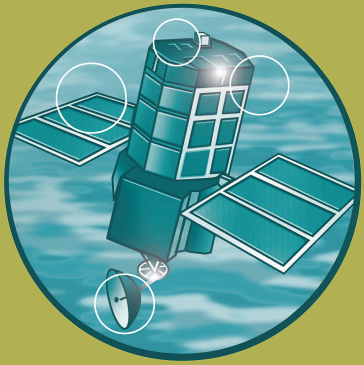
Defra/Environment Agency
Flood and Coastal Defence R&D Programme

Use of Joint Probability Methods in Flood Management

A Guide to Best Practice

R&D Technical Report FD2308/TR2



Environment Agency

defra
Department for Environment
Food and Rural Affairs

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

Use of Joint Probability Methods in Flood Management

A Guide to Best Practice

R&D Technical Report FD2308/TR2

March 2005

Author:
PJ Hawkes

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 R&D output from the Joint Defra / Environment Agency Flood and Coastal Defence R&D Programme.

Dissemination Status
Internal: Released Internally
External: Released to Public Domain

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

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.

Publishing organisation
Defra - Flood Management Division
Ergon House
Horseferry Road
London SW1P 2AL
Tel: 020 7238 3000
Fax: 020 7238 6187
www.defra.gov.uk/environ/fcd

Published by the Department for Environment, Food and Rural affairs. Printed in the UK, March 2006 on recycled material containing 80% post-consumer waste and 20% totally chlorine-free virgin pulp.

PB No. 11209

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) within the joint Defra / Environment Agency research theme Risk Evaluation and Understanding of Uncertainty, headed by Ian Meadowcroft. Availability of this report implies no endorsement by the Department of the report's results or conclusions. The Project Managers for HR Wallingford, CEH Wallingford and the Proudman Oceanographic Laboratory were Dr Peter Hawkes, Dr Cecilia Svensson and David Blackman, respectively. The HR Wallingford job number was CBS0706.

Acknowledgements
The authors are grateful for comments provided by Michael Owen, and for his chairmanship of the open meeting during the industry consultation. Thanks also to others involved in the industry consultation and in the end-of-project meeting, listed in Appendices 1 and 2 of companion R&D Report FD2308/TR1. Acknowledgement 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 Oceanographic 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. The hourly rainfall data were made available by the Environment Agency for use by HR Wallingford. The daily precipitation and wind speed data from the HadRM3 regional climate model and the hourly sea surge data from the shelf-seas model were supplied by the UK Met Office for use by CEH-Wallingford under a project-specific licence.

SUMMARY

Joint probability methods could be used in most flood risk calculations, as flood risk is rarely a function of just one source variable (e.g. waves, sea level, river flow, rainfall) but more usually of two or three variables. Joint probability analysis gives the probability of the relevant source variables taking high values simultaneously and thus creating a situation where flooding may occur. Take-up of joint probability methods has been patchy due to two main reasons: lack of information on dependence between the source variables, and perceived difficulty in usage and interpretation of the methods.

This report is aimed at non-specialist users of joint probability methods, to encourage them to adopt and use joint probability methods without the need for specialist advice. The report includes a high level overview, which could be extracted together with example dependence plots, to be published separately in the form of an Environment Agency introductory booklet. The main best practice guide contains enough information for routine use of the methods.

The guide includes a summary of the desk study and analytical approaches to joint probability analysis, and a software tool for application of the desk study approach. It includes advice on data preparation, parameter selection, application of the methods in complex areas, incorporation of climate change allowances, and interpretation of the results of the analysis. The variable-pairs presented in this report, including enough information for calculations, are:

- wave height & sea level, relevant to most coastal flood defence studies
- river flow & surge, relevant to most river flood defence studies
- hourly rainfall & sea level, of potential use in drainage studies in coastal towns
- wind-sea & swell, of potential use in coastal engineering studies.

The guide includes outline case studies for each of the variable-pairs listed above, for each of the two main analysis methods. These include techniques for use in complex areas and for incorporation of climate change allowances.

This report is supported by a separate longer technical report (FD2308/TR1) containing more detailed information and description for experienced users. The technical report includes the project glossary, descriptions of the source data sets, derivation and comparison of the dependence measures used, and descriptions of the desk study and analytical approaches to joint probability analysis. It also includes a record of the industry consultation and a full set of dependence results, with confidence limits, including some additional variable-pairs not reproduced here, namely:

- wave height & surge
- tide & surge
- daily precipitation & surge.

CONTENTS

Figures

PART ONE: INTRODUCTORY USERS' GUIDE

1. OVERVIEW OF THE STAGES IN A JOINT PROBABILITY ASSESSMENT

This nine-page Introductory Users' Guide is in the form of a series of questions, each one followed by a one-paragraph answer. It is intended to raise appreciation of joint probability methods, when and how they might be used, potential advantages and problems, and interpretation of results. It is intended that it can be read as a stand-alone introduction to the concept of joint probability assessment in flood and coastal defence, and that it could be reproduced as an Environment Agency introductory booklet. However, it does not provide enough information to begin calculations, which would require use of the main best practice guide in Chapters 2-6 of this report and possibly the associated technical reports:

FD2308/TR1: Technical report on dependence mapping; and

FD2308/TR3: Dependence between extreme sea surge, river flow and precipitation: A study in south and west Britain.

1.1 Relevance to flood management

What is joint probability?
Joint probability refers to the chance of two or more conditions occurring at the same time. For this situation to be of interest, some overall outcome of interest should depend on the combined occurrence of these conditions. A simple example would be total score (the overall outcome) made by the throw of two or more dice (the conditions occurring at the same time). For this situation to be worthy of detailed study, the dependence between the two or more conditions should be non-trivial, i.e. neither independent nor fully dependent. Environmental variables such as rainfall, sea level, wave height and river flow are therefore suitable candidates for joint probability analysis of flood and other risks.

How does it help in flood management work?
High and extreme loading of flood defence structures is often caused by more than one environmental variable, so that the probability of a certain load occurring is related to the combined probability of occurrence of all of the variables concerned. There is often a degree of dependence between the variables, but the relative importance of the various hydraulic and meteorological effects causing that dependence will vary from place to place, making it difficult to estimate overall probabilities. More accurate information on dependence, whether from dependence maps or site-specific analysis, will permit more accurate evaluation of risk due to extreme events, therefore increased confidence in design and targeting of new works, and potentially better value for money in defence works.

Which environmental parameter(s) and structure variables are of interest?
In river engineering, the combined effects of river flow with one or more of tide, surge, rainfall and possibly waves, are of interest in determining the overall river water level, and consequent likelihood of out-of-bank flow and flooding. In coastal engineering, the combined effects of sea level and waves are of interest in determining overall loads on

coastal structures, and consequent likelihood of damage or severe overtopping and flooding. In urban drainage of coastal towns, the combined effects of sea level and high intensity rainfall are of interest in determining the probability of tide-locking of drains. Better knowledge of the probability of flooding would facilitate better assessment of flood and coastal defences, and subsequent decision-making, where benefit / cost is important. This would be the case, for example, in assessing the standard of service of existing defence structures, assessing the need for new defences, and comparing the relative value of different defence schemes competing for funding.

Who might need to use joint probability methods?

1. Policy makers, e.g. Defra and HSE, may need to understand the concepts and implications, in order to frame regulations about when to use the methods, e.g. in FCDPAG.
2. Regulators and funders, e.g. Defra and Environment Agency, may need to be familiar with the regulations and calculation methods, in order to check planning and funding applications.
3. Designers and flood risk estimators may need to be familiar with the regulations, calculation methods and data sources, for potential use in flood defence design, benefit /cost assessment, funding and planning applications, and flood risk evaluation (maybe for different development, climate change and other uncertainty assumptions).
4. Researchers and teachers may be involved in development, testing, evaluation and dissemination of calculation methods.

1.2 Planning a joint probability analysis

What information is there on each input variable?

The most important source variables will vary from one area to another and from one type of flood management consideration to another. They will usually include one or more of rainfall, river flow, wave height and sea level. It may also be desirable to analyse related variables, for example ground wetness, wave period, wave direction, wind velocity, surge or swell. Part of the preliminary work preceding analysis involves a review of existing time series data, probability distributions, reports, extremes predictions and established practices for the study area. In most parts of the UK, there are relevant measurements of rainfall, river flow, sea level, surge and wind velocity. However, information on waves, swell and ground wetness is more limited and it is more likely for these variables that numerical modelling would be needed.

How will marginal (single variable) extremes be estimated?

Marginal extremes can either be determined as part of a joint probability analysis or be predicted separately and provided as input to joint probability analysis. There may be established extreme values, either from previous studies in the same area, or from national guidelines. There may be local measured or synthesised data, to which the Weibull, Gumbel and/or other probability model(s) can be fitted and then extrapolated to extreme values. The analytical approach to joint probability analysis described in this report fits a Pareto probability distribution to the top few percent of data on each variable, and then simulates the distribution in the extreme tail. However, in practice these high values are usually replaced with more reliable values based on previously established extremes or those derived from longer data sets, in order to ensure that the joint exceedence extremes calculated are consistent with knowledge of the site.

Can any input variable be excluded from explicit joint probability assessment?
Quite often, an individual variable can be excluded from joint probability analysis, either because preliminary testing shows it to have little impact on subsequent calculations and decisions, or because it is highly dependent upon one of the other variables involved. Preliminary river model runs (or field data analysis) may show that either downstream sea level or upstream flow has little impact on overall water level in the area of interest, and therefore that the assumption of a constant fairly high value for that variable would be adequate. Wave energy reaching seawalls is often restricted by water depth, and it may be adequate to evaluate an extreme sea level and then to assume a depth-limited wave height. Wave period can be considered as an independent variable, but for extreme conditions it is common to assume that wave period is dependent upon (offshore) wave height. Similarly, tide and surge can be considered independently, but it is more common to work with the total sea level (i.e. tide plus surge). Variables whose extreme values are either meaningless or not relevant in flood risk, for example wind / wave direction or rainfall / flow duration, are better treated as discrete ‘conditions’ within the event definition than as ‘variables’.

What types of joint probability analysis are there?
There are several variations in approach, depending on what assumptions can be made about the variables concerned and their relative importance in the assessment being undertaken. This report concentrates on two well established approaches, referred to as ‘the desk study approach’ and ‘the analytical approach’. The latter requires significantly more data and staff time, and the desk study approach is often adequate.

When is the analytical approach appropriate?
There is no minimum length of simultaneous sequential data required for use of the analytical approach, but to achieve a significant improvement in accuracy over the desk study approach, coupled with published extremes and dependence, typically needs at least a few years of data. These need not be measured data (for example, hindcast wave data are usually used) and need not be continuous in time, but they must provide a representative sample of records. The analytical approach requires specialist software and experienced staff, but offers more accurate estimation of dependence and extremes through the use of site-specific data, more complex dependence functions and more flexible output formats. Whether to use it is a business decision, often constrained by the cost or availability of time series data. Typically, the costs and benefits of adopting the analytical approach are comparable in scale with those involved in choosing to use a wave transformation model, as opposed to desk study modifications of offshore wave conditions. There is nothing explicit in the desk study approach to indicate when the analytical approach should be preferred, but sensitivity tests using different inputs to the desk study approach may help to inform the decision.

Can the methods be applied in complex areas?
Large or otherwise complex areas may be affected by several different flood risk variables, for example locally and distantly generated waves, flows in different rivers, sea level and/or rainfall. Although limited to two or three primary variables, both the desk study and analytical approaches can be applied in such areas, but care is needed in preparing the source data and in interpreting the analysis results. This usually involves a simplification of the problem, perhaps by excluding some source variables, or by treating them either as secondary or as conditional variables, or by the use of proxy variables from which multiple flood risk variables can later be re-generated.

1.3 The desk study approach to joint probability analysis

How is the desk study approach applied?

The desk study approach requires as input high and extreme values of each of two variables, together with a simple representation of the dependence between the two. For a number of alternative joint return periods and a number of alternative levels of dependence, the method provides a list of pre-computed combinations of the two variables with the required joint return period. As these pre-computed values are expressed in terms of the marginal (single variable) return periods for each of the two variables, the approach is quite general, being neither site-specific nor variable-specific.

What do the dependence maps show?

An essential part of any joint probability analysis is an estimate of the dependence between the variables involved. In the absence of data, previous reports or site-specific analysis, dependence can be difficult to estimate, and the consequences of a poor estimate can be significant. The dependence maps provide estimates of dependence between several different variable-pairs covering most of the UK, in a format which can be used directly in the desk study approach to joint probability analysis.

What steps are involved in calculations?

5. Derive high and extreme values for Variable 1.
6. Derive high and extreme values for Variable 2.
7. Estimate the dependence between the two, for example from the relevant dependence map.
8. Decide the joint return period(s) of interest.
9. Select the appropriate table(s) for this level of dependence and return period(s).
10. Convert the listed joint exceedence extremes from marginal return periods to actual values.
11. Modify for future climate change if necessary.

What results formats are available?

Joint exceedence extremes are the only direct output of the desk study approach. They can be derived for a number of return periods, locations and assumptions about future climate change. They incorporate knowledge of the marginal (single variable) extremes and the chosen simplified dependence model.

Is there a software version of the desk study approach?

The floppy disk at the back of this report carries the software tool which reproduces the desk study approach. The user enters the marginal extremes for each of two primary variables, the dependence between the two variables and the desired joint exceedence return periods. The Excel spreadsheet calculates the joint exceedence tables and curves.

1.4 The analytical approach to joint probability analysis

What extra would be done in undertaking a rigorous analysis?

Instead of marginal extreme values, records of the simultaneous occurrence of the variables selected for analysis are acquired, checked and prepared for further analysis. Project-specific statistical analysis is applied to these data using more complex dependence models than in the desk study approach. Instead of joint exceedence extremes, the analytical approach described in this report uses Monte Carlo simulation to produce thousands of years of joint data, effectively extrapolating the joint probability density to extreme values.

What is the Monte Carlo simulation approach?

The immediate output of the analytical approach described in this report is a very large sample of data in the same format as the source data, namely records of the simultaneous occurrence of the variables selected for analysis. These data reproduce the marginal distributions, the marginal extremes and the dependence behaviour of the original source data. This offers greater flexibility and precision in handling and interpretation of results than the desk study approach.

What input information is required?

The main input is usually in the form of a large number of records of simultaneously occurring values of the two primary variables of interest (e.g. sea level and wave height, or sea level and river flow). The records need not be in time sequence, but they should provide a representative (of whatever is intended to be represented) sample. Given that information on dependence and extremes is available in published form, based on analysis of longer periods of data, it is probably not worth embarking on the analytical approach with less than about three years of source records. Additional information on marginal extremes, perhaps derived from longer periods of non-simultaneous data, can be provided as additional input for use in the joint probability analysis.

What steps are involved in preparing the data?

1. Identify the sources of time series data on each variable and check that at least three years of simultaneous data are available.
2. Purchase, synthesise or re-commission the required data.
3. Decide how the records are to be defined and extracted, e.g. one per high tide or one per hydrograph duration, whether all records or just higher values are to be extracted, and which numbers are to be noted (e.g. a peak value or an average over a period of time).
4. Gather all records corresponding to the definition chosen, discarding incomplete records; check the validity of any apparent outliers (either in terms of marginal values or dependence).
5. Possibly subdivide into statistical populations, usually by wind (or wave) direction sector (as a proxy for different types of weather conditions).

What steps are involved in the analysis?

1. Fit statistical distributions to the top few percent of values of each primary variable.
2. Fit a statistical distribution to the relationship between wave periods and the top few percent of wave height (if waves involved).
3. Fit a statistical distribution to the dependence between the two primary variables.
4. Incorporate alternative marginal extremes if appropriate.
5. Synthesise a very large data sample, based on the above distributions plus the distributions of the source data below the thresholds used for extremes analysis.
6. Optionally, convert the simulated data to structure function(s), e.g. overtopping rate, river water level, inundation area, economic loss, failure etc, using an equation or hydraulic model.
7. Extract and interpret high and extreme values of interest.

What results formats are available?

1. Joint exceedence extremes.
2. Very large sample simulation of input variables.
3. Very large sample simulation of derived structure variables.
4. Statistical confidence limits.

Is program code for the analytical approach available?

In principle, the program code and example data files are freely available and will be released on request. Several organisations have copies. However, the programs are not suitable for widespread use, as there are a number of potential problems with their portability and correct usage, and with correct interpretation of the results produced. It is recommended that the programs are not taken up without training in their operation and interpretation, and suggested that occasional users adopt the desk study approach instead.

1.5 Interpretation of joint probability results

What are joint exceedence extremes?

Joint exceedence refers to the probability that a specified value of one variable will be exceeded at the same time as a specified value of a second variable. If that probability is small, it is usually expressed in terms of its return period. This is comparable with the approach used to define single-variable extremes, and is commonly used by UK coastal engineers to present joint probability results for use in design. Note that many different combinations of the two variables will have the same joint exceedence return period, and the results are often shown as a curve or as a tabulation of up to ten alternative combinations drawn from that curve.

How are joint exceedence extremes applied?

Typically, a number of alternative joint exceedence extremes, for a given return period, are used as input to other models, e.g. river model, wave transformation model, urban drainage model, overtopping equations, defence design calculations. However, it is important to try sufficient alternative inputs to be confident that a worst case response has been found during the modelling. Even then, the return period of that response is typically two or three times lower (potentially unsafe design) than the joint exceedence return period. In practice this is usually adequately mitigated by conservative assumptions implicit in preparation of the source data, but an explicit allowance is sometimes made for this difference in return periods.

What checks should be applied?

As with any calculations, results can be checked against previous studies and prior expectations. In prediction of single-variable or multi-variable extremes, there are some simple internal checks that should be made by counting back through the higher records in the source data. For sites with good long-term records of, for example, occurrences of severe overtopping or of out-of-bank flow, analysis results can be checked by comparing the predicted and actual frequency of occurrence of such events.

How are Monte Carlo simulations applied?

The approach described in this report produces a very large (1000s of years in size) sample of data containing discrete records of the source variables in the same format as the input records. These data can be presented as scatter plots of the source variables and can be converted to structure function(s) of interest, if known. Further extrapolation to extreme values is unnecessary, and instead the extreme values are found by counting back through the higher records in the simulation. Joint exceedence extremes can be extracted from the simulation if required.

What are the uncertainties?

1. Uncertainty (and possibly systematic error) in the source data, e.g. measured river flow or predicted wave height.
2. Statistical model uncertainty introduced when an imperfect statistical distribution (or assumed threshold) is fitted to sample data.
3. Parameter uncertainty occurs when statistical model parameters are estimated from a limited length of data.
4. Uncertainty arising from the use of the return periods of joint exceedence extremes as approximations to the return periods of the associated flood risks.
5. Uncertainty in the estimation of structure function(s) from source variables, e.g. by design curves or hydraulic modelling.

Can uncertainty be evaluated?

Parameter uncertainty, and its impact on extreme values, can be evaluated analytically. Statistical model uncertainty and the mismatch between joint exceedence and response return period can be investigated through sensitivity tests. Uncertainty and systematic error in the source data and in calculation of the structure function(s) can be estimated based on general experience. These different sources of uncertainty can be combined to produce estimates of accumulated uncertainty at different stages in the analysis procedures.

Could climate change affect dependence?

Any change in the magnitude of source variables, for example sea level rise due to climate change, would not necessarily cause any change in the dependence between the occurrence of high values and those of other source variables. However, if climate change caused a change in weather patterns, storm-tracking or storm frequency, it is possible that this would affect the dependence between different source variables.

1.6 Interpretation and use of Proudman (1997) extreme sea level predictions

What is the basis of the predictions?

The source data were measurements from forty UK A Class tide gauges up to 1993, expanded to forty years of data for the whole country by spatial interpolation and numerical modelling. Tide and surge were separated, analysed separately, and then re-combined using joint probability methods to provide extreme sea level predictions on a 36km grid for open coast locations around England, Wales and Scotland.

How are the results expressed for each grid point reported?

1. 1 year return period sea levels, relative to mean sea level in 1990.
2. Conversions from mean sea level to Ordnance Datum.
3. Trend in mean sea level observed over 40 years.
4. Elevations to be added to the 1 year sea level for higher return periods.

What steps are involved in applying the results?
If Proudman (1997) alone were used, the elevations for higher return periods would be added to the 1 year return period sea level, with interpolation between grid points, and adjustment of datum and for trend as required. However, it did not incorporate information from non A Class tide gauges, local knowledge and local studies. Its authors therefore recommend re-evaluation and replacement of their 1 year return period sea level, where there is better local data available, before adding the elevations for higher return periods.

1.7 Further and related work

What is the continuous simulation approach for complex structure variables?
This is not a true joint probability approach, but can be more accurate and appropriate for derivation of complex structure variables where there is a sufficient length of input data. The source data (for example river flow and sea level) are continuously fed into a process model (for example a river model) and the structure variable of interest (for example water surface elevation at a point in the river) is continuously monitored. The resulting distribution of the structure variable is then analysed as required.

What is the contouring approach for complex structure variables?
It may be desirable to use a Monte Carlo simulation approach to joint probability analysis and calculation of a subsequent structure variable, but impractical to process vast numbers of records into equivalent structure function values where that function is complex. This is the case, for example, in computing overall economic loss due to extreme environmental conditions. It is possible to calculate a smaller number (typically about thirty) of structure function values corresponding to a regular grid of high and extreme values of the two input variables. The complex structure function can then be computed as a contour function, against which all high records in the Monte Carlo simulation can be evaluated.

Is further information and advice available?
More detailed information is given in the accompanying dependence mapping report (FD2308/TR1) and in other technical reports referenced therein. There is no programme of ‘support’ as such, but the author would be pleased to answer any minor points of clarification that may arise, or to explore the costs involved in any training, advice or consultancy work associated with joint probability analysis.

Is further research work ongoing or proposed?

1. The analysis methods are currently being applied by HR Wallingford in flood risk studies within the Environment Agency’s Thames Estuary 2100 programme.
2. The methods have been developed further and applied by HR Wallingford within Defra / Environment Agency RASP and National Flood Risk research programmes.
3. Temporal dependence (short-term sequencing) within Monte Carlo simulation will be developed by HR Wallingford within the EU FLOOD*site* programme.
4. CEH Wallingford has proposed a thorough investigation of the impacts of future climate change on dependence, based on Hadley Centre data expected in 2005.
5. Proudman has proposed to update 1997 work on extreme sea levels, using several more years of data, and finer grid modelling in the southern North Sea and Solent.
6. CEH Wallingford has proposed outline methods for the collective risk caused by multiple rivers flooding simultaneously.

7. HR Wallingford has proposed to update its 1996 swell atlas, using several more years of data, and to provide a software tool for users.

Are joint probability methods used in other countries?
Methods for estimating the joint probability of occurrence of related flood risk variables are used elsewhere, but not as frequently in the UK. The importance given to joint probability methods in the UK stems from the need to prioritise spending between different flood and coastal defence schemes, based primarily on benefit /cost ratio. In other southern North Sea countries, joint probability plays a small part in informing the national strategy towards flood risk, but is not widely used. In the Netherlands, similar approaches are also used in assessment of the probability of multiple defence failures. In the USA, joint probability is recognised as a research topic, but is not widely used in flood risk studies. The dependence mapping done during this project is UK-specific, but the two analysis methods described in this report could be used elsewhere. The analytical approach has been used in studies in Korea, the Netherlands and the USA.

PART TWO: THE BEST PRACTICE GUIDE

2. INTRODUCTION

2.1 Background to the project

For several years MAFF (now Defra) has funded a programme of research on joint probability, looking at the dependence between variables and how best to quantify their combined impact on flood and coastal defences. Work focused primarily on its applications to waves and sea levels, (at HR Wallingford) and to tides and surges (at the Proudman Oceanographic Laboratory, elsewhere abbreviated to Proudman). Joint probability methods have also been applied to rainfall, surge and river flow (at CEH Wallingford), and to wind-sea and swell (at HR Wallingford).

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 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, and the other to the lack of published information on the dependence between variables.

Specialist joint probability analysis software named JOIN-SEA was developed by HR Wallingford and Lancaster University during the Defra-funded programme of research. It has been taken up by a small number of UK consultants. Proudman's published predictions of UK extreme sea levels are widely used in the industry. However, in both cases the subtleties of application have not always been appreciated outside the originating organisations, and in some instances they have not been applied to full advantage.

The present project was developed during 2001 within the Risk Evaluation and Understanding Uncertainty Theme of the Defra / Environment Agency joint research programme. It continues a gradual programme of dissemination and appropriate take-up of joint probability methods in flood and coastal defence design and assessment. 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.

2.2 The objectives of the project

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 Oceanographic 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.

3. WHEN TO USE JOINT PROBABILITY METHODS

3.1 Introduction

3.1.1 When is joint probability relevant?

There are several situations in flood and coastal defence in which multiple variables combine to produce a response of interest. Examples are waves and sea levels combining to cause overtopping or damage to seawalls, river flow and downstream sea level combining to cause river flooding, and swell-sea combining with storm action to cause a more severe sea state. An understanding of the dependence between the variables concerned would permit a more accurate estimate of their combined probability of occurrence, therefore greater confidence in assessment of any associated risk, and potentially more efficient design of defences. However, if one variable is far more important than the other(s), or if complete independence or complete dependence would be a reasonable assumption, then joint probability analysis may not be helpful. Conversely, in complex situations, it may be necessary to separate the source data into populations, or to consider more than one variable-pair or more than one type of outcome.

As with most aspects of most flood risk studies, preliminary checks should be made before detailing the approach to be followed. These will assess the importance and availability of data on the source variables of potential interest within the study area, and how to handle and analyse those variables most effectively, where best to evaluate, and how to apply them elsewhere in the study.

3.1.2 Joint probability concepts

Joint probability refers to the chance of two or more conditions occurring at the same time. For this situation to be of interest, some overall outcome of interest should depend on the combined occurrence of these conditions. For this situation to be worthy of detailed study, the dependence between the two or more conditions should be non-trivial, i.e. neither independent nor fully dependent. Environmental variables such as rainfall, sea level, wave height and river flow are therefore suitable candidates for joint probability analysis of flood and other risks.

The joint probability analysis results usually take one of three forms.

1. Joint exceedence contours, i.e. lines of equal probability (sometimes expressed as return period) of the source variables simultaneously exceeding values indicated by any point on the contour.
2. A long time series of the relevant source variables, based on measurements or hindcasts.
3. A very long-term simulation of the joint probability density of the relevant source variables.

These results can then be applied as input to hydraulic models (e.g. numerical river model) or equations (e.g. mean overtopping rate) intended to predict responses or risks of interest.

3.1.3 The level of study

Preliminary consideration of the sources of flood risk, and of the format of results or type of decisions to be made, will establish whether or not there is a need for a joint probability assessment. A desk study approach to joint probability analysis is often adequate, and can probably be considered as the default option, in which the analyst's time and discretion are kept to a minimum. An analytical approach offers potentially greater accuracy in determining flood probabilities, and greater flexibility in application of the results, but requires more input data, and greater time and understanding from the analyst. It may be appropriate in larger studies, if greater accuracy and/or flexibility is required, and if data, time and expertise are available.

3.1.4 Benefits

Joint probability analysis takes account of the dependence between input variables, as well as the distributions and extremes of the individual variables. It can increase the accuracy of impact / failure probability estimation, whether achieved via 'design sea states', for example in the form of joint exceedence extremes, or directly via Monte Carlo simulation of structure variables. It uses the input (loading) information more effectively.

3.2 Coastal sites

The most common application of joint probability methods in coastal engineering is to waves and sea levels occurring near high tide for assessment of flood risk at sea defences. It would be possible to perform the analysis in terms of surge, either instead of or in addition to sea level, as this gives a purer indication of meteorological dependence, but this is not usually done as the results are harder to apply in flood risk calculations.

Analysis can be applied to either offshore waves and/or nearshore waves. Offshore waves give a good indication of the meteorological dependence with high sea level, which is usually indicative of dependence over the surrounding area. Nearshore waves, although perhaps more directly relevant to local flood risk, may be affected by hydraulic dependence (e.g. depth-limited wave breaking) introduced in the nearshore zone. Consequently, an analysis based on nearshore wave conditions may not be helpful in estimating corresponding conditions elsewhere in the vicinity.

Specialist analysis of extreme sea levels may involve the separate analysis of tide and surge as a joint probability application, before reporting in terms of sea level (i.e. tide plus surge). Around the UK, dependence between high surge and high tide, if any, tends to be negative, i.e. lower overall extreme sea levels than would be the case for positive dependence.

Swell wave conditions are not well represented in most wave climate and extremes predictions done in UK coastal engineering studies. At sites exposed to Atlantic wave conditions, i.e. south-west England and north-west Scotland, swell can be important in estimating the magnitude of coastal responses such as overtopping rate. For these exposed sites, it may be helpful to consider swell alone, and combinations of wind-sea and swell, as additional design conditions.

Joint probability analysis has been demonstrated using measured data on wave height and current speed, and could be important in estimating the proportion of the time that the seabed may be mobile. Alternatively, the dependence between the two could be inferred as being similar to that between wave height and sea level (as peak current speed at a particular location is approximately proportional to peak sea level at the same location).

3.3 River sites

The most common application of joint probability methods in river engineering is to upstream river flow and downstream sea level occurring near high tide for assessment of flood risk at river defences. It would be possible to perform the analysis in terms of surge, either instead of or in addition to sea level, as this gives a purer indication of meteorological dependence, but this is not usually done as the results are harder to apply in flood risk calculations.

A growing area of interest and concern is the joint probability of drainage systems being affected by high river levels. This is becoming more relevant as nearly all stormwater discharges are now consented and require flow attenuation. This results in storage units being built (often near the river) and critical storm durations becoming longer, with the resulting risk of coincidence of river flow peaks with drainage discharge affecting the ability of the drainage system to operate effectively.

Unlike coastal engineering, the two source variables are not applied at the same location (river flow is upstream and sea level is downstream). They may not be applied at the same time (each takes time to propagate along the length of the river). River flows in different rivers may combine to affect the flood risk downstream of a confluence. Hence, temporally and spatially separated dependence and event definition assume greater importance in river engineering than in coastal engineering.

Joint probability analysis of the likelihood of intense short duration rainfall coming soon after heavy long duration rainfall (as a proxy for antecedent wetness) might be of some value in catchment storage capacity and runoff calculations.

3.4 Estuary sites

Three-variable joint probability analysis may be needed in estuary locations where sea level, waves and river flow are all important. However, careful preliminary analysis can often reduce the analysis to two primary variables, by demonstrating that either river flow or waves is of relatively minor importance and therefore that it need not be analysed in detail. The three-variable approach is described in Defra / Environment Agency (2003) and in FD2308/TR1. Data preparation and interpretation of the results require particular care when three variables are analysed.

Whether or not river flow has a significant impact on water level in an estuary, it will be important where there is a barrage, either to impound water permanently, or for occasional closure to protect against high sea level. For design of the barrage it may be necessary to consider joint wave and sea level loading on the seaward side and river flow and water level on the landward side.

As at coastal sites, the joint probability of wave height and current speed could be important in estimating the proportion of the time that the seabed may be mobile, and hence prediction of morphological changes.

3.5 Urban sites

Depending on their locations, urban sites may be potentially vulnerable to inundation directly from river or coastal flooding. The same locations may also be vulnerable to localised and more frequent flooding caused by drainage systems being temporarily blocked by a high river or sea level. Joint probability analysis of high intensity rainfall (the source variable for urban drainage) and high river or sea level would assist in estimating the probability of occurrence of this type of flooding.

3.6 Larger area studies

There is an additional complication in applying joint probability methods to assess flood risks across an area, as opposed to a single site. This is the issue of spatial consistency across a large area, or in an area protected by multiple defence lengths / types, or in an area where there is more than one flood mechanism. It would be possible to apply joint probability methods to each defence length or flood mechanism in isolation, but this would not take proper account of the individual impacts combining to cause a potentially greater overall impact.

Methods are available to improve spatial consistency between events and/or outcomes, even if individual variables are not constant across an area. The easiest approach is to reduce the joint probability analysis to just two or three primary variables (for example, wind speed over an area and water level at just one location) from which other variables or values at other locations can be estimated. The other approach is much more complicated, involving a series of separate local analyses, the overall results from which cascade on into gradually larger area analyses.

4. DEPENDENCE ANALYSIS

4.1 Introduction

Dependence indicates the extent to which the value or condition of one variable can be determined solely from knowledge of the value or condition of other potentially dependent variables. (The variables may be spatially or temporally separated, but the extent of separation has to be specified).

If two variables are independent, then the value of one has no effect on the probability distribution of the other. The value shown on one die is independent of the value shown on another die. The weather in London is independent of the weather in New York (apart from both having their worst weather in the winter).

If the two variables are negatively dependent, and one takes a high value, then the other is more likely to take a low value than would otherwise be the case. For example, rainfall and sunshine tend not to occur at the same time at the same place, and so exhibit partial negative dependence.

If two variables are positively dependent, and one takes a high value, then the other is more likely to take a high value than would otherwise be the case. For example, high wind speed offshore and high wave height at the coast tend to occur together and so exhibit partial positive dependence.

The statements above are qualitatively obvious, but quantification of dependence can be difficult. Consider the hypothetical situation of river defences being designed for the water levels caused by the simultaneous occurrence of a 200 year return period upstream river flow and a 200 year return period downstream sea level. Whilst this may be an acceptable design approach, it is hard to say whether such an event would be expected about once every two hundred years, once every million years, or whether it could never occur. This degree of uncertainty then propagates through to any attempt to estimate the standard of service of an existing defence or the economic benefit of a new defence. Information on dependence, even if accompanied by a wide margin of uncertainty, would be helpful in refining flood risk calculations.

4.2 Methods and definitions

Different approaches and different measures of dependence (FD2308/TR1) were developed separately at HR Wallingford, Proudman and CEH Wallingford to meet the requirements of the particular data types being analysed. The dependence results for each variable-pair are shown in FD2308/TR1 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 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. During the industry consultation, and as the project progressed, it became clear that this approach, alone, 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 was developed, its use would have involved a loss of precision.

The ‘desk study approach’ developed during this project is no longer limited either to a particular dependence measure or to pre-computed discrete 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’ (FD2308/TR1). It also provides more precise information needed as input to the alternative approaches described in the present report.

4.3 Data sources

As far as practical, consistent sources of data were used throughout the analyses. For most variables, this involved use of at least ten years of measured data for at least twenty locations around Britain. One exception was wave data, where long periods of measured data are rare, and instead data were taken from a numerical model covering the seas around Britain. Full details are given in Chapter 2 of FD2308/TR1 but a brief summary is given here.

Tide and surge: 24 gauges, average duration 30 years, most recent year 2001

Waves and swell: 21 points chosen from a 25km grid, April 1990 to March 2002

River flow: 130 measurement stations, 1963-2001

Daily precipitation: 44 gauges, 1965-1997 on east coast, 1963-2001 on other coasts

Hourly rainfall: 14 gauges in England and Wales, mixed durations 1 to 30 years

Not all data were used in all dependence analyses, but in each analysis, the maximum length of simultaneous data on the two variables involved was used. Pre-processing was applied differently to different variable-pairs, depending on the event definition used. This involved de-clustering of otherwise dependent records, selection of peaks over threshold values, division of waves into wind-sea and swell, division of records into seasons or direction sectors, and/or division of sea levels into tide and surge.

Additional time series data on surge, wind speed and precipitation were obtained from a Hadley Centre regional climate model for use in assessment of the potential impact of future climate change upon dependence.

Details of the analysis and pre-processing methods used are given in Chapter 3 of FD2308/TR1.

4.4 Results

Details of the analysis methods and of the significance of the results for each variable-pair are given in Chapters 3 and 4, respectively, of FD2308/TR1. The results are intended to provide direct input to desk study applications of joint probability analysis. They could also provide direct input, or more likely background information and understanding, for specialists using their own data and more rigorous analysis methods.

To avoid presenting too much information, this report reproduces only those dependence maps most likely to be used in site-specific flood risk studies. This section also raises issues specific to each variable-pair, for example direction dependence, seasonal variability, time-delayed dependence, event definition and duration.

4.4.1 Wave height and sea level; wave height and surge

Dependence results involving wave height are presented in Section 4.2 of FD2308/TR1. Dependence maps and tables are given for four variations of wave height and sea level:

- wave height and sea level (all directions combined, i.e. overall climate)
- wave height and sea level (for the wave direction sector in which dependence is higher)
- wave height and sea level (for the wave direction sector in which dependence is lower)
- wave height and surge (all directions combined).

The maps include all best estimate correlation coefficient (ρ) results for 23 locations around the UK. Sea areas close to the coast are shaded using four or five different colours corresponding to the bands of dependence used in the 'simplified method' of joint probability analysis described in FD2308/TR1. Some small areas, for example the Humber, the Wash and the Severn estuary upstream of Cardiff, are deliberately uncoloured to draw attention to the fact that the open coast results may not be applicable in those areas. The tables contain the same best estimate results, but also confidence intervals, and a second series of results and confidence intervals for a different analysis parameter setting, to give an impression of the uncertainties involved.

The dependence map for wave height and sea level for all directions combined is reproduced here as Figure 1, as this is likely to be the most commonly used page of results amongst the various maps and tables. It may provide sufficient information for occasional users working on open coast locations, and for those wishing to familiarise themselves with the ideas.

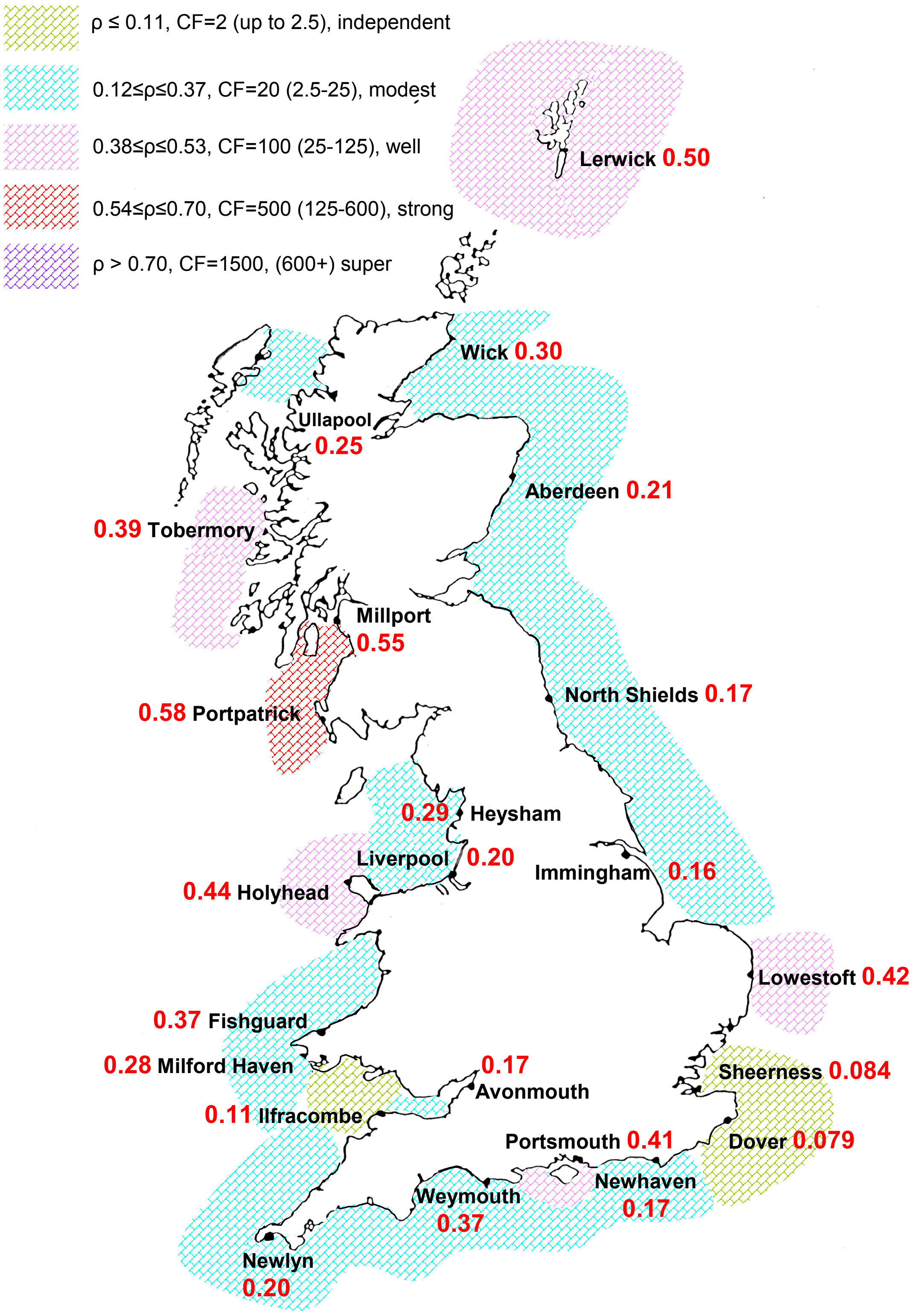


Figure 1 Summary dependence information for wave height and sea level

For coastal studies where wave direction is important, different correlation coefficients for different directions can be estimated from more extensive results in FD2308/TR1. For coastal studies involving estimation of the uncertainties involved in flood risk, it may be helpful to consider some of the alternative estimates of dependence given in FD2308/TR1.

The dependence maps offer no information on wave period. A convenient method for estimation of wave periods associated with storm wave conditions, commonly used in coastal engineering, is to assume that they are highly dependent upon wave height. A representative wave steepness ($2\pi H_s/gT_m^2$) is determined from the highest few percent of the wave conditions, which is then applied to any extreme wave heights predicted. If wave steepness cannot be estimated from data, 0.05 would be a realistic value to use for storm wave conditions in exposed offshore locations around the UK.

4.4.2 Tide and surge

The dependence between tide and surge is an integral part of the extreme sea level analyses undertaken by Proudman over the last ten years, but is not usually considered explicitly in coastal engineering studies. Dependence mapping results, not reproduced here, are given in Section 4.4 of FD2308/TR1 in terms of the degree of negative dependence between high sea levels and high surges. Negative dependence is greatest between Brighton and the Humber, in the Bristol Channel and in the Irish Sea. Negative dependence means that extreme sea levels are less severe than would be the case if there was positive dependence or even independence.

4.4.3 River flow and surge; rainfall and surge

Dependence results for river engineering are presented in Section 4.3 of FD2308/TR1. Event definition and duration were not so obvious for rainfall and flow as for the marine variables but effectively one day was taken as the record interval / length for the river engineering variables. Dependence maps and (for river flow and surge) tables are given for three pairings of rainfall, river flow and surge:

- river flow and daily maximum surge
- river flow and daily maximum surge occurring at high tide
- precipitation and daily maximum surge in catchments draining to the British east coast
- precipitation and daily maximum surge in catchments draining to the British south and west coasts.

Additional analyses and dependence maps are given in specialist report FD2308/TR3 for:

- precipitation and river flow
- auto-correlation analysis for river flow, for rainfall and for surge (giving an indication of event duration)
- inter-station dependence for river flow, for rainfall and for surge (giving an idea of the area affected by an event)
- winter / summer variations for all three inter-station dependences and for all three variable-pairs
- river flow and daily maximum surge, for various time lags
- precipitation and daily maximum surge, for various time lags.

The maps show pairs of variables / stations joined by lines where dependence exceeds a particular value (each figure contains two or three maps, corresponding to two or three alternative dependence thresholds). The tables contain the best estimate values of dependence measure χ, plus confidence intervals, and a critical value against which the significance of the χ value can be judged. It should be noted that χ (and ρ used for wave height and sea level) measures dependence in a general sense, including dependence arising from seasonality. The critical value for the significance of χ effectively tests whether there is more dependence than could be explained purely by seasonality.

The dependence information for river flow and surge is summarised in Figure 2, in a form that is likely to be of the most immediate use for derivation of combinations of sea level and river flow for use in modelling of individual rivers. Flow stations are shown by their four or five figure reference number (see FD2308/TR1 or FD2308/TR3 for station details). Each flow station is given one of five colours, approximately corresponding to those used in Figure 1, to represent the level of dependence between river flow and surge. The colour used in each case is based on the average χ dependence measure for daily averaged river flow at each flow station, analysed against daily maximum surge at the nearest (usually) two tide gauge stations. Dependence is higher on the west coast than on the east coast and, throughout the country, is higher for short steep rivers with rapid response than for long flat ones.

Figure 2 may provide sufficient information for occasional users, and for those wishing to familiarise themselves with the ideas, but it should be noted that neighbouring catchments can show very different dependence behaviour because of different response times.

For river studies where continuing heavy rainfall and/or a time lag between variables is important, the additional relevant dependence information can be extracted from more extensive results in FD2308/TR1 and FD2308/TR3. For river studies involving estimation of the uncertainties involved in flood risk, it may be helpful to carry out sensitivity tests using the confidence limits for dependence given in FD2308/TR1.

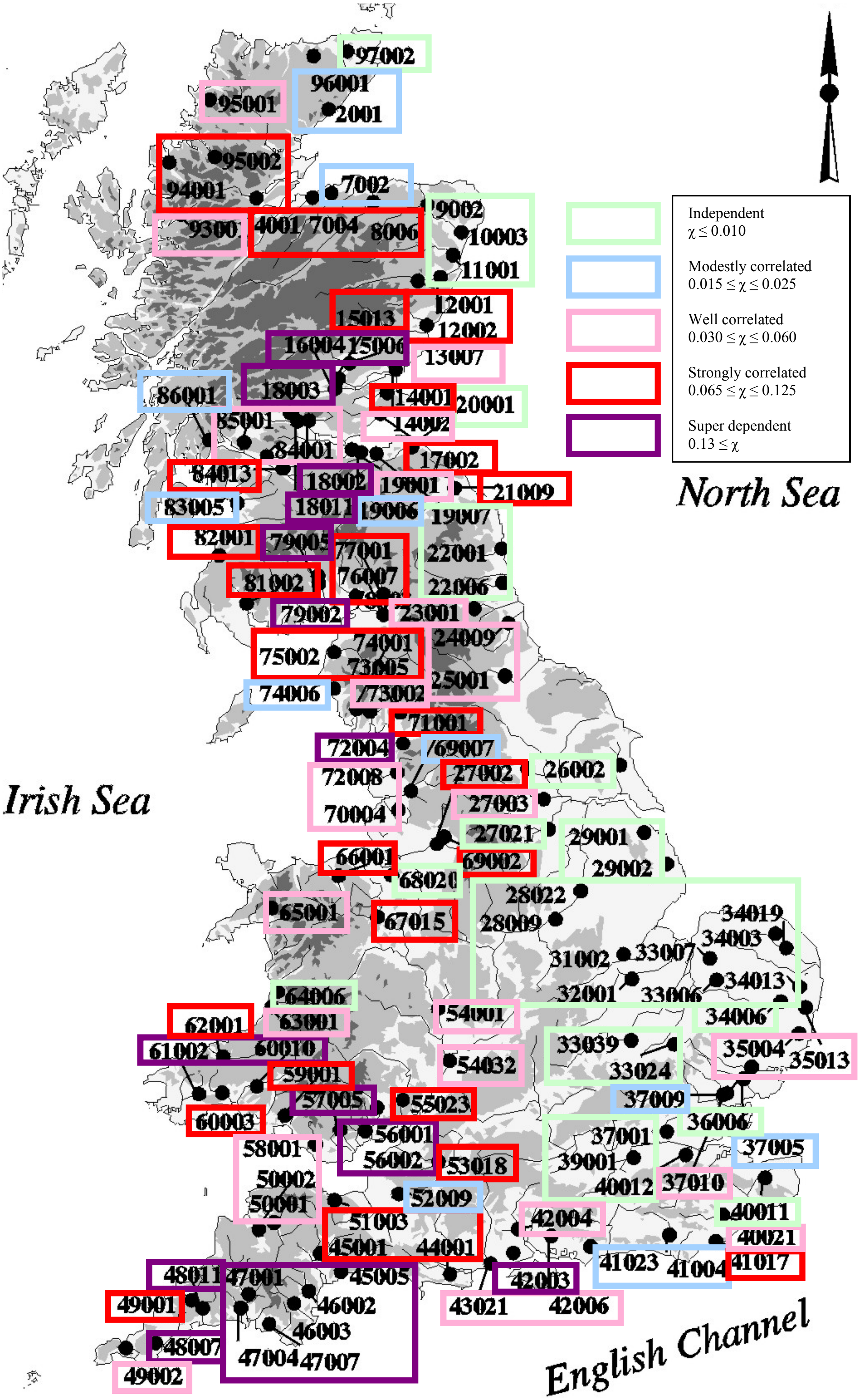


Figure 2 **Summary dependence information for river flow and surge**

4.4.4 High intensity rainfall and sea level

Dependence results for short duration rainfall and sea level are presented in Section 4.2 of FD2308/TR1. The dependence map for two-hourly rainfall and sea level in England and Wales is reproduced here as Figure 3. The same five-colour scheme is used as required for the 'simplified method' of joint probability analysis described in FD2308/TR1, but in this case only the two lowest levels of dependence are seen. On the east coast of England, although only three station-pairs were analysed, the dependence between short duration rainfall and sea level is consistently minimal (but not zero). This probably indicates that high values of hourly rainfall and sea level are uncorrelated, except that they both tend to occur only in the winter half of the year. On the south and west coasts of England and Wales, the dependence is consistently characterised as 'modest', indicating a low meteorological correlation between hourly rainfall and sea level. The corresponding table in FD2308/TR1 contains correlation coefficient results, plus confidence intervals.

4.4.5 Wind-sea and swell

Dependence results for wind-sea and swell are presented in Section 4.2 of FD2308/TR1. The dependence map for wave height and sea level for all directions combined is reproduced here as Figure 4. Sea areas close to the coast are shaded using five different colours corresponding to the bands of dependence used in the 'simplified method' of joint probability analysis described in FD2308/TR1. Some small areas, for example the Humber, the Wash and the Severn estuary upstream of Cardiff, are deliberately uncoloured to draw attention to the fact that the open coast results may not be applicable in those areas. The corresponding table in FD2308/TR1 contains correlation coefficient results, plus confidence intervals, for Scottish locations (results for England and Wales having been computed previously using a slightly different method).

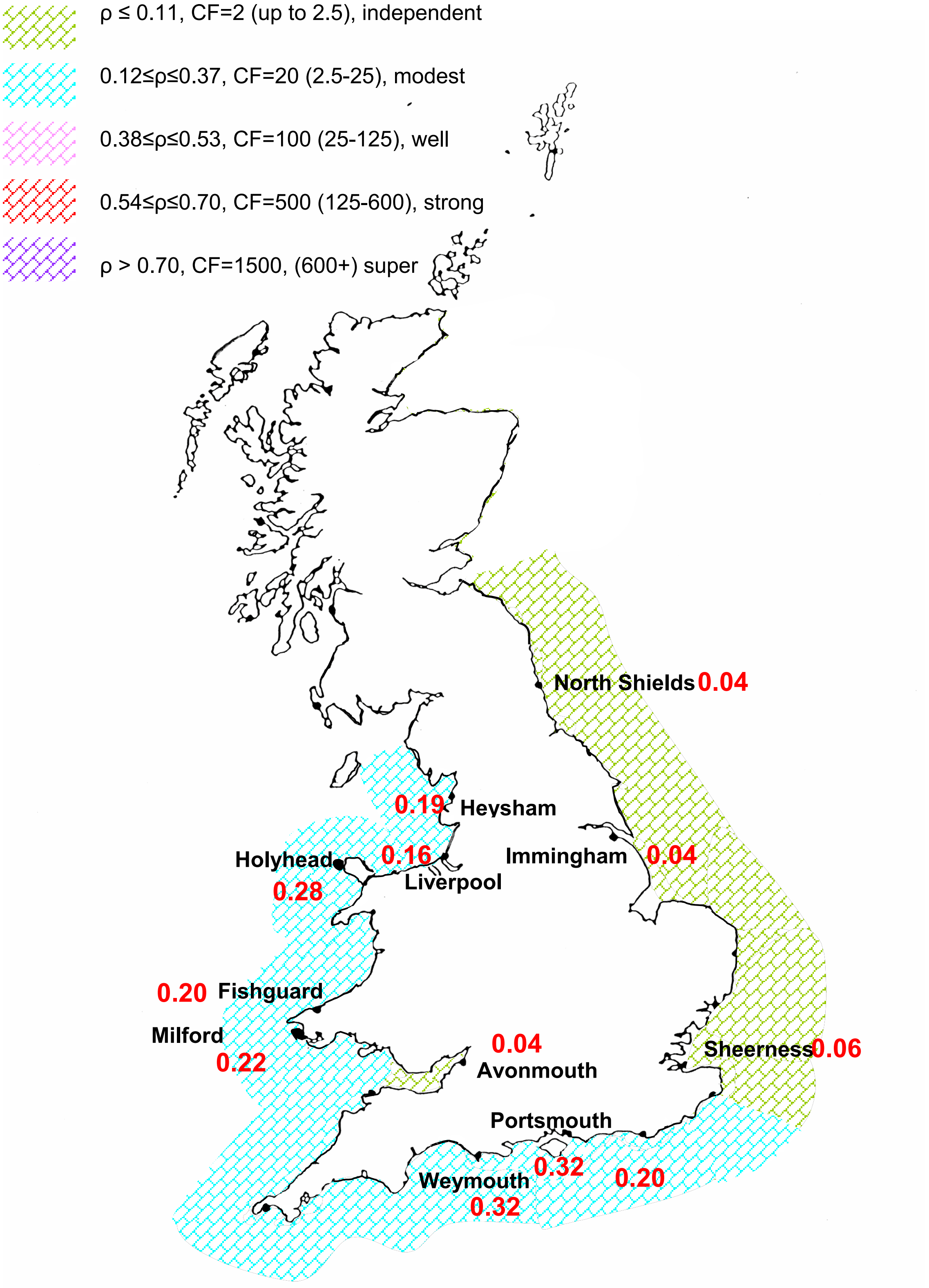


Figure 3 Summary dependence information for rainfall and sea level

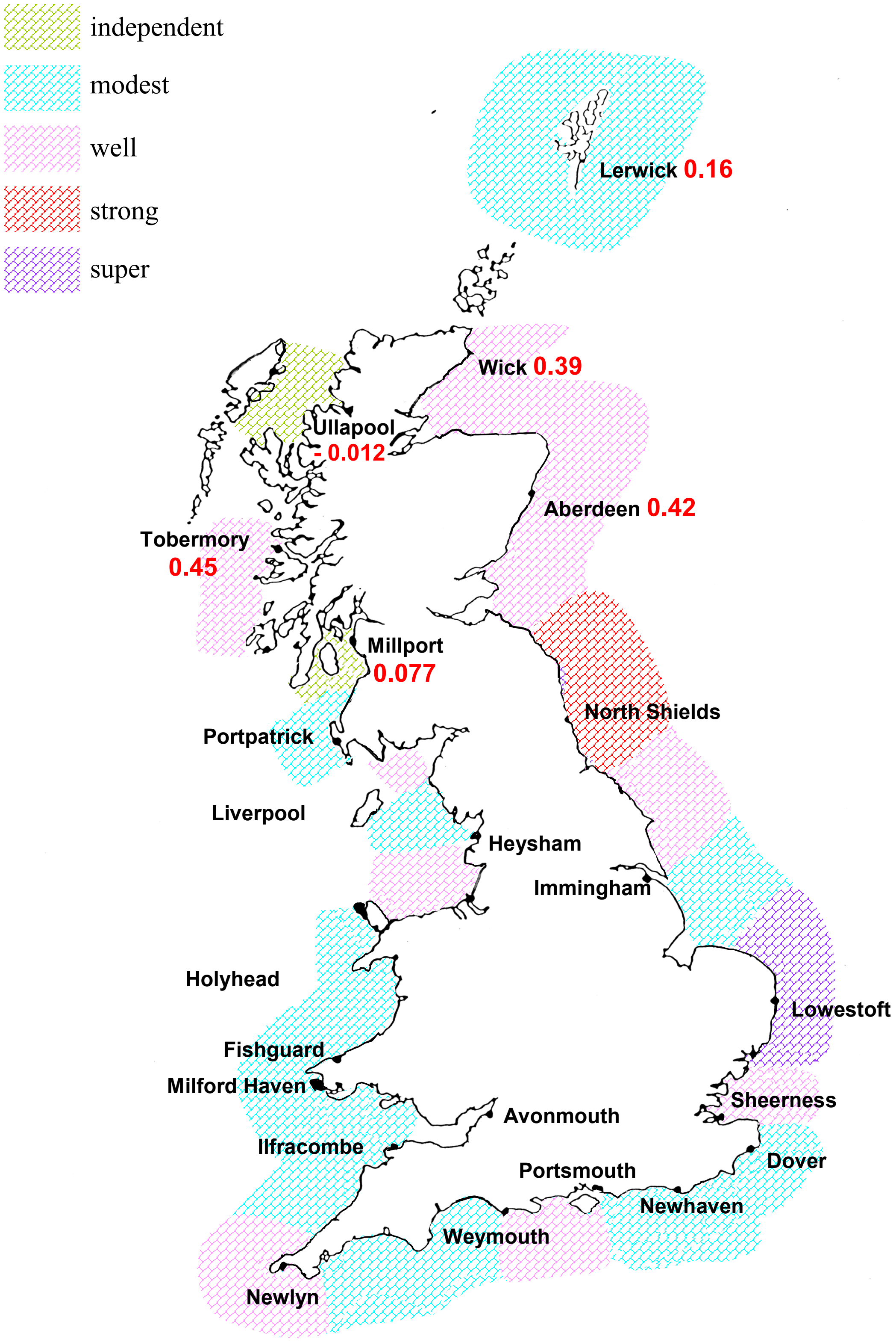


Figure 4 **Summary dependence information for wind-sea and swell**

Wave direction is not usually important for studies involving swell, because swell tends to come from a relatively narrow band of directions at most locations.

The dependence maps offer no information on swell wave period, but a typical value could be taken from the source of information used for estimating swell wave height. In some applications at some locations, if swell is critical, it may be necessary to consider different swell wave periods, for example corresponding to 'young swell' and 'old swell'. HR Wallingford (1997) offers more information on different swell wave periods, and how to combine them with storm wave predictions around England and Wales.

4.4.6 The potential impact of climate change

As with climate change impacts on individual source variables, climate change may affect dependence between them, perhaps through changes in the tracking of storms around the UK. Section 4.5 and Appendix 5 of FD2308/TR1 describe investigation of this possibility through the analysis of projections from two numerical climate models, and the comparison of results between present-day and future 30-year time slices. Even if absolute levels of dependence are less reliable than those derived from measurements, the consistency between present and future time slices means that differences in dependence between the two should be meaningful.

The German ECHAM4 model, which had approximately one-degree spatial resolution, provided wind velocity and atmospheric pressure series, used as input to numerical tide and wave models, which produced surge, sea level and wave time series. For five locations around Britain, dependence analysis was applied to wave height and sea level. The levels of dependence at each of the five locations, and the differences between the five locations, are quite similar to those determined during this project and giving confidence in the computational methods used. There was no significant difference in dependence between the present-day and future time slices at any of the five locations.

The Hadley Centre 50km grid HadRM3 model, in combination with a shelf-seas tidal model, provided information relevant to both river and coastal flood risk, for 23 locations around Britain. The dependence between high surge and high precipitation was used to represent river flood risk, and the dependence between high surge and high wind speed to represent coastal flood risk. The levels and patterns of dependence around the country are quite similar to those obtained from measurements, giving confidence in the data source.

The results consistently suggested a significant increase in dependence under the Medium-High Emissions Scenario both for river and for coastal flood risk. The surge and precipitation analysis indicated approximately a doubling in the likelihood of high surge and high precipitation occurring together on the south and west coasts of the UK, and on the east coast of Scotland, due to increasing dependence. The surge and wind speed analysis indicated about a 50% increase in the likelihood of high surge and high wind speed occurring simultaneously north of a line between Weymouth to Lowestoft, due to increasing dependence.

Results derived from the Hadley Centre model were used to form the draft precautionary allowance below for the possible impact of future climate change on the dependence between variables relevant to flood risk.

> *For river flood risk calculations involving a high sea level coupled with a high river flow, on the south and west coasts of the UK and on the east coast of Scotland, test the sensitivity of the assessment to a 100%*

> *increase in flood probability caused by an increase in dependence between the two. For coastal flood risk calculations involving a high sea level coupled with high waves, north of a line between Weymouth and Lowestoft, test the sensitivity of the assessment to a 50% increase in flood probability caused by an increase in dependence between the two.*

The level of confidence in the appropriateness of this allowance is low (although not a great deal less than in the precautionary allowances for the impact of future climate change on river flows and wave heights currently recommended by Defra, 2003). As further work has been proposed into the effects of future climate change on dependence, it is recommended that adoption of this new allowance be deferred until the results of the new work become available.

5. DESK STUDY APPROACH TO JOINT PROBABILITY ANALYSIS

5.1 The analysis method

5.1.1 Theory

The 'simplified method' for joint probability analysis is described in Section 3.5 of FD2308/TR1. It involves the use of tables of combinations of two variables, expressed in terms of their marginal return periods, pre-computed for a number of example joint return periods, levels of dependence and numbers of records per year. The method is described in mathematical terms, and contrasted with the ρ and χ dependence models, in Appendix 4 of FD2308/TR1.

The basis of the desk study approach is the same but, as a computer assisted version is available, it is not limited to example pre-computed values or to the 'correlation factor' statistical model underlying the simplified method. Most of this chapter is written assuming manual use of the pre-computed tables, but Section 5.7 describes the additional flexibility and resolution offered by the desk study software tool for joint probability analysis. All the points raised in Sections 5.1.2-5.1.5 relate to well established aspects of joint probability analysis, and all are potentially relevant to most applications of the desk study approach. The point raised in Section 5.1.6 is not in common use, but may provide a method of comparison with other joint density probability approaches.

5.1.2 Data requirements

The desk study approach requires extremes of the first variable, extremes of the second variable and a single-parameter representation of dependence between the two. Time series and climate tables are not required (although they may well have been used in the course of deriving the extreme values).

5.1.3 Steps in application

Detail will vary according to the type of study, for example whether coastal, river or urban, and the size of the area considered, but broadly the following steps will be usually be involved.

Select the variables and any conditions attached to those variables
Select the two relevant variables, e.g. waves and sea level, or river flow and surge. Determine any conditions, such as a requirement to be in a particular season, a particular wind / wave direction sector, or a particular rainfall / flow duration.

Decide how the variables will be represented
The desk study approach is fairly flexible about how variables are represented, the only condition being that the representation must be capable of interpolation and extrapolation to other high values. So, for example, waves could be represented by a height, period and direction, and river flow could be represented by a flow hydrograph.

Obtain extreme values for each variable
The desk study approach requires that high and extreme values of each source variable are provided from elsewhere, for a wide range of return period, sometimes down to quite commonly occurring conditions (tens of times per year). For waves or river flows, a separate modelling study might be needed to derive these values. For sea levels, a brief desk study of published values will usually be enough. Some sources of published extreme values of flood risk variables are listed and briefly described in Section 1.4.4 of FD2308/TR1.

Obtain the level of dependence between the variables
Previous studies and/or the results of the present project will usually provide a reasonable estimate of dependence, perhaps helped by an understanding of the physical aspects of the situation being studied. Correlation can be expressed in precise terms using one of the dependence measures used in this project, i.e. ρ or χ, or in general terms as being within one of the five bands of dependence also used in this project.

Apply the desk study approach
The desk study approach is based on a large number of tables, each consisting of about ten pre-computed combinations of the two source variables, each expressed in terms of its marginal return periods. The first part of the task is to select the most appropriate of the pre-computed tables, depending upon the joint exceedence return periods required, the degree of correlation between the two variables and the number of records per year. The second part of the task involves conversion of the marginal return period(s) in the selected table(s) into actual values of the two source variables, usually by interpolation between values known for specified marginal return periods.

Apply the results of the desk study approach
The results will be in the form of one or more lists of combinations of the two source variables, with separate lists for each return period if relevant, and separate lists for each direction sector, season or duration, if relevant. If the variables and their representations were well chosen, it should be possible to convert these lists into equivalent lists of response or risk variables, using hydraulic modelling, equations or design curves.

5.1.4 Additional options during analysis

Division of the data set into separate populations
It may be helpful to separate the source records into sub-sets, either to isolate groups of data with different statistical properties, and/or to accommodate additional secondary variables which may be important in subsequent calculations. Offshore marine data series are often separated into wind/wave direction sectors, or into storm waves and swell, both to isolate the different statistical populations, and because wave direction and/or steepness may be important in subsequent hydraulic modelling. Fluvial events might be separated by rainfall or hydrograph duration if the associated flood risk depends on duration as well as intensity. Marine or fluvial data could be separated into winter and summer conditions, if the statistical properties vary seasonally, or if the risks being considered occur only in one season. Where data sets are divided in this way, each sub-set should be analysed separately. For each sub-set, the number of records per year and the marginal extremes should be evaluated conditional upon the variable(s) meeting the defining criteria associated with that particular sub-set.

Incorporation of secondary variables
Variables previously considered to have been of secondary importance can be appended to joint exceedence results, based on the assumption that they are dependent upon one or both of the primary variables. A common example of this, where wave height is a primary variable, is the addition of wave period, usually based on the assumption of a standard wave height to length ratio determined from local wave measurements or predictions. Other examples are wind set-up and wave set-up based on wind speed or wave height, and sea current based on sea level.

5.1.5 Practical points during application of results

Use of multi-valued joint exceedence extremes
A feature of the desk study approach is that it delivers multiple combinations of the two source variables for any given return period (and climate scenario). It is necessary to test enough of these combinations to ensure that a worst case (i.e. maximum risk) has been found. This worst case combination may vary between locations, between different types of flood risk and, where applicable, would be affected by depth-limitation of wave height.

Meaning of joint exceedence return period
If a particular risk increases roughly in proportion to the value of one source variable only, then the return period of the risk and the return period of the source variable will be the same. This is not the case with joint exceedence extremes, which refer to the probability of two or more source variables each exceeding given thresholds at the same time (or sometimes with a specified time lag). This return period is (at best) only an approximation to the return period of any structure function (response or risk) derived from it. This point is explained and discussed in Section 3.5.3 of FD2308/TR1. If there is no implicit conservatism in the way that the source data are prepared or in the level of dependence assumed, then it may be wise to work with a joint exceedence return period twice as large as that of the response to be evaluated. In practice, an adequate level of conservatism is usually achieved through the implicit assumption that extreme wave conditions will necessarily occur at high tide or that peak flows will necessarily coincide with peak sea level, even though the record durations may be significantly longer than the duration of the peak conditions.

Allowance for future climate change
Precautionary climate change allowances are easy to apply if they can be expressed in similar terms and parameters to those used in the desk study analysis. Allowance for sea level is usually expressed in terms of an increase in level (e.g. add 0.4m) which can be simply applied to all sea levels in the joint probability results. Allowances for other source variables are most often expressed as factors (e.g. add 20%) which can be simply applied to all affected variables in the joint probability results. The possible allowance for future change in dependence suggested in Section 4.4.6 is expressed in the form of a change in flood probability. This would most conveniently be applied as a factor on the design return period used (or on the probability of occurrence for any particular joint probability loading condition considered).

Handling of discontinuous structure functions
Discontinuity in the flood risk function may be caused, for example, by nearshore depth limitation of offshore wave height, change from in-bank to out-of-bank flow, or a

change in flood mechanism (e.g. wave overtopping, to weiring flow, to breaching). This does not cause any particular problem, provided the discontinuities are known, because each joint exceedence condition is transformed separately to appropriate structure function(s) before identification of worst case(s).

Estimation of joint probability density from joint exceedence values
Joint probability *density* is capable of providing a more precise estimate of the probability of certain flood risks, dependent upon two or more variables, than are results expressed in terms of joint exceedence. Unfortunately, it is not easy to evaluate joint density as part of the desk study approach (except where the two primary variables are independent of each other). To a limited extent, joint probability density can be re-constructed from joint exceedence values, as described in Section 3.6.2 of FD2308/TR1. However, this is rarely done in practice, as joint probability density is more conveniently and precisely available, up to extreme values, from the analytical approach described in Chapter 6.

5.2 Case study involving waves and sea levels

5.2.1 The hypothetical coastal situation

Consider a typical small coastal engineering study, on the north-east coast of England at the north end of a bay. It has some protection from the largest open ocean waves from the north and north-east, provided by a headland, but is fully exposed to waves from the south-east. The existing seawall, fronted by a shallow foreshore, is potentially subject to damage under attack by large waves and high sea levels, and is to be tested for standard of service based on an acceptable overtopping rate criterion.

5.2.2 Select the variables and any conditions attached to those variables

Although surge and swell may be of some interest, waves and sea level would be the normal variables to select for joint probability analysis in this situation. It is necessary to consider (at least) two classes of wave condition. Broadly northerly waves would have larger heights offshore, and greater correlation with high sea levels, but would be reduced by nearshore processes and the headland, as compared with broadly south-easterly waves with more direct access to the site.

The study could involve wave transformation modelling followed by one or more nearshore joint probability analyses, or an offshore joint probability analysis followed by wave transformation modelling. Either approach would work. However, unless the nearshore conditions required can be very closely specified, it is probably better to undertake the joint probability analysis offshore, where the dependence is purely meteorological, and representative of a larger area. Wave transformation modelling introduces hydraulic effects (e.g. wave breaking) into the dependence analysis, which may be extremely site-specific. A small change in location, or in climate change or other uncertainty allowance, may mean that calculations would need to be repeated.

5.2.3 Decide how the variables will be represented

If offshore waves are divided into broadly northerly and broadly south-easterly sectors, they can be represented by specified wave heights and periods (based on waves within the relevant sector) for discrete return periods between about 0.1 year and 200 years. Sea level is unlikely to vary significantly in and around the bay, and sea level can be represented, for one location, by specified levels for discrete return periods between about 0.1 year and 200 years. Strictly speaking, separate extreme sea levels should be estimated corresponding to the times when waves are in one sector or the other, but without long-term local time series data it is impractical to implement this refinement to the analysis.

5.2.4 Obtain extreme values for each variable

Extreme sea levels would usually require a brief site-specific review of existing predictions, but for illustrative purposes, the following present-day values were drawn directly from Proudman (1997) for a location in north-east England. (Values for return periods below 1 year were estimated by the present author.)

Return period (yrs)	0.1	0.25	0.5	1	10	25	50	100	250
Sea level (mOD)	2.80	2.96	3.09	3.20	3.53	3.68	3.77	3.91	4.05

Extreme offshore wave conditions would usually require site-specific calculations, but for illustrative purposes, the following present-day values were estimated, based on a previous HR Wallingford study for a location in north-east England.

Return (years)		0.1	0.25	0.5	1	2.5	5	10	25	50	100	250
North	H_s (m)	5.0	5.8	6.4	7.0	7.8	8.4	9.0	9.8	10.4	11.0	11.8
	T_m (s)	8.0	8.6	9.1	9.5	10.0	10.4	10.7	11.2	11.5	11.9	12.3
South-east	H_s (m)	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0	6.2	6.4	6.6
	T_m (s)	7.7	7.8	8.0	8.2	8.3	8.5	8.6	8.8	8.9	9.1	9.2

5.2.5 Obtain the level of dependence between the variables

Figure 1 indicates a correlation coefficient $\rho = 0.17$, or alternatively the ‘modest’ (blue) level of dependence between large waves and high sea levels for all direction sectors combined for north-east England. However, in this instance, two wave direction sectors were used, and additional dependence information is needed from FD2308/TR1. Figure 4.2 (high dependence sector) indicates a correlation coefficient of $\rho = 0.36$ (but still modest or blue) for the northerly direction (330-45°N). Figure 4.3 (low dependence sector) shows $\rho = -0.01$ (independent, green) for the south-easterly direction (45-180°N).

5.2.6 Apply the desk study approach

The tables from which combinations of two variables with given joint exceedence return periods can be determined are given in Section 3.5 of FD2308/TR1. For illustrative purposes, consider only a 100 year joint exceedence return period. The relevant combinations are defined in Table 3.5 of FD2308/TR1, using Columns 1 and 2 (independent) for the south-easterly sector and Columns 1 and 3 (modestly correlated) for the northerly sector. Those three columns of return periods are reproduced below, together with the corresponding actual sea levels and wave conditions, based on values given in Section 5.2.4 and extrapolating where lower return period values are required.

Joint exceedence return period of 100 years							
Variable 1: Sea level		Variable 2: Wave conditions					
		South-easterly direction			Northerly direction		
Return period (years)	Sea level (mOD)	Return period (years)	Height H_s (m)	Period T_m (s)	Return period (years)	Height H_s (m)	Period T_m (s)
0.01	2.32	28	6.0	8.8		N/A	
0.02	2.45	14	5.9	8.7	100	11.0	11.9
0.05	2.66	6	5.6	8.5	60	10.5	11.6
0.1	2.80	2.8	5.4	8.3	28	9.9	11.2
0.2	2.93	1.4	5.3	8.2	14	9.3	10.9
0.5	3.09	0.6	5.0	8.0	6	8.5	10.4
1	3.20	0.28	4.8	7.8	2.8	7.9	10.0
2	3.31	0.14	4.7	7.7	1.4	7.3	9.7
5	3.43	0.06	4.5	7.6	0.6	6.5	9.2
10	3.53	0.03	4.3	7.5	0.28	5.9	8.7
20	3.64	0.014	4.1	7.3	0.14	5.3	8.2
50	3.77	0.006	3.8	7.1	0.06	4.5	7.6
100	3.91	0.003	3.6	6.9	0.03	3.8	7.1

5.2.7 Apply the results of the desk study approach

Full results would be in the form of a table, similar to that illustrated above, for each joint return period of interest. Before use at the seawall, the wave conditions need to be put through a wave transformation model to account for the headland, shallow water effects and breaking on the shallow foreshore. Following this, the lists of sea conditions can be converted into equivalent lists of overtopping rates, using appropriate equations. At this stage, there is probably no need to continue to distinguish between northerly and south-easterly waves, but the highest overtopping rate calculated for each joint exceedence return period provides an approximation to the overtopping rate with the same return period. If the point at which overtopping rate becomes unacceptable can be decided, then the corresponding joint return period provides an indication of the standard of service of the defence.

If sensitivity tests are to be made for the effects of uncertainty and/or future climate change, the changes to individual source variables are best applied to the results of the offshore joint probability analysis, before repeating the wave transformation and overtopping calculations. The precautionary allowance for possible future change in coastal variable-pair dependence suggested in Section 4.4.6 would apply to the north-east coast of England. Sensitivity to a possible 50% increase in flood probability would most easily be tested by temporary use of a return period of 150 years in the calculations.

5.3 Case study involving surges and river flows

5.3.1 The river situation

Consider water levels in the Clyde, south-west Scotland. Within Glasgow, the river is fully protected from ocean waves, and locally generated waves can be considered to have very minor impact on flood risk. River level is dependent upon downstream sea level, and upstream river flow, the relative importance of which may vary along the length of the river. The river wall at a particular location is potentially subject to overflowing under attack by high river flow combined with high sea level, and is to be tested for standard of service based on the criterion that overflow is unacceptable.

5.3.2 Select the variables and any conditions attached to those variables

One decision to make is whether to work with sea level or surge. The dependence mapping analysis in this project is based on surge, so as to focus on the meteorological dependence with river flow, and some specialists may prefer to retain surge as the variable of choice for joint probability analysis. However, in order to determine the downstream sea level required as input to a river model, it would be necessary to make some assumption about the accompanying astronomical tidal level (or to test a range of values). Sea level is more often used than surge, and is used in this case study. Extreme sea level estimates are more readily available than extreme surge estimates and, more importantly, sea level provides more direct input to subsequent river modelling. Although the dependence analysis in this report is based on surge, in the context of events occurring over more than one tidal cycle (as effectively assumed in the use of daily averaged river flows) it is not hugely conservative to assume that the same level of dependence also applies to sea level and river flow.

Although waves and rainfall may be of some interest, river flow and sea level (or surge) would be the normal variables to select for joint probability analysis in this situation. In some river flow situations, it may be necessary to consider different time lags between upstream river flow and downstream sea level for use in different parts of the river, and/or to consider the different behaviours of flood hydrographs of different durations. However, these are not important effects here and it is enough to consider just one hydrograph duration and no time lag between river flow and sea level.

The study could involve continuous simulation modelling based on several years of data on the two source variables as input to a river numerical model, or joint probability analysis of the two source variables, followed by river numerical modelling. Either approach would work, but it is probably better for prediction of extreme conditions to undertake the joint probability analysis on the source variables, where the dependence is purely meteorological, followed by river modelling, during which hydraulic dependence is introduced.

5.3.3 Decide how the variables will be represented

Extreme flows in the Clyde tend to last for about three days, and this fact would normally be taken into account in a flood risk assessment. However, for the purposes of illustrating use of the desk study approach, assume that each flow record / event lasts for one day. (The software tool accompanying this report permits the use of other durations, but the hard-copy version assumes either half-day or one-day record duration.) Flow rate during each record / event can be specified either as a daily average rate or, as nominally assumed here, the peak flow rate during the event, for discrete return periods between about 0.1 year and 200 years, at one upstream location. Sea level can be represented, for one downstream location, by specified levels for discrete return periods between about 0.1 year and 200 years.

5.3.4 Obtain extreme values for each variable

Extreme sea levels would usually require a brief site-specific review of existing predictions, but for illustrative purposes, the following present-day values were drawn directly from Proudman (1997) for a grid point just to the south of the mouth of the Clyde. (Values for return periods below 1 year were estimated by the present author.)

Return period (yrs)	0.25	0.5	1	10	25	50	100	250	500
Sea level (mOD)	2.34	2.48	2.62	3.07	3.26	3.37	3.56	3.75	3.86

Extreme river flows would usually require site-specific calculations, but for illustrative purposes, the following present-day values were estimated, approximately based on flow rates in the Clyde.

Return period (yrs)	0.25	0.5	1	10	25	50	100	250	500
River flow (m^3/s)	350	400	450	800	950	1100	1250	1500	1700

(The sea levels are a little low and the flows a little high compared to actual values in the Clyde, so these illustrative figures should not be used to draw conclusions about actual flood risk in the area.)

5.3.5 Obtain the level of dependence between the variables

Figure 2 categorises dependence between river flow and surge in the Clyde (Daldowie, Station No 84013) as 'strongly correlated' ($0.065 \leq \chi \leq 0.125$). Although this categorisation does not capture the exact site-specific value of dependence and the desk study approach does not capture the exact form of the χ dependence model, it provides a fair approximation, if the alternative would be to assume either independence or full dependence. (FD2308/TR1 gives a best estimate of $\chi = 0.10$ between the Daldowie flow gauge and the Millport tide gauge.)

5.3.6 Apply the desk study approach

The tables from which combinations of two variables with given joint exceedence return periods can be determined are given in Section 3.5 of FD2308/TR1. For illustrative purposes, consider only a 500 year joint exceedence return period. The relevant combinations are defined in Table 3.6 of FD2308/TR1, using Columns 1 and 5 (strongly correlated) taken together with the multiplier of 1.37 from Table 3.7. Those two columns of return periods are reproduced below (Column 5 values multiplied by 1.37) together with the corresponding actual sea levels and river flows, based on values given in Section 5.3.4.

Joint exceedence return period of 500 years			
Variable 1: Sea level		Variable 2: River flow Desk, strongly correlated	
Return period (years)	Sea level (mOD)	Return period (years)	Flow rate (m^3/s)
2	2.76	500	1700
5	2.93	233	1500
10	3.07	110	1250
20	3.21	55	1100
50	3.37	23	950
100	3.56	11	800
200	3.70	5.5	700
500	3.86	2.3	600

5.3.7 Apply the results of the desk study approach

Full results would be in the form of a table, similar to that illustrated above, for each joint return period of interest. Each combination of river flow and sea level would be used as input to a river numerical model (or similar) in order to predict water level at locations of interest within the fluvial-tidal reach of the Clyde. The highest water level predicted by the river model for each joint exceedence return period provides an approximation to the water level with the same return period. When water level exceeds river wall level, the corresponding joint return period provides an indication of the standard of service of the defence.

If sensitivity tests are to be made for the effects of uncertainty and/or future climate change, the changes are best applied to the individual source variables determined during the joint probability analysis, before repeating the water level calculations. The precautionary allowance for possible future change in fluvial variable-pair dependence suggested in Section 4.4.6 would apply to the west coast of Scotland. Sensitivity to a possible 100% increase in flood probability would most easily be tested by temporary use of a return period of 1000 years in the calculations.

5.4 Case study involving urban drainage

5.4.1 The hypothetical drainage situation

Consider a typical drainage analysis for a small coastal housing development, on the south coast of England. Stormwater drainage from the site discharges into the sea, and therefore the drainage system is potentially at risk of being unable to discharge stormwater runoff during intense rainfall coupled with a high sea level. It is to be tested for standard of service based on on-site flooding criteria.

5.4.2 Select the variables and decide how they will be represented

Although wave and wind conditions may be relevant in slightly raising the shoreline water level, sea level at a single location representative of the drainage outfall is the normal variable to select to represent sea conditions in this situation. The appropriate duration over which to characterise high intensity rainfall depends on the size of the area to be drained, and the expected duration of high water during a high sea level event. For illustrative purposes, two-hours is used here.

If possible, extreme values of both variables should be determined for discrete return periods from about 0.1 year up to the maximum return period of interest, say 200 years.

5.4.3 Obtain extreme values for each variable and the dependence between them

High intensity rainfall for a site is calculated as design rainfall events for the given hydrological region of the country. For illustrative purposes, the following present-day design rainfall values were calculated using the design rainfall method described in the Flood Studies Report (NERC, 1975) for a hydrological region in the south of England. Two-hour duration rainfall events were used and the values given represent the average rainfall intensity over the two-hour period.

Return period (yrs)	0.1	0.2	0.5	1.0	10	30	50	100	200
Two-hour rainfall intensity (mm/hour)	3.3	4.8	6.8	8.4	15.2	19.5	21.9	25.6	30.0

Extreme sea levels would usually require a brief site-specific review of existing predictions, but for illustrative purposes, the following present-day values were drawn directly from Proudman (1997) for a location on the south coast of England. (Values for return periods below 1 year were estimated by the present author.)

Return period (yrs)	0.1	0.25	0.5	1	10	25	50	100	250
Sea level (mOD)	2.20	2.35	2.48	2.61	2.92	3.11	3.23	3.45	3.67

Figure 3 indicates a correlation coefficient $\rho = 0.32$ (or alternatively the ‘modest’ (blue) level of dependence) between high intensity rainfall and high sea level in the middle of the south coast of England.

Drainage systems with extensive storage (for water quality or other reasons) will have long critical durations, when the dependency of the two-hour high intensity rainfall event would be less relevant. The higher dependency between river flow and surge in Figure 2 suggests that assessment of drainage storage systems, with long critical duration events, should apply a higher degree of dependency to any analysis.

5.4.4 Apply the desk study approach

Piped drainage systems are designed to discharge the 30 year return period critical duration event without surface flooding, with a 100 year return period against property flooding[1]. For illustrative purposes, consider a potential additional design situation involving combinations of high rainfall and high sea level with joint return periods of 30 years.

---

[1] Water UK and Water Research Centre, “Sewers for Adoption”, 5th edition, 2001, specifies the 30 year return period. The 100 year return period has become a standard for Environment Agency flood risk requirements, and is specified as a requirement in “Preliminary rainfall runoff management for development”, R&D Technical Report W5-074/A, 2004.

The tables from which combinations of two variables with given joint exceedence return periods can be determined are given in Section 3.5 of FD2308/TR1. As there is no table specifically for a 30 year joint exceedence return period, it is necessary to interpolate between the tables for the two nearest return periods: in this case Tables 3.4 and 3.5, for return periods of 20 and 100 years, respectively. The required results are obtained by selecting Column 1 of Table 3.4/3.5 and combining it with an interpolation between Column 3 (modestly correlated) of Table 3.4 and Column 3 of Table 3.5. The resulting return periods (excluding rainfall below 3mm/hour) are reproduced below, together with the corresponding actual rainfall and sea level conditions, based on values given in Section 5.4.3 and extrapolating where lower return period values are required.

Joint exceedence return period of 30 years			
Variable 1: Rainfall		Variable 2: Sea level Desk, modest	
Return period (years)	Rainfall (mm/hour)	Return period (years)	Sea level (mOD)
0.1	3.3	7.0	2.87
0.2	4.8	3.5	2.78
0.5	6.8	1.4	2.65
1	8.4	0.7	2.54
2	10.5	0.35	2.41
5	13.1	0.14	2.25
10	15.2	0.07	2.13
20	17.8	0.035	2.00
30	19.5	0.02	1.90

5.4.5 Apply the results of the desk study approach

Each combination of rainfall intensity and sea level would be used as input to a numerical model (or similar) in order to predict flooding at locations of interest within the development site. The highest flood level or flood extent calculated for each joint exceedence return period provides an approximation to the water level or flood extent with the same return period. When the flood level or extent exceeds the on-site flood storage, the corresponding joint return period provides an indication of the standard of service of the drainage system.

If sensitivity tests are to be made for the effects of uncertainty and/or future climate change, the changes are best applied to the individual source variables determined during the joint probability analysis, before repeating the flood level calculations. The precautionary allowance for possible future change in fluvial variable-pair dependence suggested in Section 4.4.6 would be relevant here. Sensitivity to a possible 100% increase in the probability of the 30 year return period loading conditions would most easily be tested by temporary use of a return period of 60 years in the calculations.

5.4.6 Other considerations if drainage were into a river rather than the sea

In the case of drainage and river flows, there is some difficulty in deciding upon a suitable sampling process. Rivers and sewer systems both respond to rainfall, and therefore a high level of correlation between high river flow and high sewer flow might be expected. This is particularly true where the drainage system is critical for long duration events, due to the use of extensive storage. The critical rainfall duration for drainage systems can vary greatly, ranging from 15 minutes to over 24 hours; the latter being very similar to the period for generating high river flows.

River flood characteristics are a function of antecedent and seasonal weather conditions and therefore defining a regular interval to ensure independence between events is theoretically not possible. Urban systems, on the other hand, can assume that an inter-event dry period of 24 hours ensures independence for sampling.

As in the case of river flows and sea levels, it is suggested that a 1 or 2 week period would be appropriate for sampling the source data series. Alternatively, a peaks over threshold approach based on Q5 (the river flow rates which occur less than 5% of the time) could be used. The sampling and ranking of the rainfall would be based on a duration that was relevant to the drainage system being considered.

5.5 Case study involving wind-sea and swell

5.5.1 The hypothetical coastal situation

Consider a coastal defence project in Lyme Bay, which is exposed to storm wave conditions and to longer period Atlantic swell coming from the south-west. Although the swell wave heights are not as large as the storm wave heights, their longer periods may be capable of causing greater damage to sea defences. It is not obvious whether extreme swell, extreme storm waves or some combination of wind-sea and swell represents a worst case for coastal defence design. It would therefore be helpful to have a range of different wave conditions, each with the same joint return period, for potential use in design. (Incidentally, this group of related wave conditions, now thought of as a single sea state with a given return period, could be used in joint probability analysis with sea level, if it were necessary also to consider the probability of sea level.)

5.5.2 Select the variables and any conditions attached to those variables

The two variables to select for joint probability analysis in this situation are the two components of the overall offshore wave condition, i.e. wind-sea and swell. As swell in Lyme Bay tends to come only from the south-west, it is not necessary to sub-divide the analysis into direction bands. However, it may be necessary to look separately at different types of swell, e.g. recently produced young swell with a wave period in the range 10-14s and distantly generated old swell with a wave period in the range 14-18s. (HR Wallingford (1997) contains swell data for up to four separate bands of wave period, depending on location around England and Wales.)

The study could involve wave transformation modelling followed by a nearshore joint probability analysis, or an offshore joint probability analysis followed by wave transformation modelling. Either approach would work, but it is probably better to undertake the joint probability analysis offshore, where the dependence is purely meteorological, followed by wave transformation modelling, during which hydraulic dependence is introduced.

5.5.3 Decide how the variables will be represented

An offshore wave condition can usually be reasonably well represented by specified significant wave height, mean wave period and, if relevant, a mean wave direction. However, the situation is a little different here, as the intention is to represent two components of the wave condition, which may be difficult to separate from each other. Ideally, wind-sea will be evaluated separately from the total-sea, perhaps using a wave model incapable of predicting swell, but total-sea is probably a reasonable approximation; wave height and period are needed for discrete return periods between about 0.1 year and 200 years. The swell variable is conveniently represented solely by wave height, but with separate estimates for 10-14s swell and for 14-18s swell, for discrete return periods between about 0.1 year and 200 years.

5.5.4 Obtain extreme values for each variable

Extreme offshore wave conditions would usually require site-specific calculations, but extreme offshore swell predictions for England and Wales can be taken from HR Wallingford (1997). For illustrative purposes, the following present-day values were estimated, based on a previous HR Wallingford local wave prediction study in Lyme Bay and swell wave height predictions in Lyme Bay in HR Wallingford (1997).

Return (years)		0.1	0.25	0.5	1	2.5	5	10	25	50	100	250
Wind-	H_s (m)	3.8	4.1	4.3	4.5	4.8	5.0	5.2	5.5	5.7	5.9	6.2
sea	T_m (s)	6.7	6.9	7.1	7.3	7.5	7.6	7.8	8.0	8.1	8.3	8.5
10-14s	H_s (m)	2.7	3.0	3.3	3.6	3.9	4.1	4.4	4.7	4.9	5.1	5.4
swell	T_m (s)	12	12	12	12	12	12	12	12	12	12	12
14-18s	H_s (m)	1.6	2.0	2.3	2.6	3.0	3.3	3.6	3.9	4.2	4.5	4.8
swell	T_m (s)	16	16	16	16	16	16	16	16	16	16	16

5.5.5 Obtain the level of dependence between the variables

Figure 4 indicates the ‘modest’ (blue) level of dependence between wind-sea and swell in Lyme Bay. (In this project, only one level of dependence is presented for each area, regardless of swell wave period, but more precise information for England and Wales is given in HR Wallingford, 1997, if required.)

5.5.6 Apply the desk study approach

The tables from which combinations of two variables with given joint exceedence return periods can be determined are given in Section 3.5 of FD2308/TR1. For illustrative purposes, consider only a 20 year joint exceedence return period. The relevant combinations are defined in Table 3.4 of FD2308/TR1, using Columns 1 and 3 (modestly correlated). Those two columns of return periods are reproduced below, together with the corresponding actual wave conditions, based on values given in Section 5.5.4 and extrapolating where lower return period values are required.

Joint exceedence return period of 20 years					
Variable 1: Wind-sea			Variable 2: Swell (use one or other type)		
Return period (years)	Height H_s (m)	Period T_m (s)	Return period (years)	Height H_s (m) of 12s swell	Height H_s (m) of 16s swell
0.01			N/A		
0.02	3.3	6.3	20	4.6	3.8
0.05	3.6	6.5	8	4.3	3.5
0.1	3.8	6.7	4	4.0	3.2
0.2	4.0	6.9	2	3.8	2.9
0.5	4.3	7.1	0.8	3.5	2.5
1	4.5	7.3	0.4	3.2	2.2
2	4.7	7.4	0.20	2.9	1.9
5	5.0	7.6	0.08	2.6	1.5
10	5.2	7.8	0.04	2.4	1.2
20	5.4	8.0	0.02	2.1	0.9

5.5.7 Apply the results of the desk study approach

The results are in the form of combinations of the two types of wave condition, expressed in the form of two wave heights and two wave periods, for each return period (20 year illustrated above) and for each category of swell (12s and 16s swell illustrated above). For example, the table above indicates that a wind sea of $H_s = 4.3$m and $T_m = 7.1$s would be expected to occur in conjunction with a 12s swell of $H_s = 3.5$m and with a 16s swell of $H_s = 2.5$m, once each, on average, every 20 years. (Similarly for each other line in the table.)

Before use at the coast, these wave conditions would need to be put through a wave transformation model. Appendix 3 of HR Wallingford (1997) lists FORTRAN program code capable of generating a bi-modal spectrum from the two separate offshore components, which may be helpful in preparing wave spectra for input to the wave transformation model.

Following this, the nearshore sea conditions may be used to calculate equivalent coastal responses, using appropriate hydraulic models. This is far from being a simple procedure, as most desk study and numerical methods assume a standard spectral shape, usually defined by single height and period parameters. HR Wallingford (1998) provides guidance on how to apply bi-modal sea conditions in design, and their potential importance on differently exposed coasts.

5.6 Issues involved in applying the desk study approach in a larger area

Joint probability analysis at a single location can usually be reduced to just two primary flood risk source variables. As the area covered by flood risk analysis grows larger, some of the following difficulties may develop:

- More than two types of source variable may be important, e.g. sea level, waves and river flow

- More than one value of a single type of source variable may be needed to represent conditions across the area, e.g. different wave conditions along a frontage, or different river flows in different rivers
- Control of spatial coherence between different parts of the area may become an issue for overall flood risk and estimation of extreme losses
- Time lag between peak values of different variables may become more important in prediction of their combined effect on flood risk.

The last of these points is the easiest to accommodate. An appropriate time lag is estimated beforehand and is carried through the data preparation, joint probability analysis and application stages.

The techniques below, which can be combined, offer some assistance in addressing the first three points above. They are illustrated with reference to how they might be applied in the outer Thames.

Use of conditional analyses (or division of populations)
This technique is common in coastal engineering, where joint probability analysis of large waves and high sea levels may be divided into a small number of discrete and mutually exclusive direction sectors. It might also be applied to a small number of different durations of rainfall or river flow. Wave direction (or rainfall duration) then becomes the ‘condition’, and the records meeting that condition provide a ‘population’. Each population can then be analysed separately, with the appropriate number of events per year and marginal extremes for the proportion of data meeting the condition. Each set of joint probability results should then be considered as a potential worst case.

Relevance to use in the Thames
As on most of the east coast of England, the highest sea levels in the Thames Estuary tend to be associated with winds from the north-west (and sometimes north-east). However, unlike most of the east coast of England, waves from the north-west may constitute a significant flood risk on the south bank of the Thames. In the Thames numerical case study described in Section 6.6, four wave direction sectors are used, to provide the best balance between the different characteristics of different sectors and the need for sufficient data in each sector. The same division could be used within the desk study approach, with separate analyses for the four direction quadrants. Results for the north-west and north-east sectors would be applied at locations on the south bank of the Thames, and those for the two southerly sectors on the north bank.

Evaluation of a source variable at just one representative location
This is a commonly used technique, in which the joint probability analysis is based on the value of a source variable at just one location, which is later used to reproduce dependent values at other locations. This is usually adequate for sea level, which varies in a fairly predictable way along a length of coast. Similarly it will often be appropriate to calculate wave conditions offshore and then to infer equivalent conditions at several nearshore locations. The method might also be extended, but with much greater caution, to rainfall or river flow, by assuming that conditions are strongly dependent on those occurring in neighbouring catchments.

Relevance to use in the Thames

Statistical analysis and river numerical modelling during a previous study of the Thames Estuary indicated that the tide- and surge-induced components of high estuarine water level could be estimated reliably from a single representative downstream sea level. The joint probability analysis carried out during the present project (see Section 6.6) was conducted in terms of high tide sea level at Southend. Values elsewhere within the Thames are then intended to be inferred from river numerical modelling or from summary look-up tables.

Use of proxy source variables

This is a similar technique to the previous one, but less often used. It may be possible to work in terms of a different underlying source variable, from which the flood risk variables can later be re-constructed. An example is the use of wind speed (usually conditional upon wind direction) as a proxy variable for later input to multiple wave prediction models. Similarly, rainfall (perhaps conditional upon duration) might be used a proxy variable for later input to multiple river flow prediction models.

Relevance to use in the Thames

During the joint probability analysis for the Thames Estuary, sea conditions were divided into four direction sectors, and sea levels were derived from a single source at Southend. Within each sector, it was convenient to represent wave conditions by a wind speed representative of conditions over water in the outer estuary. The same representation of winds / waves could be used within the desk study approach. The intention would be that in flood risk studies, the wind speeds (and directions) would be used as input to a series of local wave prediction models, and the sea levels would be used as input to numerical river models. This would have the effect of producing spatially coherent wave and sea level conditions throughout the lower tidal Thames (i.e. excluding parts sensitive to river flow).

Use of multiple joint probability analyses

Joint probability analyses can be linked, but only to a limited extent, as this technique rapidly becomes unmanageable. The first analysis might be applied to large waves and high sea levels to infer the approximate distribution of coastal overtopping rate. This in turn could then be regarded as a primary variable for further joint probability analysis against river level overtopping rate.

Relevance to use in the Thames

Multiple spatially correlated joint probability analyses could become very complicated in the Thames, because of the many separate river flows in tributaries, the variability of waves and sea levels, and the existence of control structures. Probably the only practical extension of the approach described in the preceding paragraphs would be to treat the combined probability of ‘sea conditions’ (described by Southend wind speed, sea level and direction sector) as a single variable. This variable could then be used as input to separate joint probability analyses with river flow in the upper Thames and in each tributary. This would admittedly give a poor representation of dependence between different flows, but with little connectivity between the associated flood risks, probably not an unacceptable assumption.

5.7 The software tool

A software tool for evaluation of joint exceedence extremes is provided on the floppy disk at the back of this report. It works in the same way as the simplified method described in FD2308/TR1 and the desk study approach described above, but is not limited to the joint return periods and levels of dependence for which tables of results have been pre-computed.

Information is entered only on the pink and blue cells on Sheet 1 ('user input') of the Excel spreadsheet.

1. Enter text descriptions (including units) of each of the two primary variables in the top two pink boxes. These are used to annotate the results tables.
2. Just below these descriptions, enter the number of records per year (e.g. 707 for high tide data or 365 for daily averaged data).
3. Enter a '1' in one of the three blue boxes (and clear the other two) to indicate the type of dependence measure to be used, i.e. either ρ, or χ or CF. Enter the corresponding value of the chosen dependence measure in the neighbouring pink box. A ρ value is automatically changed to the nearest equivalent CF value before calculations are performed, but the spreadsheet uses a separate approximation to the χ statistical model if χ is selected.
4. In the remaining pink boxes to the left of the yellow bar, enter between one and eight joint exceedence return periods, lowest at the top, using as many of the eight pink boxes as required. There is no upper or lower limit, but the return periods chosen should be consistent with the marginal return periods available (Step 5.).
5. In the first column of pink boxes to the right of the yellow bar, enter between six and eleven return periods for which the marginal extreme values are known, lowest at the top, using as many of the pink boxes in Column L as required. In the other two columns, enter the corresponding extreme values of each of the two primary variables. If possible, the marginal return periods should range from about 0.05 year up to fractionally higher than the highest joint return period of interest, but the tool will attempt to extrapolate below the range given. For best results, the marginal return periods should be approximately in geometric sequence, e.g. 0.1, 0.2, 0.5, 1, 2, 5 years etc.

Sheet 2 ('return period output') shows the table of appropriate combinations of the two variables, for each joint return period required, expressed in terms their marginal return periods. Sheet 3 ('source variable output') shows a similar table, but now expressed as actual values of the two variables, together with a plot of the joint exceedence contour for each return period selected. The chart and axis titles and the axis scales have to be edited manually if required.

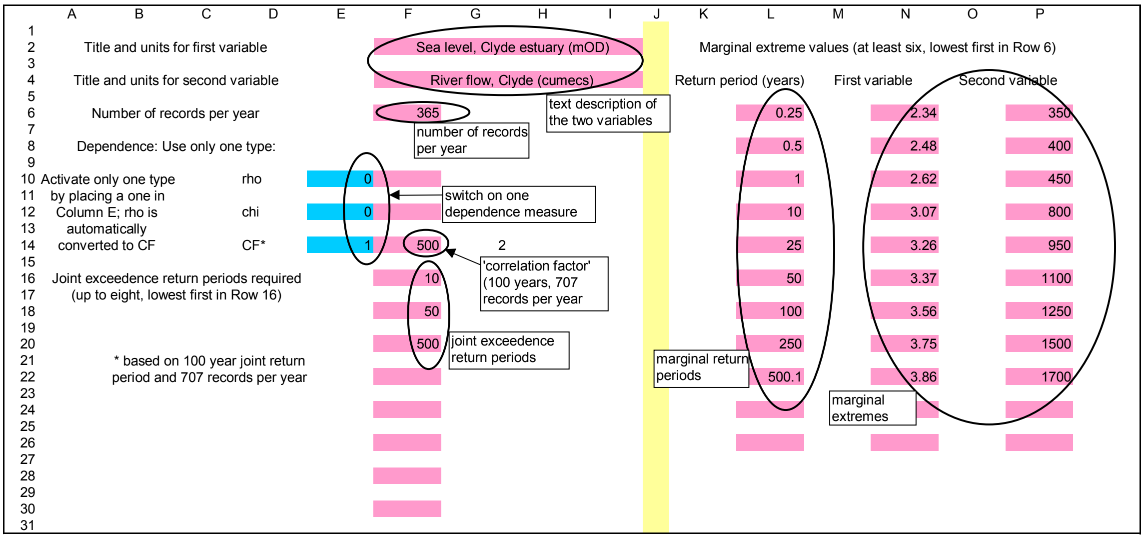
The top three rows of results in Sheets 1 and 2 involve a simple extrapolation outside the input data range; if these results are clearly wrong, they can be disregarded. The curves plotted in Sheet 3 often consist of three distinct parts, particularly when dependence is high; the two outer parts are valid, but can be disregarded for practical purposes, since the 'worst case' conditions will always lie within in the central part.

If the source data are divided into sub-sets before analysis, for example separation by direction sector, then separate runs of the tool are required for each sub-set.

5.7.1 Software tool training example

The copy of the software tool included with this report contains an example analysis using wave heights and sea levels. New users might like to begin by testing their ability to enter data, using the case study outlined in Section 5.3, involving surges and river flows in the Clyde (but now with return periods of 10 and 50 years in addition to the original 500 years). Figure 5 is a screenshot of Sheet 1 ('user input') annotated with ovals to show where data has been entered in the pink and blue boxes of the spreadsheet, and rectangular text boxes noting what has been entered within each oval.

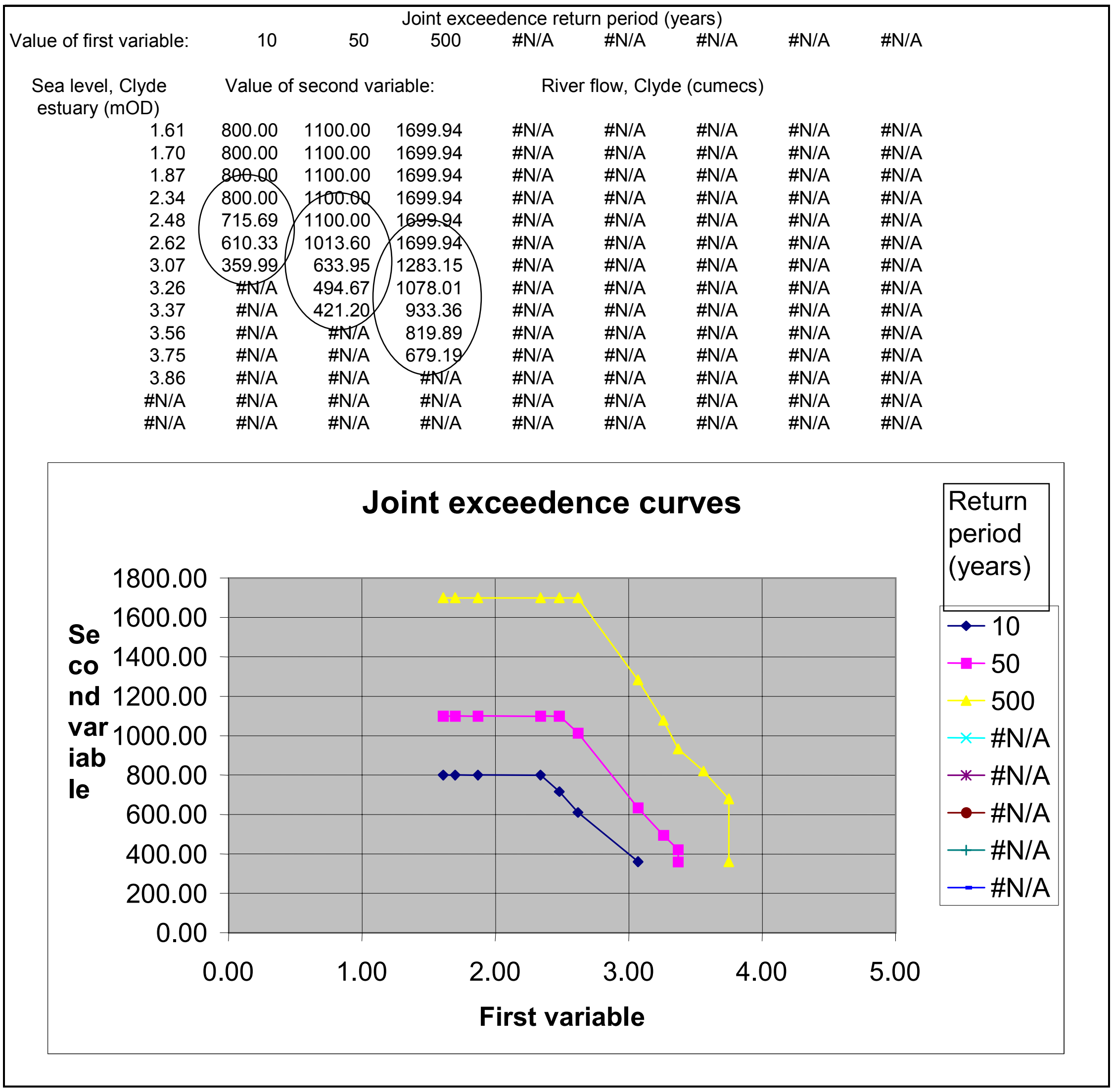
Section 5.3.5 describes the dependence between surge and river flow as 'strongly correlated', one of the five bands of dependence used in the colour-coded dependence maps. This method of specifying dependence is communicated to the spreadsheet by entering a '1' into the blue box next to 'CF' (and zeroes in the other two blue boxes). The level of dependence is entered as 500 (see Table 5 of FD2308/TR1) in the neighbouring pink box, corresponding to the value of 'correlation factor' (in its original context of a joint exceedence return period of 100 years and a number of records per year of 707).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2		Title and units for first variable				Sea level, Clyde estuary (mOD)					Marginal extreme values (at least six, lowest first in Row 6)					
3																
4		Title and units for second variable				River flow, Clyde (cumecs)					Return period (years)		First variable		Second variable	
5																
6		Number of records per year				365						0.25		2.34		350
7																
8		Dependence: Use only one type:										0.5		2.48		400
9																
10	Activate only one type			rho	0							1		2.62		450
11	by placing a one in															
12	Column E; rho is			chi	0							10		3.07		800
13	automatically															
14	converted to CF			CF*	1	500	2					25		3.26		950
15																
16	Joint exceedence return periods required					10						50		3.37		1100
17	(up to eight, lowest first in Row 16)															
18						50						100		3.56		1250
19																
20						500						250		3.75		1500
21			* based on 100 year joint return													
22			period and 707 records per year									500.1		3.86		1700
23																
24																
25																
26																
27																
28																
29																
30																
31																

Figure 5 Example input to the desk study software tool, corresponding to the outline fluvial case study in Section 5.3

Figure 6 is a screenshot of the resulting Sheet 3 ('source variable output'). 'Worst case' combinations of sea level and river flow, for any particular joint exceedence return period, will come from the diagonal part of the curve in each case. The corresponding results in the table above are highlighted with ovals in Figure 6.



Joint exceedence return period (years)

Value of first variable:	10	50	500	#N/A	#N/A	#N/A	#N/A	#N/A
Sea level, Clyde estuary (mOD)	Value of second variable:			River flow, Clyde (cumecs)				
1.61	800.00	1100.00	1699.94	#N/A	#N/A	#N/A	#N/A	#N/A
1.70	800.00	1100.00	1699.94	#N/A	#N/A	#N/A	#N/A	#N/A
1.87	800.00	1100.00	1699.94	#N/A	#N/A	#N/A	#N/A	#N/A
2.34	800.00	1100.00	1699.94	#N/A	#N/A	#N/A	#N/A	#N/A
2.48	715.69	1100.00	1699.94	#N/A	#N/A	#N/A	#N/A	#N/A
2.62	610.33	1013.60	1699.94	#N/A	#N/A	#N/A	#N/A	#N/A
3.07	359.99	633.95	1283.15	#N/A	#N/A	#N/A	#N/A	#N/A
3.26	#N/A	494.67	1078.01	#N/A	#N/A	#N/A	#N/A	#N/A
3.37	#N/A	421.20	933.36	#N/A	#N/A	#N/A	#N/A	#N/A
3.56	#N/A	#N/A	819.89	#N/A	#N/A	#N/A	#N/A	#N/A
3.75	#N/A	#N/A	679.19	#N/A	#N/A	#N/A	#N/A	#N/A
3.86	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

Figure 6 **Example output from the desk study software tool, corresponding to the outline fluvial case study in Section 5.3**

6. THE ANALYTICAL APPROACH TO JOINT PROBABILITY ANALYSIS

6.1 The analysis method

Only JOIN-SEA is described in any detail, as its development and testing has been funded by Defra over a period of several years, and to provide a focus for those wishing to begin using a well established method. Although the project team might use the alternative methods described in FD2308/TR1 it is not envisaged that they would be recommended for use elsewhere. The focus on JOIN-SEA does not imply discouragement or criticism of similar approaches developed within other organisations, and many of the points of best practice given in this chapter would also be applicable to other approaches.

6.1.1 Theory

The theory behind JOIN-SEA is outlined in FD2308/TR1, and detailed in HR Wallingford (2000a) in which development and testing of the method are also described. Anyone intending to use JOIN-SEA is recommended to obtain copies of those reports, and also the JOIN-SEA user manual (HR Wallingford, 2000b). The key stages are summarised below:

Step 1: Preparation of input data
Each record consists of simultaneous values of the two or three parameters used to represent the source variables. The records should be a representative sample of at least three years of data.

Step 2: Fitting of marginal distributions
Generalised Pareto Distributions are fitted to the top few percent of the marginal variables, e.g. wave heights and sea levels. If waves are involved, the empirical distribution of wave steepness is modelled by a Normal regression on wave height.

Step 3: Fitting of statistical models for dependence
A dependence function is fitted to the distribution of the two primary variables. Two alternative partial dependence statistical models are available, namely a single Bivariate Normal (BVN) Distribution and a mixture of two BVNs.

Step 4 Long-term simulation
This stage involves simulation of a large sample of synthetic records, based on the fitted distributions, and with the same format and statistical characteristics as the input data. JOIN-SEA incorporates any refined knowledge of the marginal extremes by re-scaling during the long-term simulation of data, thus permanently building this information into the long-term simulation.

Step 5: Analysis of joint exceedence extremes and structure functions
This stage involves analysis of the long-term simulation to extract extreme values. These can take the form of marginal extremes, or extreme combinations of the two primary variables. In addition, any structure function (e.g. overtopping, run-up, force) which can be defined in terms of constants and the variables present in the simulation can be

synthesised directly for every record. Direct analysis of the distribution and extremes of the structure variable is then relatively easy.

6.1.2 Data requirements

The analytical approach requires discrete records of the simultaneous (or separated by a specified time interval) occurrence of the variables to be analysed. There is no particular minimum length of data required to use this method, and a reasonable estimate of dependence can be made from one year of data. However, given that extreme values have been published for most marginal variables, and dependence values for most variable-pairs, both based on long time series data, three years seems a reasonable minimum length in order for a new analysis to add much value to existing knowledge. The records need not be sequential but in practice are usually extracted from time series.

Additional information on the marginal extremes and/or dependence can be easily and permanently incorporated into the analysis if required. This information will usually come from separate analysis of a longer period of data which may be available for one or more of the variables. (Some sources of published extreme values of flood risk variables are listed and briefly described in Section 1.4.4 of FD2308/TR1.) Alternatively, extreme values may have been established using a different analysis method, for example based on a different probability model, a different way of declustering the source data, involving de-trending of sea level data, or simply at an earlier stage in a project. This ability to assimilate additional information means that there is no particular minimum length of simultaneous data required for use of the analytical approach.

6.1.3 Steps in application and extent of user discretion

Selection, gathering and preparation of data

This is the stage in which the user has greatest involvement and discretion, and therefore the greatest chance to set up a successful analysis producing meaningful results. In selecting, gathering and preparing data, the user might have in mind the need to balance the following requirements.

1. To keep the number of primary variables down to a manageable two, or possibly three.
2. To keep the number of parameters needed to represent each variable down to a manageable one, or two if the second is wave period or if the second is either fixed (e.g. hydrograph duration or wave direction) or fully dependent upon the first parameter (e.g. peak current speed on high tide sea level).
3. To develop an event definition such that input records are independent of each other, and with enough parameters to re-construct a meaningful event for use in later hydraulic modelling or predictions. (One way of bringing in an additional parameter, not involved in extremes analysis, involves division of the data into separate populations to be analysed separately).
4. To use as great a length of data as possible (a minimum of about three years) to improve reliability.
5. To keep staff time and data purchase cost down to level appropriate to the value of the study.

Fitting of distributions to each variable
This task is largely automated within the JOIN-SEA *BVN* and *MIX* programs, although the user does have discretion over thresholds for distribution fitting and starting points for optimisation of distribution parameters. It is recommended that a few different fitting thresholds be tried as a sensitivity test, to find the point at which the most stable statistical model fit is achieved. It is not usually necessary to try alternative starting values for the parameter fitting unless the fitting programs report difficulty in evaluating the model parameters.

Fitting of distributions to the dependence between variable-pairs
This task is largely automated, but the user has one decision to make which is critical to the accuracy of the joint probability analysis. The user has to examine the variation of correlation coefficient (ρ) with threshold, calculated by JOIN-SEA program *BVN*. If ρ remains roughly constant above a certain threshold then the single BVN distribution can be used with this ρ value specified above this threshold. If ρ does not tend towards a constant value, then the mixture of two BVN distributions available within JOIN-SEA program *MIX* is likely to provide a better statistical model.

Simulation of a very large sample of data from the fitted distributions
The long-term simulation is automated within the JOIN-SEA *SIMBVN* and *SIMMIX* programs, but the user has discretion at this stage to bring in additional information on extreme values, and to set the size of the simulation to suit the purpose of the study. The user can tell the simulation to adjust high and extreme values of one or both primary variables to match more reliable extremes than could be determined solely from the data used in the joint probability analysis. The user can also choose the size of the long-term simulation (ten to twenty times the highest return period of interest is reasonable) and whether to store all simulated records or just those over thresholds of interest (called importance sampling).

A question raised from time to time is why the JOIN-SEA developers chose to use a Monte Carlo simulation approach, rather than an entirely analytical extrapolation of the fitted statistical distributions. Whilst this would have been theoretically possible, the combined extrapolation of three or more statistical models, some in transformed space, and subsequent conversion of results to flood risk structure functions, would be unmanageable and difficult to check.

Simulation of structure variables; evaluation of joint exceedence extremes
The user has some flexibility as to how the long-term simulation data are applied. Apart from the loss of information on sequencing and seasonality, they can be treated like a very large sample of measured data. The *ANALYSIS* program that forms part of the JOIN-SEA package is able to determine marginal and joint exceedence extremes from the long-term simulation by counting back through the highest values. The same program also contains demonstration versions of four coastal response functions, namely, overtopping rate, run-up, force and armour size.

Checking of results
The JOIN-SEA programs produce some diagnostic information, indicating when it has been difficult or even impossible to fit standard statistical models to the data. When this happens, different starting conditions can be tried in the parameter optimisation routines, or it may be necessary to re-visit the data gathering and preparation stage of

the analysis. Assuming an apparently successful long-term simulation, the user should routinely make the following checks on the results:

1. If the option to re-scale marginal extremes was used, check that the extreme values reported by *ANALYSIS* agree with expectations.
2. If the re-scaling option was not used, check that the 1 year return period value reported by *ANALYSIS* was exceeded approximately once per year of data within the source data sample. Stated more generally, check that (number_over) is approximately equal to (t/T), where t is the number of years of data in the sample and T is the return period in years.
3. Plot the joint exceedence extremes (even if these are otherwise not used) on axes representing the two primary variables. A lack of smoothness would suggest insufficient source data, too short a simulation and/or over-adjustment during re-scaling of the marginal extremes.
4. An approximate counting analysis check of the joint exceedence extremes is possible but not simple. If the marginal extremes had not been re-scaled, the number of source data points outside a joint exceedence contour should lie somewhere between $(t/T)\log_e(nT)$ if the two primary variables were independent and (t/T) if they were fully dependent, where n is the number of records per year.

6.1.4 Sensitivity and uncertainty

Sensitivity tests are helpful in understanding the magnitude and impact of any uncertainties. Aspects of the joint probability analysis that might be tested in this way are the thresholds used for statistical model fitting, the assumed level of dependence, and whether joint exceedence or joint density extremes are used. Experience at HR Wallingford over many studies suggests that joint probability results are most sensitive to uncertainties in the marginal extremes, then to the assumed dependence, and a little to interpretation of the true meaning of the results.

Defra / Environment Agency (2002) discusses the different sources and magnitudes of error and uncertainty, and how they can be combined to estimate overall uncertainty. A simplified method of combining uncertainties is illustrated by the following example. Suppose that three independent but cumulative sources of uncertainty can be expressed as 90% confidence that the true value lies within ±20%, ±10% and ±30% of the best estimate values. Then the composite uncertainty, at the same 90% confidence level, would the square root of $(20^2 + 10^2 + 30^2)\%^2$, i.e. ±37%.

There is a special version of JOIN-SEA which evaluates the uncertainty associated with each individual parameter of the fitted statistical models. Its subsequent calculation of the overall statistical uncertainty in the results is quite complex because the individual parameter uncertainties are inter-dependent. In practice this version is rarely used, partly because statistical uncertainty is usually a relatively small part of the overall uncertainty due to all causes.

6.1.5 Practical points during application of results

See Section 5.1.5 regarding the use of joint exceedence extremes and allowance for possible future climate change.

6.1.6 A note on the case studies

Sections 6.2-6.6 contain outline case studies for five different physical zones: coastal, tidal, river, urban drainage and complex area. Each is presented in a different way to illustrate different aspects and issues of the analysis procedures, so there would be value in looking at all of the case studies even if direct interest is limited to just one or two of the physical zones.

6.2 Case study involving waves and sea levels

As this is the main case study for JOIN-SEA, involving all stages of the analysis procedure, it is described at greater length than the case studies presented later in this chapter.

6.2.1 Outline of the case study

North Wales has been used as one of the main case studies in a series of Defra-funded joint probability research studies since 1990 (e.g. HR Wallingford, 2000a and 2000b). The results have also been used as the basis of the joint probability assessment in several consultancy studies on the North Wales coast. Throughout that time, the source data set has consisted of the same ten years or so of high water records of measured Liverpool tide gauge data combined with deep water predictions of wave height and wave period off the North Wales coast. It provides an interesting case study because of the large tidal range, the relatively high dependence between large waves and high sea levels, and a noteworthy flooding event for validation purposes in February 1990.

Although consultancy studies of the North Wales coast have always been based on actual sea levels, HR Wallingford's research studies, which often require release of source data, have always used an arbitrary datum for the North Wales sea levels. Even though the sea level data are no longer subject to a restrictive licence agreement, the same arbitrary sea level datum is used here for continuity with previous research case studies.

6.2.2 Selection, gathering and preparation of data

As in the present study, the original priority was to assess the dependence between the source variables (waves, sea level and surge) and to estimate the return period of known extreme events. To meet this requirement, a long period of consistent good quality data was required, including the time of the damage during February 1990. At the time of the original study, it was convenient to use tide gauge measurements at Liverpool 1970-1983 and wind gauge measurements for the same period at Blackpool to be used as input to a deep water wave prediction model.

From the point of view of achieving a representative sample, it might have been better to limit the analysis to the period 1970-1983. However, for coastal engineering purposes it would have been unacceptable to exclude the known extreme conditions during February 1990, and so an additional few days of data covering that period have always been part of the data set. The joint probability analyses have always been undertaken in terms of sea level (or sometimes surge) at Liverpool and deep water wave conditions approximately off Rhyl. When equivalent conditions at particular coastal locations were needed subsequently, the results were modified for differences in tidal range and for shallow water effects on waves.

The usual event definition was adopted, namely sea conditions at each high water. The tide gauge data were scanned to identify high waters, at each of which a record was constructed, consisting of sea level, significant wave height and mean wave period. It was noted that this assumption is conservative, in that the two or three days during February 1990 when damage occurred provided four or five separate records of high sea conditions, which would subsequently be treated as independent events. The 9272 source data are summarised in Figure 7 (records for February 1990 shown as triangles).

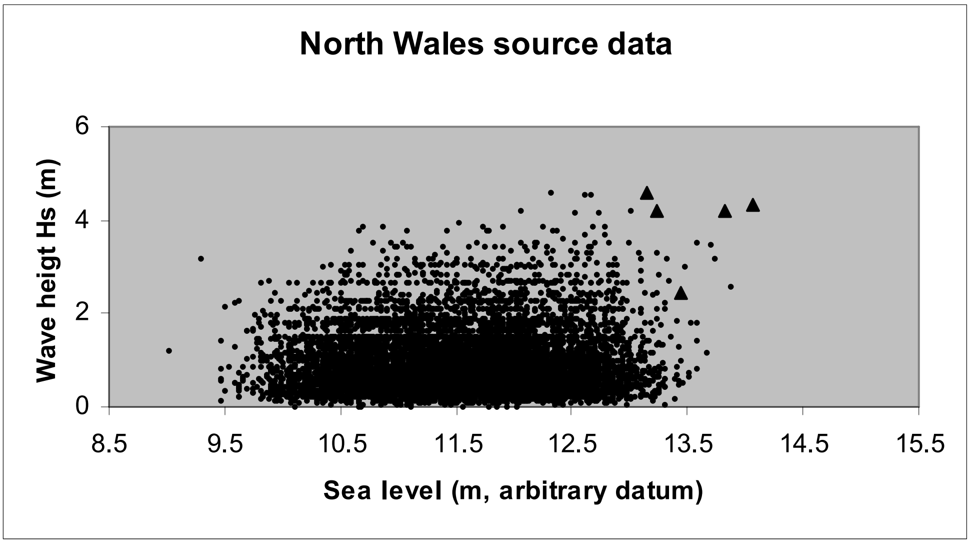


Figure 7 Coastal case study: source data 1970-1983 (9267 dots) plus records during February 1990 event (5 triangles)

It would have been possible to separate the data by wind / wave direction, perhaps into westerly and northerly sectors. However (with the possible exception of short stretches of coast not exposed to westerly waves) there would have been little point in doing so, as the westerly sector clearly provides the most severe sea conditions, and so the data set was not divided by direction.

6.2.3 Fitting of distributions to each variable

A number of different distributions have been applied during previous studies of this data set. The JOIN-SEA *BVN* and *MIX* programs automatically fit a GPD to the largest wave heights and the highest sea levels, unless told otherwise basing them on the top five percent of the source data. In practice the marginal extremes would usually be re-scaled to values derived carefully from a longer period of source data, or perhaps to values established during an earlier study. However, if re-scaling is not to be applied, then it is recommended to run *BVN* several times using alternative fitting thresholds to see which provides the most stable result. Values other than the defaults can be chosen to start the parameter optimisation routines in JOIN-SEA if recommended by the program's diagnostic messages or if extreme values look wrong, but this is not usually necessary.

6.2.4 Fitting of distributions to the dependence between the variables

The JOIN-SEA *BVN* and *MIX* programs fit dependence at the same time as fitting the marginal distributions but, unlike the marginal distributions, final selection of the most appropriate dependence model always requires some judgment by the user. Normally, *BVN* would be the first of the JOIN-SEA programs run during a joint probability analysis. It determines the best value of correlation coefficient, ρ, for a number of different thresholds, assuming that the single Bivariate Normal can be applied to the source data. Figure 8, showing the dependence results for North Wales (with confidence limits set to one standard error) shows that this would be a poor assumption. ρ increases with threshold, and there is no threshold above which it is reasonably constant, at which a representative value could be taken.

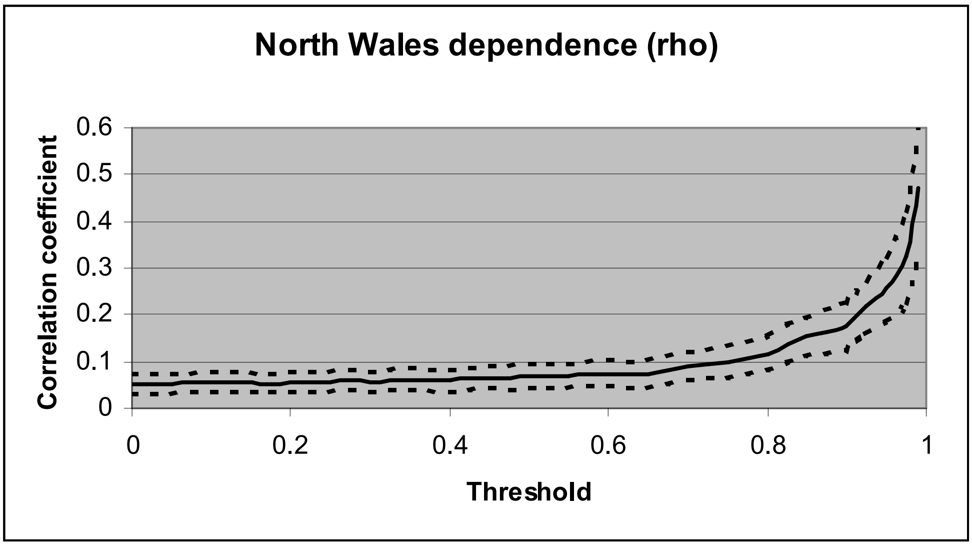


Figure 8 **Coastal case study: dependence between wave height & sea level**

In this situation, the *MIX* program is used as an alternative to *BVN*; the marginal fitting is the same in both programs, but the more complex mixture of two BVNs dependence model is used. There are seven statistical parameters to be fitted and different starting values often lead to different results. HR Wallingford (2000a) reports a best fit of 3.2% of the data with ρ = 0.48, the remainder having ρ = 0.02, whilst HR Wallingford (2000b) reports a best fit of 6.3% of the data with ρ = 0.40, the remainder having ρ = 0.02.

It is not easy to assess the goodness of fit as some of the distribution parameters refer to normalised scales. However, the proportion of data assigned to each of the two BVNs and the correlation coefficient determined for each one can be compared visually both with the original data and with the rho-values reported by *BVN*. The fitted distributions seem correct in finding no correlation for the majority of the data, but a strong dependence amongst the highest 3.2 or 6.3% of the data.

6.2.5 Simulation of a very large sample of data from the fitted distributions

The JOIN-SEA *SIMBVN* and *SIMMIX* programs produce long-term simulations based on the original data set and the statistical distributions fitted by *BVN* and *MIX*, respectively. For this illustration the mixture dependence model was used with the top 6.3% of the distribution having a strong dependence and the remainder having very low dependence. Typically, the simulation length chosen is about ten times the highest joint return period of interest, for example a thousand year simulation being used to derive conditions up to a 100 year return period. However, a shorter simulation period might be used during program testing, or a longer final simulation to increase reliability.

Figure 9 is a scatter diagram showing the 662000 records produced during a thousand year simulation. It clearly has the same general distribution as the source data, but offers higher resolution of the extreme values.

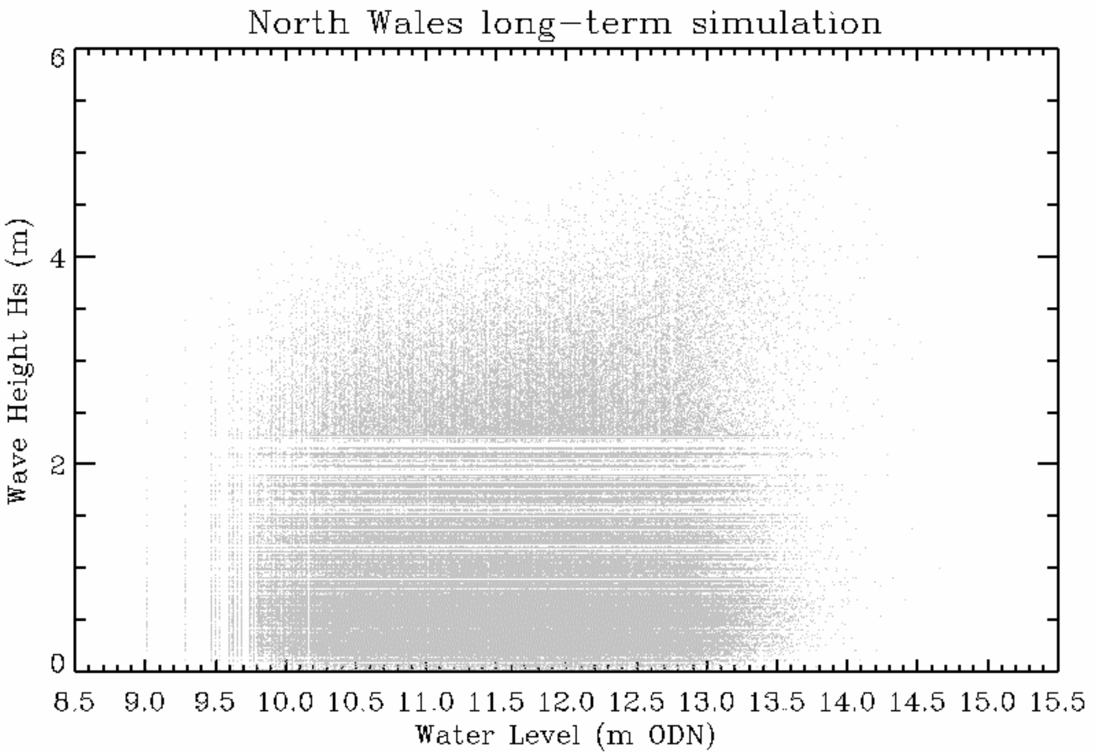


Figure 9 Coastal case study: long-term simulation of wave height & sea level

6.2.6 Simulation of structure variables; evaluation of joint exceedence extremes

The JOIN-SEA *ANALYSIS* program reads in the long-term simulation data and determines extreme values by counting back through the records. (For example the 100 year wave height would be the tenth largest value in a thousand year simulation.) On request it will list marginal extremes, joint exceedence extremes, and extremes of structure functions (e.g. overtopping or run-up) which can be calculated from wave height, wave period, sea level and any fixed structure parameters.

Joint exceedence extremes, for return periods of 1, 5, 20 and 100 years, are shown in Figure 10, overlaid on the source data.

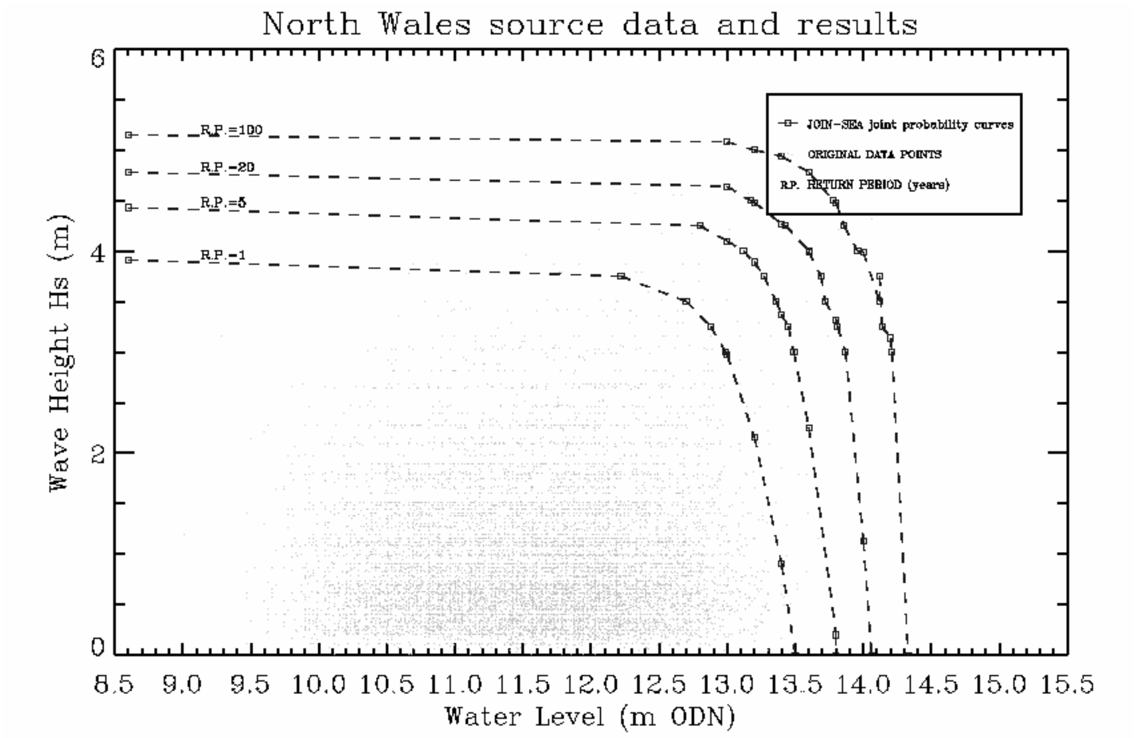


Figure 10 Coastal case study: joint exceedence curves overlaid on source data

6.2.7 Checking of results

1. Inspection of the scatter diagrams of the source data and the long-term simulation data show that they have approximately the same distribution. This is reassuring.
2. The numbers of sea levels exceeding the predicted 1, 5 and 20 year return period values in approximately thirteen years of source data, are 11, 2 and 1, respectively. This accords with expectations and gives confidence in the prediction of extreme sea levels from the source data.
3. The numbers of wave heights exceeding the predicted 1, 5 and 20 year return period values in approximately thirteen years of source data, are 11, 4 and 0, respectively. This accords with expectations and gives confidence in the prediction of extreme wave heights from the source data.
4. The numbers of records lying outside the 1, 5, 20 and 100 year joint exceedence contours in approximately thirteen years of source data, are 54, 9, 3 and 1, respectively. The final paragraph of Section 6.1.3 describes how to calculate the expected numbers, in this case as 13-84, 3-21, 1-6 and 0.1-1.4. This agreement gives confidence in the dependence fitting.

Note that Checks 2-4 above would have to be applied with care and understanding if the option to re-scale extreme values in the JOIN-SEA long-term simulation is used. Re-scaling would alter the predicted marginal and joint exceedence extremes, but would not alter the source data.

6.3 Case study involving tides and surges

For many years the Proudman Laboratory has been involved in predictions of extreme sea levels based on UK tide gauge measurements. During the 1990s, a major programme of analysis involved joint probability analysis of tide and surge data from

UK A class tide gauge sites, coupled with numerical tidal modelling to infer longer periods of data for locations between tide gauges. The results of this programme are given in Proudman (1997) in which extreme sea level estimates are presented at numbered grid points about 20km apart for England, Scotland and Wales. (The Severn Estuary and Bristol Channel were computed and presented on a finer grid.) Although these predictions are authoritative and easily accessible, it was not intended that they would be used in isolation without reference to additional local data and knowledge, particularly in areas away from the open coast such as the Solent, Wash, Thames and Morecambe Bay.

The steps involved in the intended usage of Proudman (1997) are illustrated here for a hypothetical location on the Scottish coast of the Irish Sea, protected from wave action, incorporating hypothetical local sea level information.

Decide the grid point location
With reference to Figure 8.15 of Proudman (1997), consider a hypothetical site one third of the way from Point 69 (also referenced as distance from Wick 3487km in Table 4.1 of Proudman (1997)) to Point 68 (distance 3448km). Take an average of results for Points 68 and 69, weighting the latter by a factor of two.

Extract the 1 year return period sea level
From Table 8.4 of Proudman (1997) the 1 year sea level is 4.90m relative to mean sea level in 1990. Table 8.6 indicates to add 0.23m to adjust to 5.13m relative to Ordnance Datum. Table 8.7 indicates a mean sea level trend of 0.1mm/year, which is insignificant between 1990 and 2004.

Extract the differences between the 1 year and higher return period sea levels
Table 8.3 of Proudman (1997) indicates that sea levels for return periods of 10, 25, 50 and 100 years, would be 0.43m, 0.58m, 0.67m and 0.83m, respectively, higher than the 1 year level.

Adjust for local sea level knowledge
Hypothetically, it might be known that a wall at the site, with a crest elevation 5.40mOD had been flooded 20 times in 50 years, suggesting a return period for 5.40mOD of 2.5 years. This would suggest that the 1 year return period level directly from Proudman (1997) would be 10-15cm low for this site, and so that it should be raised by, say, 0.15m.

Adjust for future climate change assumptions and summarise
A typical precautionary allowance for future sea level rise in this region would be 4mm/year, which over 50 years would mean a rise in sea level of 0.20m. The results can therefore be summarised as:

Return period years	1	10	25	50	100
Present-day sea level (mOD)	5.28	5.71	5.86	5.95	6.11
Sea level in 50 years (mOD)	5.48	5.91	6.06	6.15	6.31

6.4 Case study involving surges and river flows

6.4.1 Outline of the case study

The Severn was used as the main case study of extreme water levels in tidal rivers in a previous Defra-funded joint probability study. It is not described at length here as details are given in Defra / Environment Agency (2003) and as many of the steps are similar to those in the case study involving waves and sea levels described in Section 6.2. The Severn provides an interesting case study, as it includes areas affected only by sea level, or only by river flow, and areas affected by both variables, each of which areas is large enough to include gauging stations. It is well served by numerical models, knowledge of flood risks, many years of tide gauge data from Avonmouth and river flow data from Haw Bridge, and several years of water level data at several gauging stations.

The procedures for distribution fitting, long-term simulation, evaluation of joint extremes and checking of results are similar to those described in Section 6.2, and so are not described here. However, the procedures involved in the data preparation and water level calculation stages are different from those that would be used in coastal applications, and so are described here.

6.4.2 Selection, gathering and preparation of data

The intention was to combine the distributions and extremes of downstream sea level and upstream river flow with a good representation of the dependence between the two variables, to develop predictions of high and extreme water levels throughout the river. Preliminary river numerical modelling was used to determine a convenient downstream location below which river flow has only limited influence on water level, and a convenient upstream location above which sea level has only limited influence on water level. These two positions proved to be near to the Avonmouth tide gauge and the Haw Bridge river flow gauge, from which simultaneous sea level and river flow data were obtained for the period 1987-1998.

The appropriate event definition for extraction of (nearly) independent records from the sequential measured river flow and sea level data was not clear. This will be the case whenever the typical flow hydrograph duration is much above one day.

Had sea level been the dominant consideration throughout the study area, independent records might have been taken at every high water or on every spring tidal cycle. Had river flow been the dominant consideration, independent records might have been taken at intervals corresponding to the typical flood hydrograph duration of 7-10 days for the Severn. As both variables were important, a peaks-over-threshold definition might have been adopted, using a different threshold for each variable, causing high values of that variable to be extracted alongside simultaneous values of the other variable, whether high or not. A further consideration was that the record definition used had to include sufficient information for subsequent re-construction of an event in sufficient detail to evaluate water level at different positions within the river.

The compromise adopted was to divide the time series data into 14-day blocks corresponding to neap-spring-neap tidal cycles. One record was extracted from each block, consisting of the peak sea level and the peak river flow occurring during that period (with a manual check against 'double counting' of the same river flow event in consecutive records). These records provided the required variable-pair input to JOIN-SEA used for distribution fitting and long-term simulation. Note that as JOIN-SEA expects the input records to be in the form of wave_height / sea_level / wave_period, it was necessary to enter river flow in the wave_height position and to add a dummy 'wave_period', set to ten seconds for each record.

6.4.3 Evaluation of water level from a long-term simulation of river flow and sea level

The long-term simulation stage of JOIN-SEA produced thousands of years of records of river flow and sea level. It was found during the fitting stage that the data showed high dependence ($\rho = 0.7$) towards the top end of the distribution, which was consistent with the significant dependence ($\chi = 0.05$ and $\chi = 0.06$) between river flow and daily maximum surge found for flow stations in the Severn (Section 4.4.3).

In some circumstances, it would be possible to generate the main flood risk variable (water level in the river in this case) by a relatively simple function of the relevant source variable(s) (river flow and/or sea level in this case). For up to a few dozen cases of interest, it may be practical to run a hydraulic model simulation, but this would not have been so for the potentially hundreds of thousands of records in the long-term simulation. In this instance, an alternative structure function contouring approach (Defra / Environment Agency, 2003) was used.

A representative sample of combinations of high and extreme values of river flow and sea level was selected for use as input to a numerical river model of the Severn. The corresponding water levels at each position of interest within the model were contoured using fitted functions of river flow and sea level, retained to represent the water level function. The stored contours were then applied to each record within the range of interest in the long-term simulation, yielding an empirical distribution for water level. Extreme values were estimated from this distribution by countback analysis (e.g. within a 1000 year simulation, the average of the tenth and eleventh highest values provides an approximation to the 100 year return period value).

6.5 Case study involving urban drainage

6.5.1 Outline of the case study

Application of the analytical approach to short duration rainfall would have been impractical, at all but a handful of locations, until recently, because of the shortage of sequential hourly rainfall data. Many more hourly rainfall gauges have been deployed over the last five years, and so the type of analysis outlined here may come into use in the future.

For illustrative purposes, consider a hypothetical drainage flood risk analysis for a large coastal town, on the south coast of England. Stormwater drainage from the town discharges into the sea at a single point, and the drainage system would potentially be at risk of being unable to discharge stormwater runoff during intense rainfall coupled with a high sea level. To simplify the analysis, assume that this is the only possible failure mode of an existing drainage system. Test the system for standard of service based on two criteria, one representing the onset of street flooding, and the other representing the onset of significant flooding of houses.

6.5.2 Selection, gathering and preparation of data

The source variables of interest in this case study are sea level, at a single location representative of the drainage outfall, and short duration rainfall. The appropriate duration over which to characterise rainfall depends on the size and speed of response of the drainage system to rainfall. For illustrative purposes, two-hours is used here. As high intensity rainfall and high surge events tend to last for less than one day, a reasonable compromise between the need to use all of the source data and the need for successive records to be independent of each other is to take one record at each high tide. (As discussed in Section 5.4.6, the main difference in approach for urban drainage systems discharging into rivers, rather than into the sea, would come from the difficulty in choosing an appropriate method of sampling to characterise the dependency between high river level and high rainfall events.)

Had there been sufficient data close to the site, the data would have been prepared for joint probability analysis by extracting each high water level, and matching it with the highest rainfall intensity, averaged over two consecutive hours, within six hours of the time of the high tide. Instead, for illustrative purposes, an artificial five-year duration 'source' data sample was prepared, to be reasonably representative of the Dorset coast. The sample contained one record per high tide, i.e. 3535 records over five years. The distribution and extremes of sea level, and its dependence ($\rho = 0.32$) with rainfall, were based on those measured by the Weymouth tide gauge. High and extreme rainfall was based on the values derived in Section 5.4.3 for the south coast of England.

6.5.3 Estimation of standard of service from a long-term simulation of rainfall and sea level

Distribution fitting and Monte Carlo simulation of a thousand year sample were carried out as described in Section 6.1 and illustrated in Section 6.2. The 'source' and long-term simulation data sets are shown together in Figure 11, in the form of scatter diagrams of rainfall against sea level. In a real flood risk study, it may be desirable to use the option to re-scale high and extreme values during the long-term simulation, in order to achieve target marginal extremes for rainfall alone and for sea level alone, derived outside the joint probability analysis. This was not done here, as it would obscure the link between the source and long-term simulation data.

To illustrate how the long-term simulation could help in estimating flood risk, two hypothetical failure criteria are drawn as blue lines on each of the scatter diagrams in Figure 11. The area above the lower line ('onset of street flooding') represents combinations of rainfall and sea level where the drainage system would be over-loaded and water would begin to flood into the street. Similarly, the upper line ('onset of house flooding') represents the onset of significant flooding of houses. The positions of these lines might be based on past experience of flooding in the town, or on urban drainage numerical modelling indicating which input conditions cause flooding (and which do not). The probability of occurrence of such conditions can then be estimated by counting the numbers of occurrences above these lines in the scatter diagrams.

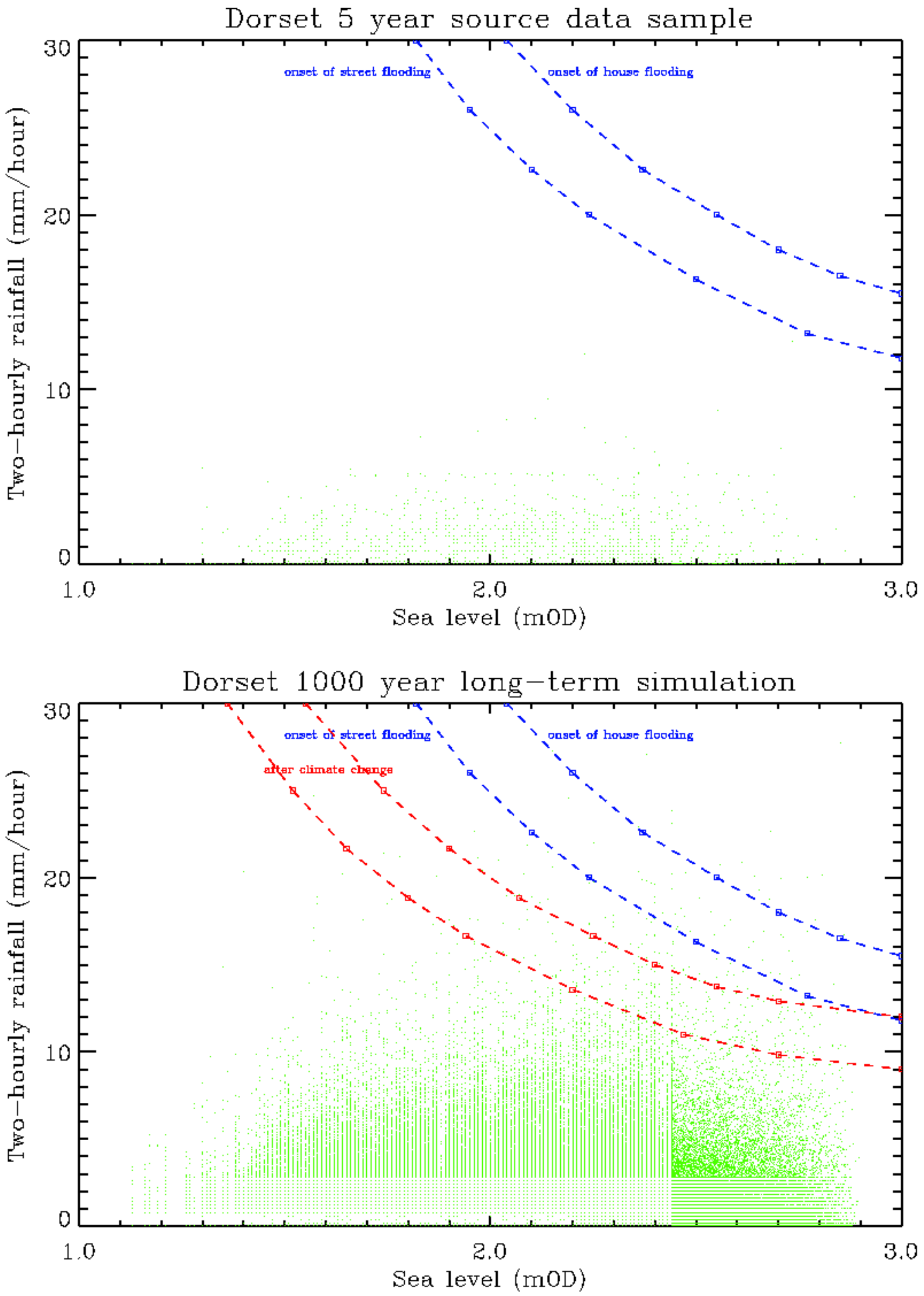


Figure 11 Drainage case study: source and long-term simulation data with overlaid failure curves (blue for present-day, red representing climate change)

The source data plot contains one data point close to the lower failure line, suggesting that street flooding would have been close to occurring once over a five year period, in turn suggesting a return period of a little over 5 years, but with a large margin of uncertainty. Similarly, the source data plot suggests that the return period for house flooding would be much higher than five years. The long-term simulation plot shows 57 points above the street flooding line, including 17 points above the house flooding line, in one thousand years. Although there may still be uncertainties about the source data, the extrapolation, and the drainage system modelling, the long-term simulation thus provides direct estimates of the standard of service, in terms of the two flood criteria, of 18 and 59 years.

6.5.4 Incorporation of climate change allowances

The sensitivity of the effectiveness of a flood defence or drainage system to future climate change can be tested by assuming that the source variables will change in line with current climate change projections but that the defence or drainage system will not change. If future climate change can be represented in terms of simple adjustments to the source variables, then the long-term simulation allows a direct calculation of the sensitivity of the failure probability to those adjustments.

For illustrative purposes, apply future climate change over the next fifty years consistent with the current Defra precautionary allowances, i.e. 20% increase in high and extreme rainfall intensity, and 0.3m increase in mean and extreme sea level on the south coast of England. In a real study, these two adjustments would be applied to every record in the long-term simulation, and the failure probabilities re-calculated. For ease of comparison in this case study, the source variables are left unchanged, but the two failure curves are re-plotted (in red in Figure 11) to reflect the climate change allowances (sea level 0.3m lower; rainfall divided by 1.2).

Using the same counting technique as in Section 6.5.3, the return period for street flooding would reduce from 18 years to 3 years, and for house flooding from 59 years to 8 years. One might conclude from this hypothetical assessment that, although the drainage system offers an adequate (20-60 year) standard of service at present, it would become inadequate (3-8 year standard) under the climate change scenario considered.

6.6 Case study involving a larger area

6.6.1 Issues involved in applying the analytical approach in a larger area

The potential difficulties described in Section 5.6 in relation to the desk study approach would also apply when attempting to use JOIN-SEA in a large area such as the Thames, where multiple source variables contribute to the overall flood risk. An additional difficulty is that simultaneous sequential data on all of the variables of interest may not be available to provide the required input to JOIN-SEA.

The options for addressing those difficulties are quite similar to those given in Section 5.6, except that there is a specific three-variable simulation version of JOIN-SEA, which eases the joint probability analysis of three primary variables. The three-variable version works in a similar way to the standard version, fitting extreme value distributions to each of the three individual variables, and Bivariate Normal distributions to each of the three variable-pairs, before producing a long-term simulation. As with the two-variable version, wave period can be included as a secondary variable, and wave direction or flow duration as a conditional variable. A longer description is given in Section 3.5 of Defra / Environment Agency (2003).

6.6.2 Introduction to the outer Thames numerical case study

The Thames estuary was chosen partly because it illustrates a number of points of general relevance to application of joint probability methods in larger areas. It was also chosen because of links to requirements and related work within the Environment Agency's ongoing Thames Estuary 2100 flood risk management programme. One of the aims of that programme is to obtain accurate estimates of overall flood probability and consequence, both for present-day and future conditions. This case study (HR Wallingford, 2004) provides a method and data set that could be used widely and on a consistent basis throughout the outer Thames estuary.

In the estuary, within about a ten kilometre radius of Southend, sea level (including surge) is the most important variable in assessing flood risk. Waves, both locally generated and those arriving from the North Sea, are also important. The impact of river flow on overall water level is relatively small, although not necessarily negligible. Conversely, moving upstream from Tilbury, river flow becomes more important, and sea level and waves less important in determining river water level, and so different methods would be used here.

6.6.3 Preliminary analysis of the outer Thames

Several source variables are of potential interest in assessment of flood risk in the Thames, some of them varying in magnitude across the study area, some of them with different durations, and having different dependencies between different variable-pairs. Preliminary analysis focused on reducing the number of primary variables for use in the joint probability analysis necessary to assist flood risk evaluation.

Initial runs of a river numerical model indicated that river flow had limited impact upon high water level downstream of Tilbury. Therefore river flow was not regarded as a primary variable, and a single fairly high flow rate could be assumed to provide adequate representation throughout all subsequent analysis.

Sea level is the most important variable throughout the area downstream of Tilbury. It is convenient to take sea level at a single point as a primary variable, with appropriate factors derived from a numerical tidal model to derive sea level at other locations, based on that primary variable.

Previous wave prediction and wave transformation model runs indicated that waves are important in estimating flood risk, and that realistic predictions could be achieved using a local wave prediction model with a wave generation area extending only a little outside the estuary. However, the shape of the Thames means that there is considerable variation in wave height throughout the study area, and dependence upon wind direction. This potential problem was solved by taking as a primary variable wind speed (sorted by wind direction sector) as a proxy for wave conditions. In other words, the statistical analysis would be done in terms of wind speed (for a limited number of direction sectors) and then wind speed would be used as input to a number of separate wave prediction models during flood risk evaluation.

The records of the source variables required as input to JOIN-SEA were defined by wind speed, wind direction and sea level. As river flow was not a primary variable, it was not necessary to take account of hydrograph duration, and records were taken at each successive high water, following normal coastal engineering practice.

6.6.4 The joint probability analysis undertaken for the outer Thames

Time series wind data (to be used in later wave modelling) and time series sea level data from Southend formed the basis of the joint probability analysis. Concurrent data were available for 11 years (1989-1999).

Times and sea levels at each high water were extracted from the Southend data set and coupled with wind speeds to form variable-pair records. Consecutive records were assumed to be independent of one another; a necessary requirement of the subsequent statistical analysis.

The records were then sub-divided into four separate sets based on wind direction. There were two reasons for this subdivision. Based on previous experience, it was anticipated that the dependence between wind speeds and sea levels would vary with wind direction, a feature that would not be captured in the analysis without division into direction sectors. Different wave prediction points which might be used in the estuary would be exposed to different wind directions, but for individual points on the south coast, the shape of the Thames is such that exposure could be reasonably approximated by ‘north-easterly’ or ‘north-westerly’. Hence it would be necessary to know the approximate wind direction for any given wind speed / sea level event. For example, if the wind is from the north-east, wave prediction points exposed to north-easterly winds would predict a wave condition (and hence a wave overtopping rate), whilst other wave prediction sites, not exposed, would assume zero wave height.

Sensitivity tests on the most appropriate directions in which to separate the data were carried out, the results of which showed the natural separation of north-east (i.e. winds from 0-90°N), south-east, south-west and north-west, to be sensible.

The wind speed and sea level data for each direction sector were run through the first stage of JOIN-SEA. Sea level and wind speed data in the north-east sector were largely independent, although a number of well-correlated observations were apparent. The south-east and south-west sectors both showed independence. The data for the north-west sector were distinctly skewed, indicative of a positive correlation between high sea levels and high wind speeds. The majority of sea levels greater than 3.5mOD occurred when the wind was from the north-west.

Prior to long-term simulation using JOIN-SEA, the extreme sea levels were re-scaled to more reliable values based on Proudman (1997) than those obtained from just 11 years of data from one gauge. However, as the JOIN-SEA analysis was divided into four categories based on wind direction, the re-scaling of these separate direction sectors was not straightforward. When re-scaling sea levels for individual direction sectors, consideration of the total distribution of high sea levels is required. That is to say when the simulated data from the four direction sectors are added together the created distribution of sea levels should be the same as the target overall distribution.

The north-east and north-west sectors contained the majority of high sea level events, and so the extreme values based on Proudman (1997) were considered appropriate for these direction sectors. The south-west sector contained fewer significant high sea level events, and so the appropriate extreme values were not obvious. However, it was thought, on balance, that this direction sector did contribute significantly to the overall distribution of extreme sea levels, and so the values based on Proudman (1997) were again used. The south-east sector was considered an exception and the long-term simulation was run without re-scaling.

It would also have been possible to re-scale the high and extreme wind speeds. However, since the extreme values predicted from 11 years of data for the prevailing south-west direction agreed well with values published in offshore design guidelines, no re-scaling was applied.

6.6.5 Intended use of the outer Thames data set in flood risk studies

The long-term simulation stage of JOIN-SEA was run separately for each of the four direction sectors to produce the proportion of 100000 years of data appropriate for each direction sector. These four separate simulations were combined to form an overall 100000 year data set of sea level (at Southend), wind speed (over the Thames Estuary) and wind direction sector.

This combined data could be applied in a consistent way at different points throughout the outer Thames (in combination with numerical wave and river modelling) and throughout alternative climate scenarios (if defined in terms of simple adjustments to wind speed and sea level). As water level is not sensitive to river flow in the outer estuary, it is sufficient to use a single representative flow rate where necessary. The advantage of using this single simulated data set is that it maintains spatial homogeneity and temporal consistency throughout the study area and across different climate scenarios.

Site-specific flood risk assessment usually focuses on the flood risk variables such as waves and sea levels) and direct structural responses such as overtopping rate or structural failure. The defence design process is then focused on reducing to an acceptable level the probability of a structure variable exceeding a specified (failure) value. Often, there is an explicit functional relationship between the flood risk variables and the structure variable, for example in the form of wave overtopping rate equations. In this situation, the probability of failure can be estimated, from a long-term simulation of the source variables, through direct evaluation of the structure variable.

In complex areas such as the Thames, it may be better to consider the overall flood risk, taking account of discrete (perhaps partially linked) floodplain areas, each protected by a different series of different types of defences, all of which are subject to inter-related loading conditions. In this situation, it may be helpful to determine both the probability and the consequences of defence failure, in terms of economic losses associated with flooding. The structure function (economic loss) is no longer an explicit function of the source variables, but is implied, through a series of complex modelling procedures. In this situation, an approximation to the structure function can be developed as a contouring function based on a limited number of hydraulic and economic loss model runs. Each record in the long-term simulation of the source variables is then converted to the structure variable through application of the contour function. This procedure has been used within the Environment Agency RASP and National Flood Risk programmes, and will be used within the Thames Estuary 2100 programme. It enables a full assessment of flood risk for a range of loading conditions and defence failure scenarios.

6.6.6 Related work within Thames Estuary 2100 Phase 2

Developments within the present project will be applied within TE2100 Phase 2 under a specific project on extremes and joint probability analysis. The scope of the project is outlined here as some parts may be of generic interest for studies in other complex areas. The project will provide information to other tasks within TE2100 to help address the potential range of different flood risks and management responses in the Thames tideway and its tidally affected tributaries.

A significant preparatory stage, before any statistical analysis is undertaken, will be to decide which variables and data really need to be analysed. Initially, this will involve a review of which source variables are of greatest interest elsewhere in TE2100, what measurements are available for each one, and which could be treated either as ‘secondary’ or as ‘conditional’. The hope is that the list can be reduced to about ten primary data sets for detailed analysis, e.g.

- 3 fluvial, Thames and two tributaries
- 1 wind, over-water in the outer Thames
- 1 or 2 tidal, including Southend
- 1 or 2 waves, possibly measured in ‘the Warp’ and/or hindcast within TE2100
- 2 or 3 rainfall, including one north and one south of the Thames.

The main analysis stage will be in four parts.

1) Analysis of marginal extremes for each data series, and assessment of their year-by-year variability.

2) Analysis of dependence between relevant variable-pairs (or ‘triads’), with additional conditional variables (e.g. wind duration and direction) where relevant. Initial thoughts are that these might include:

- 1 fluvial / tidal / wind
- 1 tidal / waves
- 2 rainfall / fluvial
- 1 main river fluvial / tributary fluvial.

3) Clarify climate change allowances to be applied to tidal, fluvial, rainfall, waves and wind. Initial thoughts are that the scenarios described might represent:

- 4 based on the UKCIP02 scenarios
- 1 based on the Defra precautionary allowances.

4) Transfer of source data and analysis results into the TE2100 GIS system.

7. REFERENCES

Defra (2003). Supplementary note on climate change considerations for flood and coastal management. http://www.defra.gov.uk/environ/fcd/pubs/pagn/Climatechangeupdate.pdf. *Current appropriate precautionary allowances for the possible effects of future climate change.*

Defra / Environment Agency (2002). Risk, performance and uncertainty in flood and coastal defence: A review. Defra / Environment Agency Report No FD2302/TR1 (also referenced as HR Wallingford Report SR 587). *Definitive background reading on the definitions, evaluation and implications of risk and uncertainty in flood management.*

Defra / Environment Agency (2003). 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 Report SR 645). *In-depth additional information for regular users of joint probability methods in tidal rivers.*

Defra / Environment Agency (2005a). Joint probability: Dependence mapping and best practice: Technical report on dependence mapping. Defra / Environment Agency R&D Technical Report FD2308/TR1 (abbreviated to FD2308/TR1 in this report, and also referenced as HR Wallingford Report SR 623). *Essential information additional to the present report for frequent users of joint probability analysis.*

Defra / Environment Agency (2005b). 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 (abbreviated to FD2308/TR3 in this report, and also referenced as C Svensson and D A Jones (2005). Dependence between extreme sea surge, river flow and precipitation: a study in south and west Britain, CEH Wallingford Report*). In-depth additional information for hydrologists interested in dependence.*

HR Wallingford (1997). Swell and bi-modal wave climate around the coast of England and Wales. HR Report SR 409, November 1997. *Source of swell data and in-depth additional information for regular users of swell and bi-modal wave conditions.*

HR Wallingford (1998). Impact of bi-modal seas on beaches and control structures. HR Report SR 507, February 1998. *In-depth additional information for users of swell seas in coastal engineering.*

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. *In-depth additional information for users of JOIN-SEA.*

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. *Essential additional information for users of JOIN-SEA.*

HR Wallingford (2002). Coastal defence vulnerability 2075. HR Report SR 590, February 2002. *In-depth additional information on the impact of climate change on coastal flood risk.*

HR Wallingford (2004). Joint probability issues within estuaries: A numerical case study for the tidal Thames. HR Wallingford Report TR 143, August 2004. *In-depth additional information on the outer Thames waves and sea level analysis.*

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

Natural Environment Research Council (1975). Flood Studies Report. Institute of Hydrology. *Source of extreme rainfall data and in-depth additional information for regular users of extreme rainfall predictions.*

Proudman Oceanographic Laboratory (1997). Estimates of extreme sea conditions: Spatial analyses for the UK coast. Proudman Oceanographic Laboratory Internal Document No 112. *Source of extreme sea level data and in-depth additional information for regular users of extreme sea level predictions.*

PB 11209

Nobel House,
17 Smith Square,
London SW1P 3JR

www.defra.gov.uk

Department for Environment
Food and Rural Affairs