}$$

A power series expansion of this for small p gives

$$p_b \approx \chi p + (1-\chi)(1 - \tfrac{1}{2}\chi)p^2 + ... \tag{5}$$

and hence the leading term is linear in p, except when $\chi = 0$ when p_b is (an exact) quadratic.

5 The re-revised correlation factor model

The corresponding correlation factor model would have

$$p_b = \exp\left(-\left\{\left(-\log p_1\right)^{1/\alpha} + \left(-\log p_2\right)^{1/\alpha}\right\}^{\alpha}\right), \tag{6}$$

where, to correspond to the earlier revision (i.e. to give the required result when $p_1 = p_2$),

$$\frac{2}{1+\delta} = 2^{\alpha}, \qquad \alpha = 1 - \frac{\log(1+\delta)}{\log 2}.$$

When $p_1 = p_2 = p$,

$$\begin{aligned} p_b &= p^{2^{\alpha}}, \\ &= p^{2/(1+\delta)}. \end{aligned} \tag{7}$$

Hence in this case p_b is a fractional power of p, where the power varies from 2 (for independence) to 1 (for complete dependence).

6 The Bi-variate Normal Threshold model

For the Bi-variate Normal Threshold model, the probability law being modelled is restricted to a range such that p_1, p_2 are both small. Let p_L be the upper limit of the exceedence probability range within which the bi-variate Normal distribution applies, and let ρ be the correlation of this component distribution. Then, for p_1, p_2 both less than p_L,

$$p_b = \Phi_2\left\{\Phi^{-1}(p_1), \Phi^{-1}(p_2); \rho\right\}, \tag{8}$$

where Φ^{-1} is the inverse of the standard uni-variate Normal distribution function and Φ_2 is the standard bi-variate distribution function,

when $p_1 = p_2 = p$,

$$p_b \approx C_\rho p^{2/(1+\rho)}, \tag{9}$$

where

$$C_\rho = (1+\rho)^{3/2}(1-\rho)^{-1/2}(4\pi)^{-\rho/(1+\rho)}.$$

The above approximation is derived from Bortot and Tawn (1997; p37), which in turn is based on Ledford and Tawn (1996). The additional paper Ledford and Tawn (1997) is also related.

Hence the behaviour of p_b in this case is much like that of the re-revised correlation factor method, being a fractional power of p, except that a multiplying factor is involved in addition which depends on ρ.

7 Comparison of models

The most readily understood comparison of the behaviour of the models is obtained by taking the case $p_1 = p_2 = p$ and comparing the power-series expansions for small *p*. The revised χ model has (Equation (5))

$$p_b \approx \chi p + (1-\chi)(1-\tfrac{1}{2}\chi)p^2 + \dots \tag{5}$$

which says that p_b is a linear function of p for p close to zero, and the slope is non-zero if χ is positive. The re-revised correlation factor model has (Equation (7))

$$p_b = p^{2/(1+\delta)} \tag{7}$$

which says that p_b is a power of p and the slope of this relationship is zero when p is zero. The Bi-variate Normal Threshold model has (Equation (9))

$$p_b \approx C_\rho p^{2/(1+\rho)}, \tag{9}$$

which again says that p_b is a power of p and the slope of this relationship is zero when p is zero. However, there is an extra multiplying factor compared to that for the re-revised correlation factor model.

Thus it is apparent that the structures of the three models are radically different and one can expect that the relationships between the return periods T_1, T_2, T_b under the three models will be radically different. In addition, there is no direct way of relating the parameters measuring dependence χ, δ, ρ of the models because the structures of the models are so different.

8 Matching parameters across models

The following approach to finding corresponding parameter values for the three models is suggested on the basis of the way in which the χ model is presently being fitted in the current CEH Wallingford work on measuring dependence. For this, thresholds for the two variables are chosen to have a selected exceedence probability corresponding to 2.3 events per year. This equates to $p_{fix} = 2.3/K$, or

$p_{fix} = 0.0063$ for daily data,
$p_{fix} = 0.0033$ for tidal-peak data.

Then, to estimate χ, the procedure is equivalent to estimating p_b and solving Equation (4) for χ. This gives

$$\chi = 2 - \frac{\log\left(p_b - 2p_{fix} + 1\right)}{\log\left(1 - p_{fix}\right)}$$

One way of matching parameters across the models is to require that the values of p_b for a given p_{fix} should agree for the three models. Thus, to match χ and correlation factor models, one can start with γ or δ (which are related by Equation (2)), use Equation (7) to determine p_b and then use Equation (10) to find χ as

$$\chi = 2 - \frac{\log\left(p_{fix}^{2/(1+\delta)} - 2p_{fix} + 1\right)}{\log\left(1 - p_{fix}\right)}.$$

It is clear that different pairs (χ, δ) will match for different selections of p_{fix}. Similarly, for the Bi-variate Normal Threshold model,

$$\chi = 2 - \frac{\log\left(\Phi_2\left\{\Phi^{-1}\left(p_{fix}\right), \Phi^{-1}\left(p_{fix}\right); \rho\right\} - 2p_{fix} + 1\right)}{\log\left(1 - p_{fix}\right)}.$$

9 Tables of pairs of parameters constructed on this basis

```
Table for pfix= 0.00630   K=  365.25

    delta        corrfact    chi
    0.000            1.00  0.00000
    0.050            1.69  0.00394
    0.100            2.86  0.00962
    0.150            4.83  0.01749
    0.200            8.18  0.02807
    0.250           13.82  0.04191
    0.300           23.38  0.05957
    0.350           39.53  0.08163
    0.400           66.84  0.10868
    0.450          113.02  0.14128
    0.500          191.12  0.17999
    0.550          323.17  0.22533
    0.600          546.45  0.27781
    0.650          924.02  0.33788
    0.700         1562.48  0.40594
    0.750         2642.06  0.48239
    0.800         4467.58  0.56753
    0.850         7554.43  0.66166
    0.900        12774.13  0.76499
    0.950        21600.35  0.87773
    0.999        36143.29  0.99746
    1.000        36525.00  1.00000
```

```
Table for pfix= 0.00330   K=  706.00

    delta        corrfact    chi
    0.000            1.00  0.00000
    0.050            1.75  0.00240
    0.100            3.05  0.00606
    0.150            5.34  0.01141
    0.200            9.33  0.01896
    0.250           16.30  0.02929
    0.300           28.49  0.04302
    0.350           49.78  0.06085
    0.400           87.00  0.08350
    0.450          152.04  0.11172
    0.500          265.71  0.14627
    0.550          464.35  0.18793
    0.600          811.49  0.23746
    0.650         1418.15  0.29561
    0.700         2478.35  0.36311
    0.750         4331.15  0.44064
    0.800         7569.09  0.52886
    0.850        13227.69  0.62837
    0.900        23116.61  0.73974
    0.950        40398.43  0.86346
    0.999        69816.15  0.99714
    1.000        70600.00  1.00000
```

```
Table for pfix= 0.00630   K=  365.25

    rho      chi
    0.000  0.00000
    0.050  0.00295
    0.100  0.00687
    0.150  0.01196
    0.200  0.01845
    0.250  0.02659
    0.300  0.03666
    0.350  0.04896
    0.400  0.06385
    0.450  0.08172
    0.500  0.10303
    0.550  0.12833
    0.600  0.15829
    0.650  0.19374
    0.700  0.23579
    0.750  0.28597
    0.800  0.34652
    0.850  0.42111
    0.900  0.51670
    0.950  0.65071
    0.999  0.94958
    1.000  1.00000
```

```
Table for pfix= 0.00330   K=  706.00

    rho      chi
    0.000  0.00000
    0.050  0.00182
    0.100  0.00437
    0.150  0.00784
    0.200  0.01246
    0.250  0.01850
    0.300  0.02624
    0.350  0.03603
    0.400  0.04826
    0.450  0.06337
    0.500  0.08190
    0.550  0.10448
    0.600  0.13186
    0.650  0.16500
    0.700  0.20516
    0.750  0.25406
    0.800  0.31419
    0.850  0.38962
    0.900  0.48795
    0.950  0.62811
    0.999  0.94604
    1.000  1.00000
```

10 Comparison of return periods for various models

The plots show contours of base-10 logarithms of return periods, where the axes are the marginal return periods also on a base-10 logarithm scale. Thus 0 corresponds to 1 year, 2 to 100 years.

There are two sets of contours on each plot, one set (shown in black) are return periods from the χ model while the second one (shown in colour) are return periods for either the "correlation factor" model (shown in red) or for the bi-variate normal model (shown in green). On each plot the parameters of the models have been matched to give the same return period at a return period of 2.3 samples per year (-0.36 on the base-10 logarithm scale).

The parameters have been selected so that a plot having contours for the correlation factor model can be compared with one having contours for the bi-variate normal model by finding the plot for which the value of χ nearly matches.

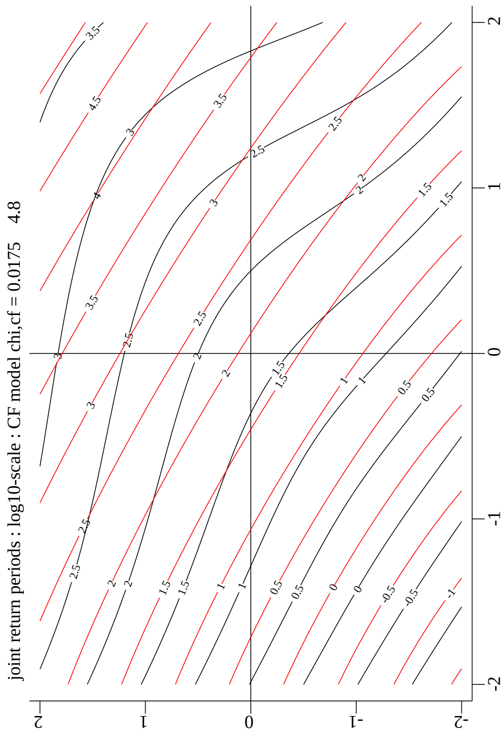
11 References

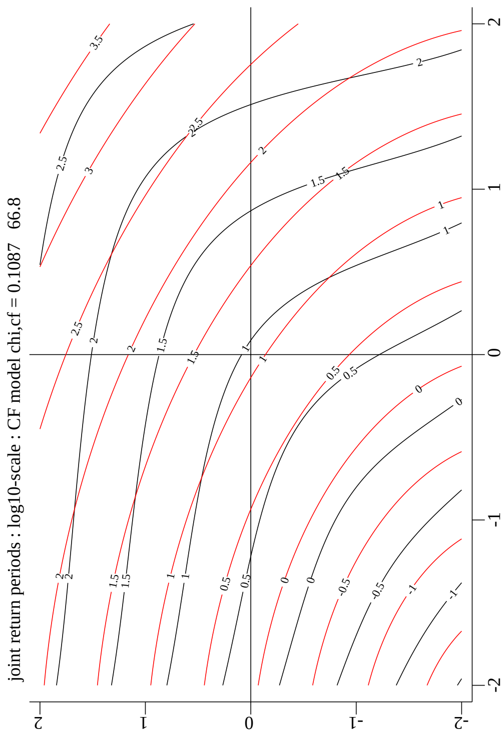
Bortot, P and Tawn, J.A. (1997). Joint probability methods for extreme still water levels and waves. In "The Joint Probability of Waves and Water Levels: JOIN-SEA", HR Wallingford Report SR 537 (November 1998).

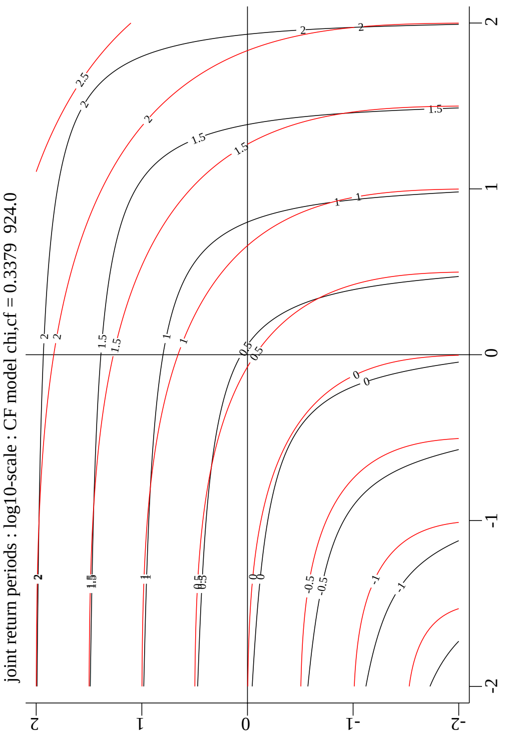
Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. Biometrika, 83(1), 169-187

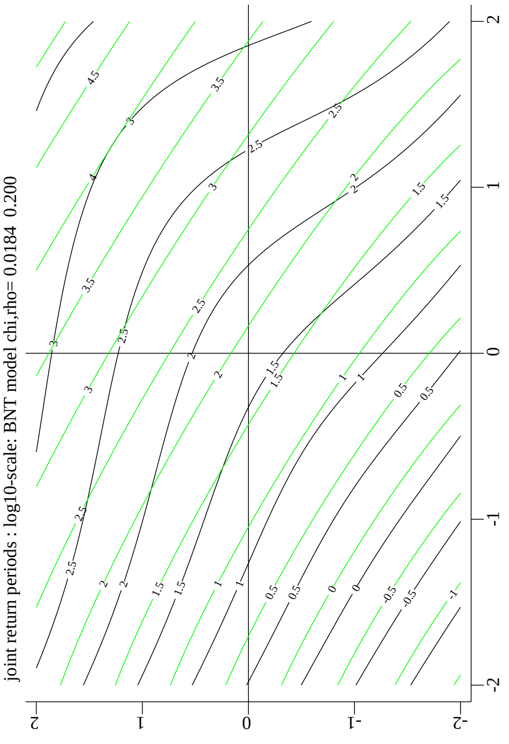
Ledford, A.W. and Tawn, J.A. (1997) Modelling dependence within joint tail regions. J. R. Statist. Soc. B, 59 (2), 475-499

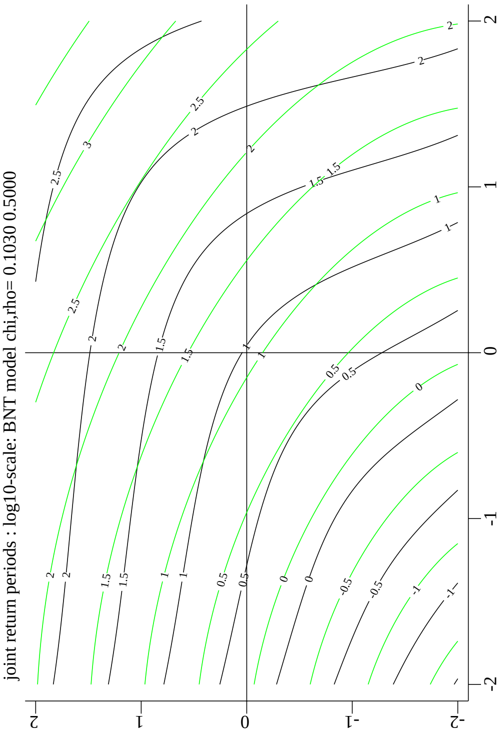
CEH Wallingford 25/07/2002


joint return periods : log10-scale : CF model chi,cf = 0.0175 4.8
-2
-1
0
1
2

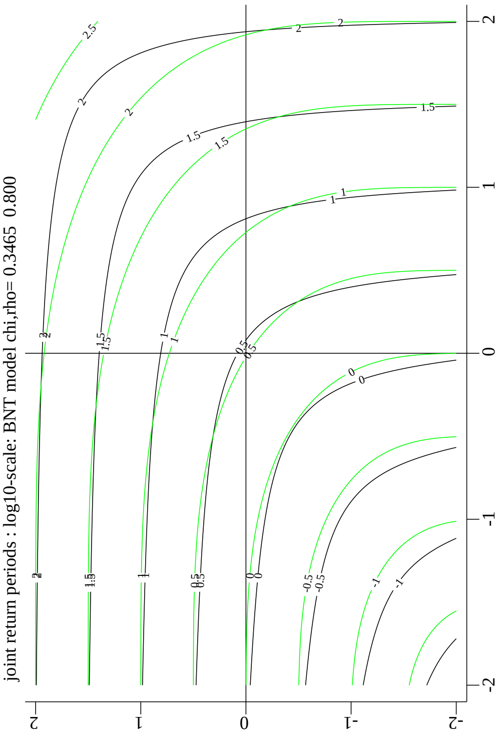

joint return periods : log10-scale : CF model chi,cf = 0.1087 66.8
-2
-1
0
1
2


joint return periods : log10-scale : CF model chi,cf = 0.3379 924.0


joint return periods : log10-scale: BNT model chi,rho= 0.0184 0.200



joint return periods : log10-scale: BNT model chi,rho= 0.1030 0.5000



joint return periods : log10-scale: BNT model chi,rho= 0.3465 0.800