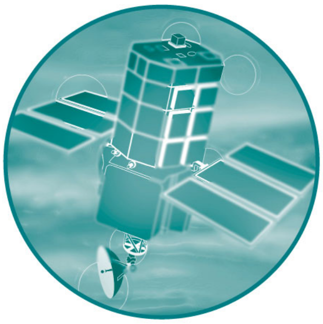
Defra/Environment Agency
Flood and Coastal Defence R&D Programme



Extreme Water Levels in Estuaries and Rivers

The combined influence of tides, river flows and waves

R&D Technical Report FD0206/TR1

defra
Department for Environment
Food and Rural Affairs

Environment Agency

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

Extreme Water Levels in Estuaries and Rivers
The combined influence of tides, river flows and waves
R&D Technical Report FD0206/TR1

December 2003

Dr P J Hawkes

Publishing organisation

Defra Flood Management Division
Ergon House
17 Smith Square
London SW1P 3JR
Tel: 020 7238 6178
Fax: 020 7238 6187
www.defra.gov.uk/environ/fcd

Dissemination Status
Internal: Released Internally
External: Released to Public Domain
This report is also catalogued as HR Wallingford Report SR 645

Statement of use
This document provides information for Defra and Environment Agency Staff about the prediction of extreme water levels in estuaries and rivers using joint probability methods, and constitutes an R&D output from the Joint Defra / Environment Agency Flood and Coastal Defence R&D Programme.

- Keywords: flood risk, dependence, joint probability, extremes, waves, sea level, surge, tide, river flow, estuaries
- Name, address and contact details of the research contractor: Dr Peter Hawkes, HR Wallingford, Howbery Park, Wallingford, Oxon OX10 8BA, email pjh@hrwallingford.co.uk.
- Client project manager: originally John Goudie of MAFF (now Defra) and then Peter Allen-Williams, Defra, Ceres House, 2 Searby Road, Lincoln, LN2 4DW, email peter.allen-williams@defra.gsi.gov.uk.

Contract Statement

This work was funded by MAFF (now Defra, the Department for the Environment, Food and Rural Affairs) under project code FD0206 (Joint probability of extreme estuarine water levels). Publication of this report implies no endorsement by the Department of the report's results or conclusions. The Project Manager for HR Wallingford was Dr Peter Hawkes; the analysis was undertaken by Ben Gouldby and the report written by Peter Hawkes. The HR Wallingford job numbers were CCS14W, CCS24W, CCS34W and CCS44W. Although the project predates the joint Defra / Environment Agency combined research programme, completion of this report falls within the programme.

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre, c/o
WRc, Frankland Road, Swindon, Wilts SN5 8YF

tel: 01793-865000 fax: 01793-514562 e-mail: publications@wrcplc.co.uk

EXECUTIVE SUMMARY

The overall water level in estuaries and rivers may be dependent upon river flow, astronomical tide, surge, waves and wind. For situations where two or more of these variables are important, this report addresses the issue of how those variables should be combined, before input to hydraulic models or design methods. It would be relatively easy to deal with the combination of either independent or fully dependent variables, but in practice the variables are usually slightly correlated to an extent best determined from site-specific data.

Joint probability methods developed for the combined action of large waves and high sea levels on sea defences were adapted to the potentially more complex problem of the combined action of river flows, sea levels and waves in estuaries and rivers. No new hydraulic modelling techniques were developed, but the statistical methods developed, tested and validated should allow better estimation of the frequency of occurrence of extreme conditions driven by two or more of the primary input variables.

The present project followed on from several years work on joint probability analysis methods aimed at refining predictions of extreme sea conditions at the coast. These methods have now come into routine use in HR Wallingford's coastal studies, and have been disseminated for use by other UK coastal engineering consultants. Further development of the methods to the more complex situation of overall water levels in estuaries was a natural continuation of the joint probability theme, funded by MAFF (now Defra) over a 3-year period beginning in Summer 1997.

Preliminary exploratory analysis is an important part of the joint probability approach, to check that all relevant input variables are included in an appropriate way, and to discover if any potentially relevant variables can be excluded from full analysis without reducing accuracy. This preliminary stage also investigates dependences and time lags between peak values of variables, and determines a suitable 'event' definition needed for data handling.

The statistical analysis determines the distributions and extremes of the input variables, and any dependences between them, before using these fitted distributions to synthesise a very large sample of data with the same joint probability density as the original input data. This sample can then be used to derive joint exceedence extremes, if required. The extrapolated joint density or the joint exceedence extremes are converted to end variables of interest (overall water level, failure, economic loss etc) using hydraulic models or structure function formulae.

These joint probability developments are not intended to replace better established methods for flood studies, but rather to assist in selecting appropriate combinations of inputs, and to refine the corresponding prediction of overall return periods.

For further information on this project, please contact Dr Peter Hawkes of the Coastal Group at HR Wallingford.

CONTENTS

Table

Figure

Appendix

Appendix 1 The Severn Estuary Case Study

GLOSSARY

Auto-correlation
The correlation of one variable with itself, as a function of time lag

Bi-Variate Normal (BVN) distribution
A two-variable joint distribution, each one having a Normal marginal distribution

Continuous simulation
Continuous (in time) hydraulic modelling to hindcast the time series of a structure variable

Correlation (coefficient)
A linear form of dependence specified by a coefficient between -1 and +1

Cross-correlation
The correlation between two variables, optionally as a function of time lag

Dependence (function)
The extent to which one variable depends on one or more other variables

Event (in context)
A noteworthy occurrence, implying a degree of selection from amongst records, often defined as a record in which a threshold of interest is exceeded within a given period of time

Extremes
Very high values, often outside the range of an observational sample

Flood plain
The area alongside a river that takes up the out-of-bank flow

(River) Flow
Primary variable, being the volume per time flowing through a river cross-section

Generalised Pareto Distribution (GPD)
A standard extreme value distribution

Hydraulic model(ling) (in context)
Numerical modelling driven by flow and sea level

Hydrograph
A graph of river flow against time, used to represent extreme flow events

iSIS
A numerical river model

JOIN-SEA
A joint probability analysis method

Joint density
The probability that two related variables will simultaneously lie in specified ranges

Joint exceedence extremes
Refers to the probability of two or more variables simultaneously exceeding given values

Joint probability
Refers to the simultaneous probability of two or more variables

(Time) Lag
The time difference between comparable observations of two variables or between one variable and the response

Marginal (distribution/extremes)
Refers to the probability of just one variable

Normal distribution
A standard distribution, symmetrical about its mean value

One-, two- or three-variable
Indicates the number of variables involved in a method, distribution, result etc

Out-of-bank flow
Refers to flow in the flood plane when a river overtops its banks

Preliminary analysis (in context)
Study phase in which the relative importance of different variables is assessed and the analysis approach is chosen

Primary variables (in context)
Sea level, flow and wave height

Proxy variables (in context)
Surge and wind speed

Record (in context)
Record of one or more variables at a particular time and place, regardless of the values of the variables

Records per year
The average number of records (some of which may be events) per year

Return period
The average time between occurrences of a particular extreme event

(Flood) Risk
Combines the probability and consequence of occurrence of a particular event, usually in terms of impact

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

Secondary variables (in context)
Wave period, wind set-up, wave set-up

Sensitivity
The extent to which a response (e.g. risk) depends on one or more input variables

(Long-term) Simulation (in context)
A very large sample of synthesised data with the marginal and dependence characteristics of the original smaller sample

Statistical model(ling)
Mathematical modelling of probability distributions

Structure function/variable
Refers to a response (e.g. overall water level) caused by multiple input variables linked by an equation or by hydraulic modelling

Surge
Sea level minus predicted tide, indicating the non-astronomic component of sea level

Threshold (in context)
The value of variable(s) above which statistical modelling is applied

(Astronomical) Tide (or tidal level)
Primary variable, being predicted sea level due to astronomical tide

Tri-Variate Normal (TVN) distribution
A three-variable joint distribution, each one having a Normal marginal distribution

Uncertainty
Represents the extent to which the exact value of a variable cannot be known

Variable–pair
Refers to the two variables to which joint probability analysis will be applied

(Overall) Water level (in context)
Referring to water level within an estuary or river due to the combined effect of all environmental variables

Wave height
Primary variable, representing the short-term variability of the water surface elevation

Wave period
Represents the time between passing of individual waves

Wave set-up
Localised increase in the overall water level within the surf zone (if any)

Weibull distribution
A standard extreme value distribution

Wind set-up
Regional increase in water level directly due to wind stress

1. SCOPE OF THE PROJECT

1.1 Background

MAFF (now Defra) has been funding joint probability research for several years, focusing previously on large waves and high sea levels, astronomical tides and surges, and confluence flows. The present project extends the joint probability theme to determination of extreme water levels in estuaries. This situation has an additional complication compared to similar predictions on the coast, in that river flows and their joint occurrence with high sea conditions (due to astronomical tides, surges and waves) may need to be considered.

Astronomical tides can be predicted reliably around most of the UK, and associated tidal levels can be forecast many years ahead. There is some information on surges, flows and waves around most of the UK, but forecasts can be made at best only a few days ahead, and predictions of extremes at nearshore and inland locations are usually done on a site-by-site basis as the need arises.

The joint probability of astronomical tides and surges in determining extreme sea levels was the subject of MAFF-funded research at the Proudman Oceanographic Laboratory and at Lancaster University (POL, 1997). The joint probability of waves and sea levels was the subject of MAFF-funded research at HR Wallingford and Lancaster University (HR, 2000a and 2001). Robust and general methods were developed for use in coastal engineering studies. Dependence between several variable-pairs relevant to flood risk assessment (e.g. wave height & sea level, river flow & surge, wind-sea & swell) was mapped around England, Wales and Scotland (Defra / Environment Agency, 2003a) as part of another ongoing Defra-funded joint probability project. Risk and uncertainty in flood prediction has been studied in several recent research projects, (e.g. Defra / Environment Agency, 2002) and is a continuing theme in Defra's research priorities.

The Institute of Hydrology (now CEH Wallingford) has undertaken MAFF-funded research on the joint probability of combined fluvial and marine flooding and of the overall extreme water levels at river confluences (Jones, 1998). It has also undertaken MAFF-funded work into the dependence between river flow and surge around England, Wales and Scotland (Svensson and Jones, 2000; Defra / Environment Agency, 2003b). HR Wallingford has assessed the joint effect of river flows and sea levels during several site-specific flood risk studies for the former National Rivers Authority. Examples are the Severn Estuary, Truro, Whitby, and the Rivers Arun, W Cleddau, Colne (Essex), Gt Ouse, Rhymney, Ribble, Stour (Dorset), Taff, Tees, Thames and Trent.

The intention of the present project was to combine the best of the existing approaches into a new generally applicable methodology for assessment of overall extreme water levels in estuaries and tidal rivers.

1.2 Outline of the project

1.2.1 Overall outline

HR Wallingford managed and led the research; the Centre for Ecology and Hydrology (Wallingford) was involved as a sub-contractor; Professor Jonathan Tawn from

Lancaster University provided specialist input; Mr Michael Owen and Mr Ian Meadowcroft were involved in an advisory capacity. The research was undertaken over a 3-4 year period 1997/2001, in three overlapping phases.

The project was intended to address:

- the best sources of data and/or prediction/simulation methods for sea levels, flows and waves;
- the dependence between the three primary variables;
- extrapolation of primary variable distributions and dependence functions;
- high and extreme values of overall water level;
- the representation of a flood plain within the methodology for extreme estuarine water levels;
- uncertainties involved;
- practical and acceptable guidelines for using the methodology.

1.2.2 Outline of Phase I

Phase I was undertaken between about September 1997 and January 1999. It involved mainly theoretical developments.

- review of data sources and requirements;
- review and development of simulation and prediction methods for single variables;
- development of methods for assessing and representing the degree of correlation between the variables;
- review and development of methods for extrapolating marginal distributions and dependence functions;
- review and development of methods for extrapolation of two- and three-variable joint probability density;
- development of prediction methods for overall extreme water levels.

1.2.3 Outline of Phase II

Phase II was undertaken between about November 1998 and November 1999. It involved further development, testing and preparation for dissemination of the methods, in particular:

- testing and validation of the methods to ensure that they are practical and accurate for use in consultancy work;

- development of guidelines and examples to assist in application of the methods;

- brief testing of ‘sensitivity’ and ‘what if?’ scenarios to provide extra guidance on use and reliability of the methods, e.g.:

    - uncertainty in the input data;

    - use of simpler prediction methods where data and/or budget are limited;

    - incorporation of flood plain effects;

- assessment of the uncertainties involved in the methods.

1.2.4 Outline of Phase III

Phase III was undertaken between about September 1999 and March 2001. It involved case studies, refinement of techniques, and report writing, including specifically:

- completion of the Severn and other case studies;

- refinement of preliminary analysis methods, issues and decisions, including ‘event’ definition;

- production of an interim project report, focusing on the issues and methods involved in preliminary analysis, variable and data selection, joint probability analysis and interpretation of results, using elements of several case studies for illustration.

1.3 Outline of this and related reports

Chapter 2 of the present report summarises the issues and variables in the project, and the decisions and alternative analysis methods for extreme estuarine water levels. Chapter 3 discusses data requirements and availability for each variable in turn. Chapter 4 introduces the range of alternative joint probability methods, and how a choice might be made based on data availability and position within the estuary. Chapter 5 presents some case studies. Chapter 6 contains the discussion and recommendations for use of the methods.

Some of the developments in joint probability analysis technique during the present project were partly funded under a parallel MAFF-funded research project at HR Wallingford (FD1704, joint probability industry testing and dissemination). More details, for example on the methods for three-variable simulation and for statistical uncertainty estimation, are given in HR Wallingford (2001).

As part of the same commission under which the present report was prepared (FD0206, joint probability of extreme estuarine water levels) the Centre for Ecology and Hydrology at Wallingford undertook an analysis of surge, river flow and rainfall for eastern Britain. About seventy time series, each of about thirty years duration, were examined to determine any correlations between the highest values in data sets both of the same type (e.g. surge/surge) and of different type (e.g. surge/rainfall). A dependence parameter was evaluated for each data set pairing, both as a function of time lag and as a function of distance between measurement stations. Svensson and Jones (2000 and 2002) present the results, together with an analysis of their significance and possible meteorological explanations. During subsequent Defra-funded work, the CEH Wallingford analysis was extended to cover the south and west coasts of Britain (Defra / Environment Agency, 2003a and 2003b). The results may provide a method for estimating the dependence between different inputs to later joint probability consultancy studies without the need for site-specific analysis of local data.

At the same time as the present study, Halcrow Water led a project called 'Forecasting extreme water levels in estuaries for flood warning' for the Environment Agency (1999 and 2000). Although superficially a similar topic to that of the present report, there is little common ground between the two, as the Agency project focuses on operational forecasting issues, while the present project focuses on statistics and extremes predictions.

Following on from the present study, HR Wallingford led a further Defra-funded project called 'Joint probability: Dependence mapping and best practice'. The project has so far produced two reports (Defra / Environment Agency, 2003a and 2003b) on variable-pairs relevant to flood risk assessment, mapping dependence around England, Wales and Scotland, and providing a source of information on dependence for subsequent site-specific joint probability studies.

2. PROCEDURES INVOLVED IN A TYPICAL APPLICATION OF THE METHODS

2.1 Outline methodology

2.1.1 The problem to be addressed

The task is to determine extreme water levels in estuaries and rivers due to the combined effects (where applicable) of tide, surge, river flow and waves. Predictions may be for a single location or for a length of the bank. The introduction of river flow, where potential flood events may persist for several days, and the time lag between heavy rainfall and increased water level, poses additional problems to those which arise in the application of joint probability methods to waves and sea levels at the coast.

2.1.2 Appreciation of the site

A physical appreciation of the issues at a particular site is essential. This will be based on general and local experience, previous reports, possibly some preliminary hydraulic modelling, data gathering and a site visit to check the defence crest level and the likelihood of out-of-bank flow. The purpose is to determine what data and hydraulic models are available or need to be prepared, which input variables (e.g. sea level, flow and waves) are important at the site, and to form a preliminary opinion about the need for flood defence improvements. Potential three (or more) variable problems can usually be simplified during stage of a study.

2.1.3 Gathering of data

It is prudent to determine what data, reports and models already exist for a site, even if a conscious decision is then made not to gather or use them. Some of the site-specific issues may have already been addressed in previous studies, there may be direct measurements of water levels near the site, or it may be necessary to commission new measurements in preparation for model validation. Depending on the analysis approach to be used, it may be necessary to acquire time series data on the relevant input variables.

2.1.4 Preliminary analysis

If not already determined, some preliminary analysis may help to decide which input variables are important. River numerical modelling might be used, for example, to show that water level at the site is either not sensitive to river flow or not sensitive to sea level. Statistical modelling might be used, for example, to show that surge should be considered separately to tide, or that waves and river flow are uncorrelated, or that correlation is more significant when a time lag is assumed. The value of the project and the budget for the study will also be important considerations.

2.1.5 Event definition

In the case of waves and sea levels at the coast, each high water can reasonably be taken to be an independent record with the potential to cause flooding. This convenient assumption is usually inappropriate when river flow is involved, but various alternative event definitions are possible, based around peaks and durations of river flow hydrographs.

2.1.6 Range of statistical methods

Several alternative statistical analysis methods were described during this study. The key distinguishing features between the methods are in terms of the number of input variables (and whether any can be combined before statistical analysis) and the form of the predictions. Results can be expressed in terms of extrapolated combinations of the input variables or can be converted, using either an equation or a hydraulic model, to a single ‘structure variable’.

2.1.7 Checks on the results

As with any extremes predictions, a ‘counting analysis’ check should be made against the original data. For examples, if there are ten years of source river flow data, the predicted 1-year return period flow should be exceeded about ten times, and if flooding is predicted at a 10-year return period then there should be at least anecdotal evidence of it occurring roughly once every ten years. There are some slightly more complicated procedures that can be applied to joint exceedence predictions (HR Wallingford, 2000a).

2.1.8 Interpretation of the results

Joint probability analysis and extrapolation is usually done in terms of the distributions and extremes of the input variables. It is usually not trivial to infer the distribution and extremes of some structure variable, dependent on the predicted input variables. It may be necessary to use a hydraulic model to convert from sea level, flow and waves to overall water level, or formulae to convert from sea level and waves to overtopping.

In the case of joint exceedence extremes predictions, there are multiple combinations of the input variables with the same return period, and some caution is required in applying them. Also, the return period of any structure variable derived from those predictions is often significantly less than the joint exceedence return period.

2.2 Preliminary analysis

2.2.1 Decisions to be made

Exploratory analysis of the site and of the structure function plays a key role in deciding what form of methods should be used. The following questions should be considered (even if perhaps not resolved) for each new site or area, before undertaking detailed analysis or modelling:

- What end-variable(s) (e.g. water level, flood area, economic loss) is/are required?

- What general level of modelling and analysis required?

- What data, reports, established extreme values etc already exist?

- Which input variables (e.g. river flow, tide, surge, wind, waves) are important?
- Which input variables are correlated with each other (with time lag if appropriate)?
- Can any input variables be combined or considered as being of secondary importance?
- What is a typical hydrograph duration?
- Can out-of-bank flow occur?
- Will river control structures be used?
- What event definition is appropriate?
- In what form will the extremes results be expressed?
- What checks will be made on the extremes results?
- How will the extremes results be used to determine the end-variable(s)?
- Are any source data to be gathered or generated, for analysis or validation?

2.2.2 River modelling

Some preliminary runs of a numerical river model, even with unrealistically high input variables and/or a simplified river structure, may help to determine which input variables are most important. If overall water level at the study site does not respond to changes in downstream sea level, that would suggest that sea level is unimportant. If it responds differently to the same peak downstream sea level composed firstly of astronomical tide and secondly of astronomical tide plus surge, that might suggest that tide and surge should be considered separately. If it does not respond to changes in upstream flow, that would suggest that flow is unimportant. It if responds to differences in hydrograph duration, that would suggest that duration will be important in determining the event definition.

2.2.3 Dependence analysis

Before embarking on a complex multi-variable analysis, it is useful to undertake auto-correlation (single variable) analysis, and cross-correlation analysis for any relevant variable-pairs. This provides an indication of cycles within individual variables, for example on a tidal, spring tidal or seasonal scale; also an indication of dependences between other variable-pairs, possibly with a time lag between them. This information can be helpful in assessing whether surge should be treated as a primary variable, whether there is any trace of tidal level in upstream flow, and whether there is any trace of flow in downstream water level. Also, if one variable is shown to be independent of the others, it may be possible to perform a full joint probability analysis of the other variables, and to re-incorporate the independent variable in later calculations.

2.2.4 Data gathering and preparation

Time series data and/or distributions of variables and/or established extreme values may need to be gathered or generated for preliminary analysis, full analysis or validation purposes. Data requirements, if any, should be assessed early in any joint probability study.

Assuming that sufficient data have been gathered and that a statistical approach is appropriate, then probably the most important consideration in data preparation is event definition, in terms of which input variables are needed and the interval between records. If possible, successive records should be independent of each other, and should comprise sufficient variables to describe a potential flood event at the site of interest. There is not necessarily a 'correct' event definition for a particular study, and plausible alternatives should all produce similar end results. One example would be simultaneous values of river flow, sea level, wave height and wave period, taken either near the peak of each tidal cycle or near the peak of each flow hydrograph.

2.3 Choice of statistical methodology

The main points to consider when selecting from amongst alternative statistical methodologies are the number of input variables (usually one, two or three, with wave period as an optional extra) and the form of results required, usually continuous simulation, joint exceedence extremes or extrapolated joint probability density. More detail on the different analysis approaches is given in Chapter 4, and Figure 1 offers some guidance as to which one might be selected.

2.4 Form and interpretation of results

Continuous simulation uses a continuous (over several years) feed of information on the input variables, passed through a structure function or hydraulic model, to produce a continuous stream of structure variable predictions. This form of results directly addresses the end variable of interest, but may be limited by the period of simultaneous input variable data available.

Joint exceedence extremes consist of combinations of structure variable values, each of which is expected to be simultaneously exceeded once, on average, in a given return period. This form of results is commonly used in coastal engineering, and has the advantage that it can be re-used in separate applications to several different end variables. Disadvantages are that joint exceedence predictions are multi-valued (i.e. there are several combinations of input variables with the same return period) and that the joint exceedence return period provides only an approximation to the return period of any derived structure variable.

Joint probability density results consist of a very long simulation of records having the same form as the input records, but including more extrapolated values. Although much larger and less readable, this is the most flexible form of results in that it retains all of the information, and that it can be re-used for different structure variables, for which it provides a direct estimate of the return period. Both joint exceedence and joint density forms will usually require either a structure function or a hydraulic model to convert between input variables and the end variable(s) of interest.

3. DATA AND STATISTICAL METHODS

3.1 Data types and requirements

The variable that this project seeks to predict more accurately is the maximum overall water level at particular locations in an estuary or river. If there happens to be a water surface elevation gauge in the vicinity it may be possible to work directly from its measurements. Otherwise it is necessary to estimate the increase in water level at particular points induced by a series of other variables such as sea levels, river flows and waves, possibly measured or synthesised at other locations.

Assuming that water level data directly at the site are not available, the three primary variables to be analysed are sea level (meaning astronomical tide *plus surge*), river flow and wave height (although there are situations where only one or two of these are relevant). Additional variables, usually of secondary importance, include wave period, wind set-up and wave set-up. In some cases it may be more appropriate to work in terms of proxy variables, for example wind speed in place of waves, or surge in place of sea level.

Data gathering and analysis is important in all studies, but earlier experience with joint probability analysis suggests that a full statistical treatment is probably justified only where there are at least three years of reliable simultaneous data on all the relevant input variables. The data need not be particularly close to the site of application, but should be such that site conditions can be inferred using modelling techniques and such that the degree of correlation with other variables is well represented.

In statistical analysis it is often more efficient to work in terms of the original variables, even if of different types, as this retains the greatest knowledge of the problem, and later to convert to end products such as the overall induced water level. This can be harder to do, and provides more scope for errors in interpretation, but offers the possibility of greater insight into the problem and more faith in the extrapolations. General statistical methods for extremes analysis in one-, two- and three-variable situations are described in Sections 3.3-3.6. A range of approaches specific to estuarine water levels are considered in Chapter 4, in some cases converting to overall water level before statistical analysis and in some cases after.

3.2 Data sources

3.2.1 Primary variables

Wave height

In assessing extreme wave heights in an estuary it is probable that some wave energy from the open sea will propagate into the estuary. Waves generated by local winds will combine with the externally generated waves to give overall wave conditions. Wave conditions are not often measured in estuaries, and even where they are they may be representative of only one location. Other standard sources of wave data, i.e. satellites, ships and large scale meteorological models are also unlikely to be available in estuaries. Therefore the most likely source of wave data will be a wave transformation model to predict the propagation of waves generated in the open ocean, into the estuary, perhaps combined with a local wave generation model to predict development of waves within the estuary. At least fifteen years of wind data to drive the models are available from various coastal anemometer stations and from the archive of the UK Met Office Wave Forecasting Model.

Sea level
Unless there are sufficient measured sea level data available in the estuary, from which extrapolation to extreme values may be possible, alternative sources have to be considered. POL (1997) derived values of extreme sea levels around the coast of the UK using data from a numerical tidal model as well as measured data from 'A' class tide gauges. The results are detailed at the model grid points, which may be some distance from the coast since the grid size is 36km by 36km. In order to transform accurately these extremes to a location within the estuary, a fine mesh tidal flow model of the area may be required. If time series sea level data are available at a site nearby, as above, then these data could be transferred to a site of interest through the use of the tidal flow model. Two further complications in estuaries of extrapolating measured data are that extreme water levels may be limited by the river bank elevation, and that hydraulic interactions between astronomical tide, surge and flow may be different in extreme conditions.

River flow
Flow rates are routinely monitored in many rivers and in some cases many years of time series flow data are available at nominal cost from the Environment Agency. Institute of Hydrology (1998) gives details of the locations of gauging stations for the whole of the UK, together with summary statistics such as mean annual flow and 10-percentile flow. In some cases there may be no flow data available or there may be considerable inflow downstream of the gauging station. It may therefore be necessary to undertake a hydrological study to develop design hydrographs from runoff data. Methodologies for such studies are described in Institute of Hydrology (1999). However, as with measured sea levels, it may be necessary to 'move' the data to the point of application and to convert them to equivalent induced increases in water level. This will usually require a site-specific numerical river model (e.g. iSIS) and may be done before or after statistical modelling, and possibly in conjunction with sea level transformation, depending on the analysis approach adopted.

3.2.2 Secondary variables

Wave period
Where wave period is considered important it is measured or modelled at the same time as wave height. It may be allowed to vary freely, or it may be assumed to be either partially or fully dependent upon wave height, either assuming a constant wave steepness (CIRIA, 1996, Section 3.3.2) or, for example, taking a typical steepness with some variation, independent of wave height, about that value.

Wind set-up
The numerical model used to determine the extreme sea levels given in POL (1997) uses meteorological parameters to predict the extent of increase in surface elevation above the level one would expect under average meteorological conditions. This is known as surge. The surge component calculated will have wind set-up incorporated within. The same argument also applies to measured sea or water level data. Therefore if the location of the sea or water level data is relatively close to the site of interest, then

wind set-up need not be considered explicitly. However, if the measured sea level data are some distance from the estuary, the extent of wind set-up could be different at the point of interest. In this case wind set-up can be calculated either during the tidal flow modelling by incorporating a wind stress parameter, or from a simple formula involving wind speed, water depth and fetch length (CIRIA, 1996, Section 3.1.4).

Wave set-up
This is a much more localised phenomenon, which may already be accounted for in any design method to be used. However, it may occasionally be necessary to make explicit allowance for wave set-up, for example using a simple formula involving wave height, wave period and bed slope (CIRIA, 1996, Section 3.1.5).

3.2.3 Proxy variables

Surge
If the analysis is not already over-burdened with input variables, surge (the non-astronomical components) of sea level could be considered in addition to either astronomical tide or sea level. In some cases, surge propagates differently to astronomical tide, and in some instances the effect on the overall estuary water level may be different. A preliminary assessment of an individual site of interest will usually show whether or not this approach would be useful.

Wind speed
Wind speed, usually in conjunction with duration and direction, is the driver for waves and for wind and wave set-up. To keep the number of variables to a minimum, it may be convenient to undertake a joint probability analysis using wind speed as a primary variable (perhaps conditional upon the wind being in a certain direction sector). Wind speeds can then converted to wave conditions (and to overtopping rates or overall water levels) as appropriate during the subsequent structure variable calculations.

River water level
Prediction of river water level is, of course, the aim of this project, but the variable is listed here just as a reminder that it may be possibly to work directly in terms of water level measurements close to a site of interest.

3.3 One-variable extremes

In the context of the present study, one-variable extremes prediction methods are relevant in three situations. The first is where measurements of overall water level have been recorded close to the site of interest, and where these measurements can be extrapolated directly. The second is where hydraulic modelling is used prior to statistical analysis to derive the equivalent of direct measurements based on primary variables recorded or simulated nearby. The third is where additional non-simultaneous data are available on one or more of the input variables, which can be extrapolated and subsequently used to refine the joint probability statistical analysis (based only on the simultaneous portion of the data).

The relative merits of the Generalised Pareto Distribution (GPD) and the three-parameter Weibull distribution were discussed and illustrated during an earlier joint probability study described in HR Wallingford (2000a). Extremes analysis is a three-stage procedure, firstly identifying a threshold level above which all data are treated as extreme values, secondly the fitting of a probability distribution to the extreme value data, and then extrapolation of that distribution to extreme values. One point to be careful with is that the number of records per year (events above a threshold value if used) needs to be known in order to determine the probability associated with any particular return period. Both distributions have been used extensively in research and consultancy studies, and no new validation work was undertaken during the present study.

3.4 Two-variable extremes

Joint probability typically refers to two or more partially related environmental variables occurring simultaneously to produce a response of interest (e.g. flooding). Examples are large wave heights and high sea levels, large river flows and high sea levels, and large surges and high astronomical tides. In the present context, two-variable statistical analysis is required where two of the three primary variables (sea level, river flow and waves) are to be extrapolated to a joint probability density and extremes. Joint probability was the subject of earlier MAFF-funded research at HR Wallingford and Lancaster University, where the focus was on large waves and high sea levels. That work was reported in detail in HR (2000a and 2000b) and in outline in Owen *et al* (1997) and Hawkes *et al* (2002) but will be summarised here, as the methods are still appropriate.

Joint probability combinations of wave heights and sea levels with a given extreme chance of occurrence are often defined in terms of sea conditions in which a given wave height is exceeded at the same time as a given sea level being exceeded. The new method (HR Wallingford, 2000a), conversely, works in terms of the joint probability *density* of the two variables and, in addition to producing joint exceedence extremes if required, can predict the return periods of the structure function(s) (where known, e.g. overtopping on a particular seawall) more directly.

Wave heights and surges are usually partially correlated (since both are related to the local weather conditions). The extent of the correlation depends on a number of factors, and is best estimated using simultaneous data on the two variables. However, the most extreme conditions (usually including a high surge component) will tend to be more correlated, particularly where the astronomical tidal range is low. Similar arguments about the most extreme values being more correlated might be expected to apply, but probably to a lesser extent, to river flows with either wave heights or surges, as all are related to weather conditions.

The joint probability analysis method (described here in application to waves and sea levels, but in principle also applicable to other variable-pairs) has five main elements.

1. Preparation of the input data, consisting of many independent records of wave height, wave period and sea level. A convenient way of satisfying the requirement for the records to be both temporally independent, and relevant, is to use only those records representing conditions at the peak of each tidal cycle (i.e. one record every 12 or 13 hours). (This may be an over-simplification in the present project where river flows may need to be gathered over a period of several days.) At least three

years of data, representative of the type of sea states of interest, are needed to justify the effort involved in applying the new approach.

2. Fitting of statistical distributions to the wave heights, the sea levels and the wave steepnesses. GPDs are fitted to the top few percent of each of the marginal variables, i.e. wave heights and sea levels, whilst the distribution of wave steepnesses is modelled by a Normal regression on wave height.

3. Fitting the dependence between wave heights and sea levels. Two alternative partial dependence statistical models were developed to represent the dependence between wave heights and sea levels. These consist of a single Bi-Variate Normal (BVN) Distribution and a mixture of two BVNs. These models were chosen since the dependence and extremes characteristics of the BVN were already well understood.

4. Simulation of a large sample (typically thousands of years) of sea conditions consisting of wave height, wave period and sea level data, using the fitted distributions, and therefore with the same statistical characteristics as the input data. Joint probability analysis is based on simultaneous information on the variables of interest. It is quite likely that there will be additional non-simultaneous data on at least one of the variables, with which to refine the extremes predictions for that one variable. The present method incorporates any refinements by re-scaling during the long-term simulation of data, thus permanently building this information into the synthesised sea state data to be used in subsequent structural analysis. In the present project these refined predictions might come from published extreme sea levels (POL, 1997), established extreme hydrographs or wave predictions based on long-term wind data.

5. Extremes analysis based on the simulated data. Results can take the form of extreme wave heights (and associated periods), extreme sea levels, or extreme combinations of the two. In addition, any given (known) structure function which can be defined in terms of analysed variables can be synthesised directly for every record in the simulated data sample. Direct analysis of the distribution and extremes of the structure variable is then relatively easy: extreme values can be estimated from the appropriate sample exceedence probability in the synthesised data.

During the present project, some developments were made within element 5, to cope with structure variables too complex to be represented by equations, but rather requiring hydraulic models for calculation of overall water level, economic loss etc. For up to a few dozen joint exceedence extremes of interest, it may be practical to run a hydraulic model simulation for each case, but this would not be so for the potentially hundreds of thousands of records in a long-term simulation. Instead, a structure variable contouring method was developed and tested. A representative sample of combinations of high and extreme values of the input variables is selected for use in the hydraulic model. The corresponding structure variable values from the model are contoured using fitted equations, retained to represent the structure function needed to process the long-term simulation into an equivalent structure variable distribution.

3.5 Three-variable extremes

A full three-variable statistical analysis of sea levels, flows and waves will be needed only quite rarely. It would require that all three primary variables are of sufficient importance, that simultaneous data existed for all three, and that no alternative measurements were available which might already represent the combined effect of two or even all three of the variables. Nevertheless (and partly because this was originally the main thrust of the project) a rigorous three-variable approach was developed, along the lines of the two-variable (three if the secondary variable of wave period is included) JOIN-SEA joint probability methods described in Section 3.4.

After transformation of the marginal primary variables to Normal distributions, all variable-pairs can be modelled by a Bi-Variate Normal model above a threshold. For the sea level and wave height variables this is the model adopted previously. If the flow variable is also included then the other two pairs can also be modelled above a threshold as BVN. Combining these pairwise dependence models gives the Tri-Variate Normal (TVN) model. The input data are prepared in a similar way to the previous study, each record now consisting of simultaneous values of each of the *three* variables (four if wave period is retained and analysed as a secondary variable as in BVN). At present the distribution fitting stage is achieved by running the *BVN* analysis program three times, corresponding to the three possible pairings of variables. This is rather inefficient both in terms of staff and computer time, as it means that some of the analysis work is repeated across separate runs. However, it results in GPD fits for the tails of all three variables, and correlation coefficients for each of the three pairings (plus the wave steepness statistics if wave period is included).

A new TVN simulation program (*SIMTVN*) was developed based on the earlier *SIMBVN* program. A long-term simulation is undertaken using the TVN with *three* representative correlation coefficients (applied only above a single chosen threshold), the transformation back to the original variables, and the fitted GPD's for the upper tails. Again, any refinement of the marginal extremes which may be available from other data sources can be incorporated during the simulation. The long-term simulation then has the correct marginal distribution characteristics and the correct degrees of correlation, but has been extrapolated to higher values than occurred in the original sample. Extreme joint probabilities and structure functions (e.g. overall water level) can then be estimated from the simulated data. Where it is too complex to be specified in the form of equations, the structure function may involve hydraulic modelling.

Coding of the TVN approach has been thoroughly tested to check that it reproduces results from the BVN programs. It has also been demonstrated on three-variable field data from the estuary at Truro, producing a plausible three-variable joint distribution of sea level, flow and winds (as proxy for waves). To demonstrate the final stage in the joint probability analysis procedure, a simplified structure function was applied to the simulated data, and three-variable joint exceedence extremes were derived. At present the TVN analysis procedure is rather cumbersome to use, and it would be possible to develop more efficient computer programs it if were to come into more frequent use.

The three variables used as input to the TVN method need not be of the same type (e.g. 'water level') or even have the same dimension(s) (e.g. 'length') but there is a constraint on the three separate correlation coefficients in that they must be reasonably consistent with each other. For example, two pairings each with a high degree of correlation would be inconsistent with there being no correlation between the third pair. This constraint can be expressed in the form of an equation which has been included as a check within the *SIMTVN* program code, and is typically satisfied when sample-based estimates of the correlation parameters are used.

3.6 Incorporation of secondary variables

If wave height is treated as a primary variable in the statistical analysis (whether of one, two or three variables) then wave period, wind set-up or wave set-up can be incorporated as an additional secondary variable. Realistic wave periods can be estimated from a constant wave steepness value applied to each wave height of interest. (This assumes, rather arbitrarily, that wave height and wave period are completely correlated.) However, a better approach was developed for JOIN-SEA (see Section 3.4) in which wave steepness is allowed to vary during the long-term simulation. Below a selected threshold, its value is taken from the sample distribution, independent of wave height; above the threshold, it varies as a function of wave height, with some variability about its mean value for that wave height.

The present approach to wind set-up and wave set-up is comparable with the ‘constant wave steepness’ approach mentioned in the last paragraph. Wind set-up depends on various site-specific fixed values, and is proportional to wind speed (if the wind is from a relevant direction) which in turn is roughly proportional to wave height (again if the wind is from a relevant direction). So, following some site-specific preliminary calculations, and assuming that wind set-up is not already present in the sea level data, it can be estimated as a (usually small) fixed proportion of wave height. Wave set-up (where required to be explicitly estimated) is handled in a similar way, although here the relation between set-up and wave height is more direct (and again depends on other site-specific fixed values).

4. METHODS FOR OVERALL EXTREME WATER LEVELS

4.1 Preliminary assessment: common to all methods

Before undertaking any detailed analysis, it is necessary to undertake a preliminary physical appreciation of the site and the project requirements, and to search for available data and existing hydraulic models. This exploratory analysis of the problem, and in particular of the structure variable (overall water level, potential economic loss etc), plays a key role in deciding what form of methods should be used. Section 2.2.1 lists the questions to be considered at this stage.

Some simple numerical modelling and dependence analysis may be needed at this stage, to assist in determining the relative impact of different primary variables on overall extreme water levels. This assessment will help in deciding which, if any, statistical/hydraulic modelling approach is to be preferred. The examples already assessed in this way show that this preliminary stage is a cost-effective way of selecting the appropriate analysis method, and that a full three-variable approach will not often be needed.

A numerical river model is driven by upstream flow and downstream sea level. It may also include secondary flows, control structures, out-of-bank flows and storage ponds. Some preliminary runs, even with unrealistically high input variables and/or a simplified river structure, may help to determine which input variables are most important, and the likelihood of out-of-bank flow and the impacts of control structures.

If the numerical model shows that overall water level at the study site does not respond to changes in downstream sea level, that would suggest that sea level is unimportant. Conversely, if it does not respond to changes in upstream flow, that would suggest that flow is unimportant.

If overall water level at the study site responds differently to the same peak downstream sea level composed firstly of astronomical tide and secondly of astronomical tide plus surge, that might suggest that tide and surge should be considered separately. It if responds to differences in hydrograph duration, that would suggest that duration will be important in determining the event definition.

It is useful to undertake auto-correlation (single variable) analysis, and cross-correlation analysis for any relevant variable-pairs, if time series data are available at this preliminary stage. Auto-correlation analysis provides an indication of cycles or persistence within individual variables, for example on a tidal, hydrograph, spring tidal or seasonal scale. This can be helpful in setting the event definition, to meet the potentially conflicting requirements of independence between records and the need to include all significantly high values of all input variables.

Cross-correlation analysis provides an indication of dependences between variable-pairs, possibly with a time lag between them. This might indicate, for example, that surge is significantly correlated with the other variable(s) but that downstream sea level is not, and therefore that surge should be treated as a primary variable. It will indicate whether there is any trace of tidal level in upstream flow, and whether there is any trace of flow in downstream water level, again helping to determine their relative importance in assessment of overall water level. Also, if one variable is shown to be independent of the others, it may be possible to perform a full joint probability analysis of the other variables, and to re-incorporate the independent variable in later calculations.

Another preliminary step involves gathering, checking and preparation of the data to be analysed. Time series and/or distributions of variables and/or established extreme values may need to be gathered for preliminary analysis, full analysis or validation purposes. An important decision at this stage is to determine what constitutes an 'event', in terms of which input variables are needed and the interval between records, so that data records can be compiled for the data set. In the case of waves and sea levels at the coast (HR Wallingford, 2000a) the wave height and sea level at each successive high tide are usually used.

In the case of flows and sea levels, the situation is not so clear. There may be a time lag as the flow and sea level data will not be measured at the same position in the river, and the effect of flows may be accumulated over several hours or even days. For a site dominated by sea level, it may be good enough to regard each high water as an independent record, and for a site dominated by fluvial flow to treat each spell of high flow (of whatever duration) as a separate event. However, in the more general case where both flow and sea level are important, a peaks over threshold approach may be the best way to identify events. This might be done using one threshold for high sea level (with an accompanying, possibly time lagged flow rate) and a second one for high flow rate (with an accompanying peak sea level during the flow event). The most appropriate event definition will vary from one site to another. There is not necessarily a 'correct' event definition for a particular study, and plausible alternatives should all produce similar end results.

4.2 Range and choice of joint probability analysis methods

4.2.1 Classification of methods by number and type of primary variables involved

The statistical analysis approach adopted is highly dependent on what data are available, on which are the most important variables at the location of interest within an estuary, and on the intended use of the analysis results and the decisions to be made. Hydraulic modelling may be an integral part of the procedure, to reduce the number of variables prior to analysis, or may be needed to make practical use of two- or three-variable statistical outputs.

The classification system adopted here is based on the number and type of primary variables (sea level, river flow, waves and/or estuarine water level) used in the statistical analysis. Table 1 summarises the defining characteristics of six alternative classes of analysis method.

Table 1 Classification of alternative statistical approaches

Method No	No of primary variables	Type of input variables	Product of statistical method
1	1	Measured overall water level data near to the site (e.g. nearby tide gauge)	Direct 1-variable extrapolation of measured overall water level
2	2 (reduced to 1)	Flow and sea level data transformed by hydraulic modelling to equivalent overall water level data at the site	Direct 1-variable extrapolation of synthesised overall water level
3	2	Measured flow and sea level data, used directly in two-variable statistical analysis	2-variable simulation and extreme combinations of flow and sea level; use as input to hydraulic modelling
4	2	Synthetic wave data, and measured or transformed sea or river level data, used directly in two-variable statistical analysis	2-variable simulation and extreme combinations of waves and sea or river levels; may be directly useful or may require the use of structure functions
5	3 (reduced to 2)	Synthetic wave data, plus flow and sea level data transformed by hydraulic modelling to equivalent water level (due to sea level and flow) data at the site	2-variable simulation and extreme combinations of waves and water levels; may be directly useful or may require the use of structure functions
6	3	Synthetic wave data, and measured flow and sea level data, used directly in three-variable statistical analysis	3-variable simulation and extreme combinations of waves, flows and sea levels; use as input to hydraulic modelling
Note: Secondary variables such as wind set-up, wave set-up and wave period can be incorporated as a function of wave height, but for inclusion of wind set-up in Methods 2 and 3, additional calculations may be needed.			

Methods 1 and 2 are single-variable approaches, working directly in terms of the structure variable of interest. Method 1 could be used only if long-term measurements of water level were available close to the location of interest, when an appropriate single-variable extrapolation method could be applied directly to the measured data. Method 2 relies on similar long-term water level data being able to be generated from corresponding long-term flow and sea level measurements, probably by continuous simulation using a numerical river model, followed again by single-variable extrapolation of the river water level data.

Methods 3 and 4 are two-variable joint probability approaches, working in terms of the joint distribution of flow and sea level, or of waves and sea level, respectively. (Waves and flow could be treated similarly, but this is an unlikely combination). A two-variable approach would be applicable if it had been shown in preliminary work that the two variables concerned were important and that the third variable (waves in Method 3 and flow in Method 4) was not important. It also requires that simultaneous sequential data are available for the two variables (in the case of waves, this would probably mean the wind data needed for hindcasting, rather than measured wave data being available). In Method 3, the resulting combinations of flow and sea level (now with estimated probabilities of occurrence) would probably be used as input to numerical river modelling. Similarly for the results from Method 4, except that structure functions such as overtopping or armour size would be used instead of a river model.

Method 5 is applicable where sea level, flow and waves are all important, but where the three variables can be reduced to two, using continuous simulation numerical river modelling to reduce the separate flow and sea level data to equivalent river water level predictions for a particular location. The resulting two-variable analysis is then applied to water level (due to sea level and flow) and waves. The resulting combinations of water levels (due to sea level and flow) and waves would probably be used as input to structure functions such as overtopping or run-up.

Method 6 is a full three-variable joint probability approach for use where sea level, flow and waves are all important enough to be considered as partially dependent primary variables. How to use the resulting three-variable combinations of flow, sea level and waves may not be obvious, and should be considered before embarking on the analysis. It would probably involve a mix of hydraulic modelling and structure functions, applied to a large number of combinations of conditions, perhaps in order to determine an overall probability of flood risk or an overall flood management decision.

4.2.2 Selection of an appropriate analysis method

Figure 1 highlights some of the factors to consider in selecting an appropriate class of joint probability analysis method for use in estuaries and tidal rivers. It is not intended to be prescriptive, but rather to illustrate the importance and relevance of some of the main decisions involved in making an appropriate choice from amongst Methods 1-6. Figure 1 includes four questions (in diamond-shaped boxes) and allows for six alternative answers (in rectangular boxes with rounded corners) to the question ‘What data are available’ leading to one of the six alternative methods (in bold in Figure 1).

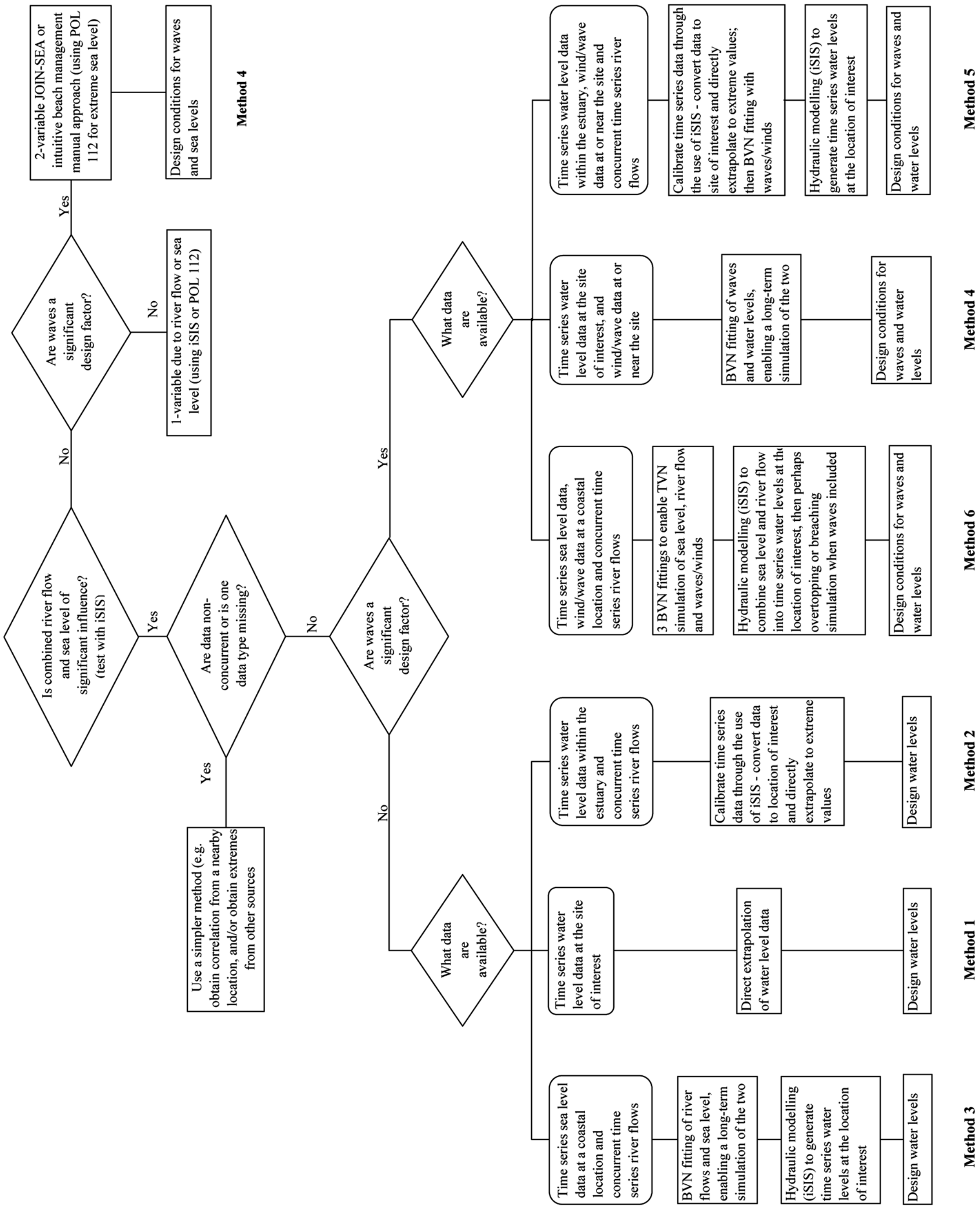


Figure 1 Selection of analysis method from alternative statistical approaches

4.3 Method 1: Measured overall water level data close to the site

In many estuaries and rivers, several years of water level measurements have been recorded by the Environment Agency and others. If available at a suitable location, these measurements will already include the combined effects of all components of the overall still water level (i.e. excluding waves). Rather than attempting to extract the separate components, it is better to work directly with the measured data, using a one-variable extremes method. An obvious potential difficulty with this approach is when the river changes from in-bank to out-of-bank flow, where there will be a discontinuity in the distribution of overall water level. This might be accommodated in analysis by using only out-of-bank level data in the analysis, or by manually reducing the predicted extreme water levels where they would exceed the bank crest elevation.

It may also be necessary to make any small changes to the results for expected differences in wave set-up or flow-induced water level between the measurement and application locations. Another potential complication is that hydraulic interactions between astronomical tide, surge and flow may be different in extreme conditions to those which occurred in the data sample.

4.4 Method 2: Continuous simulation of water level

Where concurrent time series upstream flow data and downstream sea level data exist it is feasible to use these data directly as input to a hydraulic model (e.g. iSIS). The hydraulic model would then take the place of equations which might otherwise have been used to construct a simpler structure variable such as run-up or overtopping rate. Corresponding time series data (flow or water level) can then be derived at the location of interest in an approach often referred to as continuous simulation. These data can then be statistically analysed directly to provide estimates of extreme values (see Section 4.3 above).

The benefits of this approach are:

- It is not necessary to quantify the nature of the dependence between the input variables and is therefore relatively straightforward to apply;

- Measured data are input to the model; errors in the resulting calculation of extreme water levels are only (but this is a big condition) a function of model limitations and extrapolation methods;

- A product of the continuous simulation approach is time series data at every cross-section within the model. Extreme water levels can therefore be calculated at each cross-section with relative ease, without the need to construct separate structure functions for each location.

The drawbacks of this type of approach are:

- Model construction and runs are time consuming;

- The accuracy of the results is limited by the length of available concurrent time series data;

- Inherent in this approach is the assumption that the river response is the same for extreme events as it is for more commonly occurring events. For example, the respective influence of the input variables may vary considerably when the river flow is ‘out-of-bank’. The time series data output from the model run may include only a few ‘out-of-bank’ events. Therefore it is reasonable to question the validity of extrapolating ‘routine’ data to high return period events.

4.5 Method 3: River flow and sea level data

4.5.1 Introduction

The method described here is an alternative to that described in Method 2 above. This approach differs in that the loading conditions, in this case upstream river flows and downstream sea levels, and any dependence, are extrapolated (joint density extrapolation) and the extremal behaviour of combinations at intermediate sites subsequently derived.

4.5.2 Data processing

The concurrent time series data are split into separate time blocks (records). Each record is summarised by extracting a single value for each variable. These summary statistics are ultimately used to derive values of the main variable of interest. It is therefore essential that the length of the time block and the manner in which the variables are summarised be given considerable thought before carrying out the statistical modelling. Previous joint probability studies (HR Wallingford, 2000a) regarding large wave heights and high sea levels have considered each high tide as an individual record. This is a convenient method of time blocking that concentrates on the time of interest and is appropriate where the response (e.g. overtopping) will persist for a short time period (2-3 hours). However, extreme fluvial events can persist for several hours to over ten days. The decision regarding an appropriate time scale in which to split the time series data will therefore vary considerably from site to site. Additionally, the relative importance of river flow and sea level will vary with location within the estuary. It is therefore important that due consideration be given to both variables when processing the data into time blocks.

The authors tested a number of possible data selection schemes, involving peaks over threshold for one or both variables, or division of the data sequence into equal time periods. None of the peaks over threshold schemes were really satisfactory as they failed to achieve a representative distribution for both variables. The approach of dividing the data series into equal time periods was satisfactory in this respect, but required a careful balance between too small a time period potentially retaining dependent records and too large a time step potentially missing significant independent occurrences. The best approach, assuming river flow and sea level to be of roughly equal importance, seemed to be to choose a time period representative of the interval between potential river flow ‘events’, conveniently represented by the length of the hydrograph for a particular location. In practice, this meant dividing the data sequence into time periods of between about twelve hours and about two weeks, treating each period as an independent record, and selecting the peak flow and peak sea level to represent that period of time.

4.5.3 Data simulation

Having processed the data, the statistical models are fitted (GPDs and BVN). Any dependence between the variables is quantified and the extremes for the marginal distributions are estimated. At this stage it is possible to improve the estimates of the marginal extremes if additional, non-concurrent data are available. The fitted statistical models are then used to generate a large set of different records (pairs of river flows and sea levels).

Having simulated the long-term distribution of the input primary variables, the next step is to convert the simulated records to the structure variable of interest. In previous applications of these joint probability methods to wave heights and sea levels (HR Wallingford, 2000a) this has been a relatively straightforward process. Equations exist to relate the wave height, wave period and sea level to structure variables (e.g. overtopping, run-up). In such situations it is a simple process to calculate a structure variable value for each simulated data record. However, the estimation of estuarine water levels from pairs of values of upstream river flow and downstream sea level is not straightforward. For example, flow duration, differences between in-bank and out-of-bank flow, and any time lag between sea level, flow and induced water level may be difficult to represent by formulae and may require the use of a hydraulic model.

4.5.4 Structure function evaluation

Essentially there are two methods of estimating estuarine water levels from river flow and sea level. The first makes use of a simplified empirical formula, and the second is through the use of a hydraulic numerical model such as the HR Wallingford / Halcrow model (iSIS). Jones (1998) investigated both approaches and recommends the use of a hydraulic model in practice. If it were necessary to study the complete distribution, but impractical to run all of the data through a hydraulic model, it would be possible to use formulae for commonly occurring conditions and the hydraulic model for high and extreme conditions. However, this is unlikely to help in the estimation of extremes, and so this option is not considered further in this study.

As discussed above, individual events must be defined in terms of single values of the loading conditions. The statistical modelling thus removes all time variance of the loading conditions (both within an individual event, and between consecutive events). However, to represent as accurately as possible the physical processes of the sea level/river flow interaction, through the use of a hydraulic numerical model, it is important to consider the time varying nature of flood events. Time variance of the downstream boundary can be well estimated from a peak sea level, through the use of tide curves. However, the inclusion of a time varying upstream boundary is complex. Extreme fluvial events are generally represented by hydrographs for different duration events. The shape of the hydrograph is a function of the size and shape of the river and catchment. In essence they are an attempt to categorise the river response to different storms. Although it would be possible to study the joint distribution of flow maximum and duration, it will usually be necessary to make an assumption regarding the most appropriate duration (hydrograph shape) with which to represent the upstream boundary. This decision also gives guidance on the most suitable length of time block when processing the data prior to the statistical modelling.

For some rivers standard hydrographs are well defined for different duration events. In this instance a decision as to the most appropriate duration is required. For rivers where there are no existing hydrographs, analysis of the time series flow data will provide some insight, from which a standard 'design' hydrograph can be estimated.

4.5.5 Structure function contouring

During this study, a new contouring method was developed for the common situation where an analytical approach is desirable, but where the structure function (in the form of a river model or economic loss model) is too complex to permit more than a few dozen cases to be modelled in detail. Structure variable values are determined over a grid defined in terms of about six values of each of the two primary variables (upstream river flow and downstream sea level) covering the range of marginal return periods of interest. Structure variable contours are fitted across the grid and equations representing those contours are stored. The stored contours are then applied to each record within the range of marginal return periods of interest in a JOIN-SEA long-term simulation (or any other large sample of data). This yields an empirical distribution for the structure variable, from which extreme values can be determined, for example 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).

The contouring approach has been applied successfully to derivation of extreme estuarine water level (driven by river flow and sea level) in the Thames and Clyde and to economic loss (driven by waves and sea level) on the Arun to Pagham south-facing coast of England.

4.6 Method 4: Wave and sea level data (flow unimportant)

In many estuarine situations it will not be necessary to consider explicitly the influence of river flows. For example, in areas where the river flow input is small in comparison with the tidal outflow, the effect of large river flows on the overall water level may be negligible. In such circumstances the problem of calculating extreme water levels (excluding waves) diminishes to a single variable extrapolation. Nearby measured sea level data can be transformed to the site of interest, through simple conversion factors, or, where the measuring station is a greater distance from the site of interest, a tidal flow model can be employed.

Having derived the time series sea level data the structural response of interest can be derived through extrapolation of the joint density of the wave and sea level data as outlined in Section 3.4 and described in more detail in HR Wallingford (2000a).

4.7 Method 5: Wave and river water level data using a 'structure function' approach

This approach is the same as that described in Section 4.4 but is applied in locations where wave action is significant. Time series river flows and sea levels are input and reduced to one variable (water level) at the location of interest through the use of a hydraulic model (e.g. iSIS).

Having derived the time series water levels, the structural response of interest can be derived through extrapolation of the joint density of the wave and water level data as outlined in Section 3.4.

4.8 Method 6: Wave and measured flow and sea level data

This is an alternative to Method 5 and an equivalent approach to Method 3, with the additional consideration of waves. The three variables (river flow, sea level at the downstream tidal boundary and waves) are modelled using the BVN distribution for each of the three different pairings. The dependence between the three pairings is assessed and used in a TVN simulation of many different combinations of the three variables, as outlined in Section 3.5.

The simulated data then have to be converted from the input primary variables to the structure variable of interest. This can be achieved using either the structure variable simulation approach or joint exceedence approach discussed in Section 4.9 below. If the former is chosen, then a modification of the structure variable contouring method described in Section 4.5.5 can be used to determine efficiently the high and extreme distribution of the structure variable. Sufficient combinations of high river flow and high sea level could be used as input to a hydraulic model to contour the corresponding distribution of still water level (i.e. without waves) at a location of interest within the river. Then for each three-variable record, the river flow and sea level would be converted to still water level using the fitted contours, and an overall structure variable such as overtopping rate would be estimated using wave condition and still water level as input to a structure function expressed as formulae. Finally, extreme values of the structure variable would be predicted from the derived sample distribution.

4.9 Extremes analysis based on extrapolated joint probability density data

The discussion in this section focuses on how to use the simulated long-term distribution of river flow and sea level data in conjunction with a hydraulic model, but in principle the arguments also apply to the other two- and three-variable methods. There are two main approaches to the derivation of extremes from the simulated long-term distribution of the input variables, i.e. the joint exceedence method and the structural variable method.

4.9.1 Structure variable simulation

The structural response simulation method involves estimating the overall water level at the location of interest for all the records in the long-term distribution of the input river flow and sea level data. It is too time consuming and therefore not practical to run the hydraulic model for each data pair. This problem can be overcome by the generation of 'look-up' tables. The hydraulic model is run for a range of different river flows (hydrographs) and sea levels (tide curves). The results are then tabulated and can thus be considered as the 'structure function'. The structure variable is then evaluated with reference to the tabulated results, for each data pair in the simulated data file. Extreme values are then predicted from the accumulated distribution of the structure variable.

4.9.2 Joint exceedence

The joint exceedence approach, as the name suggests, involves working in terms of the joint exceedence of the input primary variables (for a full description see HR Wallingford, 2000a). Different combinations of river flow (hydrograph) and sea level (tide curves) with the same joint exceedence return period are derived from the simulated data (through the countback procedure) and run through the hydraulic model. The highest calculated value of the response is then assumed to take the same return period as the joint exceedence combinations (NB In practice an allowance is made for the fact that joint exceedence extremes under predict the value of the response (HR, 2000a)).

4.9.3 The choice between structure variable simulation and joint exceedence

One of the benefits of using the structure variable simulation method as opposed to the joint exceedence approach is that the estimated extremes are derived directly in terms of the return period of the response. The drawbacks of this type of approach are that the accuracy of the results is restricted by the number of hydraulic model runs used to derive the ‘look-up’ tables. Additionally the subsequent assimilation and analysis of the results is time consuming.

The advantages of using the joint exceedence approach are that a comparatively small number of hydraulic model runs are required. This type of approach can also be adapted to sites where no time series data exist but extremes for the input variables are available. The drawback of this method is that the extremes for the variable of interest are initially calculated in terms of the return period of the joint exceedence of the loading conditions. This is not the same as the return period for the response. An adjustment of the joint exceedence return period is therefore required to account for this difference (see HR Wallingford, 2000a). It is not practical to calculate the magnitude of this adjustment analytically as it depends on the dependence characteristics of the data, and an estimate is thus required. Experience with two-variable analyses suggests that the necessary adjustment factor will typically be about two, but it can be significantly higher in three-variable analyses or for two near-independent variables.

4.9.4 The advantages of a joint probability approach

The advantages of two-variable (or three-variable) extrapolation (Methods 3-6) as compared to single variable extrapolation (Methods 1 and 2) are:

- Additional data that are not concurrent can be used in the calculation of marginal extreme values and thereby become included within the data simulation;
- An assessment of the correlation between flows and sea levels is carried out which gives further insight into the physical processes causing the overall extreme water levels;
- Hydraulic modelling is carried out on the extreme events, so that the extrapolations are physically based and can accommodate features such as overflow ponds, control structures and out-of-bank flows;

- It is possible to separate sea level into its surge and astronomical tide components and statistically model the two variables separately. This is shown to be a more rigorous and accurate approach by POL (1997) and by Jones (1998).

The disadvantages are:

- The assumed hydrograph shape and duration which in practice will usually be necessary with a joint probability approach is a considerable simplification, resulting in the loss of uniqueness of different flood events;

- The statistical modelling and final analysis method are far more time consuming and hence costly than the alternative single variable extrapolation used, for example, in continuous simulation (Method 2).

4.10 Options when data are insufficient

There are several stages in a joint probability analysis, and it is possible to intervene to add or refine information between the stages. It may be possible to infer and use distributions, extremes or dependences from previous studies or nearby sites. It may be possible to use non-simultaneous data to provide some of the information required. It may be worth running a joint probability analysis with one variable less than intended, to gain some insight into the joint behaviour of the other input variables.

If data are insufficient, usually due to shortage of simultaneous data covering all of the input variables of interest, a simplified approach might be considered. CIRIA (1996) Section 3.5.3 provides a method for estimating appropriate joint exceedence combinations from a knowledge of the distribution and extremes of each of the input variables and an assumed ‘correlation factor’ between them. Defra / Environment Agency (2003a) expands upon this simplified approach, making it applicable to a wider range of return periods and providing dependence information on several different variable-pairs of interest in flood and coastal defence. Additional information on marginal extremes might come from previous consultancy studies, offshore design guidelines or POL (1997) for extreme sea levels.

5. EXAMPLE APPLICATIONS OF METHODS

5.1 Site selection

The Severn from Gloucester to Avonmouth was chosen for the main case study, because of the wealth of existing data, models and reports (including previous joint probability studies) and because of the potential importance of astronomical tide, surge, waves and river flow. Several other rivers were chosen where alternative joint probability methods could be demonstrated during consultancy studies. For a regional study of the lower Thames, river flow was found to have little impact on overall water level: downstream sea level and wind speed provided sufficient input, via wave, river and flood area models, to determine the distribution of economic loss due to flooding. The Taff and Cardiff Bay provide an example analysis of a flashy (short duration hydrograph) river. The Clyde provides an example of the 'contouring' method developed for use with complex structure functions. Truro provides an example where both sea level and river flow act on a tidal flood defence barrier. Carlyon Bay provides an example where the joint probability of waves, sea level and river flow was considered. Whitby provides an example where good preliminary work reduced the calculations to a single-variable problem.

The main text of Chapter 5 summarises the key actions, decisions, event definitions, analysis methods, strategies, comparisons between methods, potential difficulties, conclusions etc.

5.2 The Severn Estuary case study

5.2.1 Introduction

The River Severn was chosen as an example study site for a number of reasons:-

- A SalmonF (predecessor to iSIS) model already existed;
- The river had been studied in detail (HR Wallingford, 1981) and the processes contributing to extreme overall water levels were therefore reasonably well understood;
- A joint probability study had previously been carried out within the estuary (HR Wallingford, 1993) and so the dependences between processes had been considered previously;
- The river is of considerable interest from a flood risk point of view, and there is a relatively large number of flow gauging stations within the estuary with which model results can be compared.

An iSIS numerical river model was developed from the earlier SalmonF model. The upstream model boundary is at Haw Bridge, approximately 5km north of Gloucester. This site was chosen as the boundary as it is far enough upstream to be largely unaffected by tides, and because there is a flow gauge sited there. Measured data from Haw Bridge are available in digital format from 1987 onwards. The model fully represents flood plain effects and also represents out-of-bank storage areas as flood plain cells. Further measured water level data exist at Minsterworth (1993 onwards) and Sharpness (1993 onwards). The downstream boundary of the model is at Avonmouth where the effects of high river flows on the overall water level are negligible. Sea level and surge data from Avonmouth were obtained from 1987 onwards. The focus of the study was on tide-surge-flow interaction.

5.2.2 Preliminary analysis

Initial runs of the iSIS model indicated that the effects of tide are felt no further upstream than Haw Bridge, about 5km north of Gloucester, and that the effects of river flow on water level are felt no further downstream than Avonmouth. The iSIS model confirmed the expectation that both tide and flow influence water level at the three intermediate gauging stations, and that the relative influence of river flow increased moving upstream from Sharpness to Epney to Minsterworth. These preliminary runs suggested that Minsterworth would provide the most interesting comparison between measurements and different prediction methods.

One of the preliminary investigations focused on the potential influence of surge, and whether it needed to be considered separately to sea level. Measured water level at Minsterworth appeared not to be significantly affected by the magnitude of the surge component within a given high sea level, suggesting that surge and tide did not need to be modelled separately in order to predict upstream water level.

Auto-correlation analysis was used to assist in choosing an appropriate time interval to achieve independence between successive sea level, river flow or water level records. Several different auto-correlations were detected, for example the daily cycle of the tide, the fourteen day spring/neap tide cycle, the persistence of high river flow over periods of up to about ten days and the annual cycle of higher winter rainfall. One record per spring/neap tidal cycle was judged the best balance between the need for independence between records and retention of all high values in the data sets.

Related to this was the issue of how to represent each record, and later to reconstruct the corresponding event, using a minimum number of parameters. The peak tidal level and the peak river flow within the fourteen day period proved to be an adequate representation of differences between records. Single representative but scaleable Avonmouth sea level and Haw Bridge river flow profiles over a fourteen day period were sufficient to reconstruct realistic events from the two parameters.

More details of the preliminary analysis (and the subsequent joint probability and extremes analysis) are given in Appendix 1. Please note that the results of similar preliminary analyses in other estuaries and tidal rivers might yield very different results, particularly in regard to appropriate event definition.

5.2.3 Example application of Methods 1-3 at Minsterworth

A longer description of the data sets, analyses and results is given in Appendix 1, but some comparisons of results obtained using different statistical methods at Minsterworth are summarised here.

Method 1: Measured overall water level data close to the site

Six years of water level measurements at Minsterworth (1993-1998) were available for direct prediction of extreme water levels at Minsterworth. Using the same interval between records as in Appendix 1, the maximum water level during each spring tidal cycle was noted. The highest 24 values, i.e. an average of four per year, were used for extremes analysis. Weibull and Gumbel fits gave almost identical results, with predicted 5, 20 and 100 year return period values of 9.87, 10.10 and 10.36mODN.

Method 2: Continuous simulation of water level

Appendix 1 describes how twelve years of data from the Avonmouth tide gauge and the Haw Bridge flow gauge (1987-1998) were used as input to continuous simulation modelling of the water level at Minsterworth. Extreme values predicted from the continuous simulation predictions are 10.10, 10.16 and 10.21mODN, for return periods of 5, 20 and 100 years.

Method 3: River flow and sea level data

Appendix 1 describes use of the same sea level and flow data as in Method 2, but now using fitted distributions and a long-term simulation provided by JOIN-SEA. Extreme values predicted from these data are 10.08, 10.20 and 10.26mODN, for return periods of 5, 20 and 100 years.

Comments on results

Results from the three alternative analysis methods are in good agreement. Results from Method 1 provide an independent check on the results from Methods 2 and 3, as they are derived from different source data covering a different period of time. This suggests that all three methods are satisfactory, if several years of data are available and if appropriate data preparation methods are used.

5.3 The Thames case study

5.3.1 Introduction

The Thames Estuary was chosen partly because it illustrates a number of points of general interest and partly because HR Wallingford has been involved in a series of studies of flows, sea levels and waves in the Thames, meaning that models and data were readily available.

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, the Thames Barrier being an additional complication in hydraulic modelling and flood risk assessment.

5.3.2 Preliminary analysis

A number of different source variables were of potential interest, 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 little impact upon high water level in the part of the estuary being studied. Therefore river flow was not regarded as a primary variable, and a single fairly high flow rate provided adequate representation throughout all subsequent analysis.

Sea level was the most important variable throughout the area studied. It was 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.

Initial wave prediction and wave transformation model runs indicated that waves were 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 meant that there was 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 then be used as input to a number of separate wave prediction models during flood risk evaluation.

Of the six methods of joint probability analysis summarised in Table 1, a variation of Method 4 (usually two-variable analysis of large wave heights and high water levels) was chosen. In this case, however, wind speed was used as a proxy for waves, sea level was represented by its value at Southend, and separate analyses were performed (and then combined) for a small number of separate wind direction sectors.

Records 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.

5.3.3 The joint probability analysis undertaken

Time series wind data to be used in the wave modelling and the 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;

- The different wave prediction points to be used in the estuary were 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 was necessary to know which direction the wind is blowing for any given wind speed /sea level scenario. For example, if the wind is from the north-east, the wave prediction points that are exposed to north-easterly winds will predict a wave condition (and hence a wave overtopping rate), whilst other wave prediction sites, not exposed, will 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 water 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 water levels greater than 3.5m 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 POL (1997) than those obtained from just 11 years of data from one gauge. However, as the JOIN-SEA analysis is 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 POL (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 POL (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.

The long-term simulation stage of JOIN-SEA was run separately for each of the four direction sectors to produce the proportion of 50000 years of data appropriate for each direction sector. These four separate simulations were combined to form an overall 50000 year data set of sea level (at Southend), wind speed (over the Thames Estuary) and wind direction sector.

This single combined data was then applied in a consistent way throughout all of the scenarios considered in the subsequent elements of the project. Mean water level was adjusted where necessary to match assumptions about future sea level rise. The wind velocities were applied as necessary at the different wave prediction points. High tide levels were adjusted as necessary for tidal range differences across the estuary. As water level was not sensitive to river flow in the outer estuary, it was sufficient to use a single representative flow rate where necessary.

5.4 The Cardiff Bay study

The Rivers Taff and Ely flow into Cardiff Bay. The bay is impounded by the Cardiff Barrage, water being released in a controlled way at lower tidal levels. The potential risk of flooding arises from high river flow being unable to leave the bay quickly enough during high tidal conditions. The marginal distributions and extremes of river flow and sea level had already been predicted. This was then a joint probability problem requiring appropriate combinations of high river flow and high sea level, as input to hydraulic modelling and accurate assessment of the risk of flooding in Cardiff Bay (Samuels and Burt, 2002).

There was insufficient length of simultaneous measured river flow and sea level data to predict high return period events in the bay by continuous simulation. There were, however, sufficient data to perform a dependence analysis. 32 years of hourly tide gauge data at Avonmouth were converted to equivalent sea levels in Cardiff Bay using an established regression relationship. Continuous river flow data at Pontypridd (on the Taff) were not available in digital format, but were available in a form that permitted the extraction of the 20 highest flow rates over the same period of time.

Approach 1 demonstrated the advantage of what might be regarded as a preliminary dependence analysis. The first part of the two-stage analysis showed that there was no dependence between the magnitudes of the 20 highest river flows and the corresponding highest sea levels on the same days. The second part showed that the distribution of high sea levels during these 20 events was similar to the long-term distribution of sea levels. Once independence had been demonstrated in this way, it was relatively easy to estimate the probability of occurrence of various combinations of high sea levels and high river flows to be used in subsequent hydraulic modelling. (A low level of dependence, had there been any, could probably have been accommodated through manual adjustments of the probabilities, but a more complex dependence relationship would probably have needed a computationally intensive Monte Carlo simulation.)

Approach 2 involved dependence analysis of daily (9am to 9am) mean river flows (exact timing of peak flow now unknown) at Pontypridd against the highest sea level during the same period (9am to 9am). This showed a slight negative dependence, confirming that independence was a realistic assumption. However, when the analysis was repeated, introducing a nine-hour time lag (sea levels now taken between midnight and midnight) a positive dependence was found. This was not a situation of interest within the flood risk assessment, but it suggested the plausible conclusion that both flow and sea level respond to certain weather conditions, but in this particular case that flow takes around nine hours longer to develop.

5.5 The Clyde study

The intention in this study was to predict extreme water levels at several locations along the Clyde, under the combined action of river flow and sea level, making as few simplifying assumptions as possible. Marginal extremes of upstream river flow and downstream sea level were determined with reference to all available data on each variable.

The flow hydrograph duration was about three days, and so this was chosen to be the interval between successive records, both at the data preparation and at the long-term simulation stages of JOIN-SEA. This was implemented at the data preparation stage of JOIN-SEA by defining one input record at the time of every sixth high water. (The impact of any extreme occurrences being overlooked during this procedure would be corrected by later use of the re-scaling option in JOIN-SEA.) Dependence was determined by analysis of these records. The desired record interval was taken into account at the long-term simulation stage by generating 118 river flow / sea level records per year, i.e. one per six tidal cycles.

Thus far, the approach adopted would be classified by Table 1 as Method 3 (two-variable joint probability of river flow and sea level). However, the special ‘contouring ‘ version was used to cope with a structure function which could not be defined in a simple form. Water level within the river was calculated using the iSIS numerical river model, but it was impractical to run separately the many thousands of records in the long-term simulation. Instead a representative sample of river flow / sea level conditions, covering the range existing within the long-term simulation, was run through the iSIS model. These results provided a regular grid of water levels, as a function of river flow and sea level, for each location of interest.

The special contouring version of JOIN-SEA reads in the grid of water level values, fits a complex contouring function, and then applies that function to each record in the long-term simulation. In this way a long-term simulation of water level was produced for each location of interest, from which the extreme values could be determined by counting back through the highest values in the simulation.

5.6 The Truro study

Analysis at Truro involved extreme estuarine water levels caused by the combined action of high river flow and a high sea level. HR Wallingford was asked to consider the probability of river flooding on the landward side of a proposed tidal barrier which could occur following closure of the barrier in response to a forecast high sea level.

The optimum barrier operating strategy and storage pond layout was determined with the help of a numerical river model. All combinations of different rainfall durations, different rainfall return periods, different high sea level return periods and different relative phases between peak flow and peak tidal level were used as input to further model runs. The overall probability of exceeding a threshold water level landward of the barrier was estimated by adding up the probabilities corresponding to those combinations causing the water levels to exceed that threshold. It was initially assumed that rainfall duration, rainfall return period, high sea level return period and relative phase were independent of each other. Subsequently, field data were analysed to discover a slight dependence between high rainfall and high sea level, and the probability of the threshold level being exceeded was accordingly revised upward.

The study illustrated the importance of planning the necessary hydraulic model runs efficiently and of analysing dependence even where none was initially apparent.

5.7 The Carlyon Bay, Cornwall study

In Carlyon Bay, a small river flows across the sea defence before flowing down the beach. A proposed development had the potential to allow more wave energy to propagate upriver, landward of the sea defence, than had previously been the case. This in turn had the potential to increase flood risk.

A desk study version of joint probability Method 4 was applied to the sea conditions, to assess the probability of large waves occurring simultaneously with high sea levels. Although it was clear that only the highest sea conditions were of interest, and that river flow was relevant, it was not clear whether high, medium or low river flow would provide the greatest flood risk.

A desk study version of Method 6, three-way joint probability assessment of river flow, waves and sea level was undertaken, to determine the highest reasonable river flow for given sea conditions and overall return period. In the physical model tests of the proposed development, river flow was gradually increased up to this highest reasonable value, during which observations were made of the time of greatest possibility of flooding. A full length test was then made for this worst case flow rate.

5.8 The Whitby study

Whitby is one of several studies illustrating the benefit of thorough preliminary analysis. A numerical river model, required for other purposes in this study, was used to test the relative importance of the two primary variables (upstream river flow and downstream sea level) at various locations of interest along the river. This showed that at all of the locations of interest, one or other of the primary variables dominates, and therefore that a determination of their joint probability was unnecessary.

The authors have been surprised to find during preliminary analysis on a number of studies that a joint probability assessment of high river flow and high sea level is quite often not needed. At locations where flow dominates, it may be enough just to assume a nominal high sea level (for example the 1 year return period level) for all cases considered, without introducing significant error. Similarly where sea level dominates, it may be enough just to assume a nominal high river flow (for example the 1 year return period flow). The length of a river in which flow and sea level are both important, and where dependence is critical, may be relatively short.

6. DISCUSSION

6.1 The challenge of overall water level in rivers and estuaries

Determination of overall water level is a potentially complex problem, involving river flow and sea level, possibly with waves and wind set-up. There may be meteorological and/or hydraulic links between the variables. Accurate analysis of extreme values and dependences between variables is important in assessment of flood risk. Fortunately, in practice, most cases can be simplified during appropriate preliminary analysis and planning.

6.2 Preparatory work

6.2.1 Understanding and data gathering

A clear understanding of what is important, what is required and what is feasible is essential to efficient determination and analysis of extreme overall water level in rivers and estuaries. This might be based on physical reasoning, previous studies, existing hydraulic models and data availability. If a certain type of data or model is required for a particular analysis approach, then availability should be checked and acquisition (or creation) put in hand as soon as practical.

6.2.2 The importance of preliminary analysis

In the majority of cases, preliminary analysis is able to reduce the potential complexity involved in a full joint probability assessment of river flow, sea level and waves. It may be possible to show by analysis of measured data, by trial runs of a hydraulic model or by trial calculations of some impact function, that one or more of these variables has negligible effect within a particular area, and therefore that it need not be considered further. Even if the impact of one of the loading variables is not negligible, it may be sufficiently small that it can be represented by a simple allowance, as opposed to a full analysis.

A further stage of preliminary analysis might show either that two variables are independent of each other, or that the assumption of full dependence between them would not have too great an impact on the final results. The assumption either of independence or of full dependence will usually simplify the later stages of analysis.

6.2.3 The importance of event definition

The event concept is important as it provides the continuity between the data preparation, the statistical analysis and the eventual application in flood risk assessment. With the partial exception of continuous simulation modelling, all of the analysis approaches described in this report involve event definition. One parameter is the (fixed) total number of records (each of which is a potential ‘event’ if one or more thresholds are exceeded) per year, sometimes expressed as a duration or interval between records. The other parameters represent the magnitudes of the variables which characterise the individual records, i.e. river flow, sea level, wave height etc.

The number of records per year needs to be set high enough to capture a representative distribution including all noteworthy events, but small enough to avoid multiple-counting of the same noteworthy events. It might be thought that peaks over threshold would be a good approach, but this proved difficult to implement for multiple variables.

If possible, to keep the joint probability analysis manageable, the magnitude of each variable characterising a record should be represented by a single number. Although sea level changes during a tidal cycle, a record can be reasonably well represented by a single number, namely the sea level at the peak of the tide. Wave conditions in estuaries respond fairly quickly to changes in winds, and they can be adequately represented by the wave condition (usually averaged over a period of one hour) at the time of a record defined in terms of some other variable (e.g. at high tide). The overall water level, in turn, will respond rapidly to any change in sea level or waves, and in most situations is not dependent upon the length of time for which the inputs are applied. These arguments do not generally apply to river flow, where 'events' are usually expressed in terms of hydrographs occurring over periods of hours or days. They are usually represented by a peak river flow rate (a primary variable in the present context) and a duration (often fixed for a particular site, and could be used as part of the event definition, but could be treated as a secondary variable).

As a river flow effect takes time to travel downstream from the point of flow measurement, there may be a time delay involved, and so for example a sea level measured downstream should perhaps be combined with a flow rate measured some hours earlier. This time lag between variables can be incorporated into the event definition, provided this is remembered through the stages of data preparation, statistical analysis and application in flood risk assessment.

6.2.4 Data preparation

The data used in statistical analysis should be a representative sample, preferably taken over a period of at least a few years. This will usually mean 'representative' of the entire distribution of all variables concerned, from the lowest values to the highest values. However, it is possible for 'representative' to be limited to a particular category (or separate categories in separate data sets), for example the highest 10% of occurrences or corresponding to a particular type of weather condition, provided that this carries right through the analysis and the probabilities are adjusted accordingly. Data should be checked for quality and outliers. If a genuine trend (as opposed to natural variability) exists, it is usually helpful to de-trend the data to equivalent values at a common date.

In preparing multiple-variable data for dependence analysis it is usually best to discard any records where one (or more) variable is missing. The timing, construction and format of records should be consistent with the event definition being used.

6.3 The choice of joint probability analysis method

After the thorough preparatory work discussed in Section 6.2, the main joint probability analysis method will often have chosen itself, depending on which variables are important and which data are available. The number of primary variables should be kept to a minimum, and in any case no more than three. If possible, less important variables should be treated as secondary variables (for example wave period being dependent only on wave height) or proxy variables might be introduced (for example wind speed subsequently being used to produce waves at several different points in a river). The main joint probability analysis will often take less time than the preparatory work.

Figure 1 provides guidance on the choice of method and Table 1 gives a brief description of the different classes of statistical methods available. Six methods are described, depending on the number of primary variables involved, and on whether that number can be reduced by hydraulic modelling prior to statistical analysis. Variations are introduced, depending on how the results are to be expressed (joint probability density or joint exceedence extremes) and how they are to be used (for example as input to structure functions or hydraulic models).

6.4 Reliability of results

6.4.1 Discussion of uncertainties

Uncertainties and sensitivities are of interest to users of analysis methods (particularly statistical methods). The sensitivity of structure variables or decisions to changes in particular inputs, for example systematic increases of 10% in wave height or of 0.2m in sea level are fairly easy to calculate using the statistical simulation methods, just by repeating the analysis with the change(s) included. This approach might be used to test the impact of uncertainty in the loading variables and/or allowances for future climate change. In the situation of defences at risk due to wave overtopping, it would usually be necessary to raise the wall crest level by more than the assumed increase in sea level to maintain the same standard of defence under future sea level rise. Frequency of overtopping and standard of defence can be very sensitive to change in sea level, so full re-calculation of structure variables and decisions is necessary to understand the impacts of changes or uncertainties in the loading variables.

Reliable estimation of uncertainties, due both to potential errors in the input data and to statistical uncertainties, is more difficult, as discussed in Defra / Environment Agency (2002). Statistical uncertainty, due to sample size, sample variability and goodness of fit, can be evaluated objectively. Uncertainty in the source data used in any statistical analysis is usually estimated subjectively. Often the largest uncertainty arises in calculation of the structure function, e.g. overall water level based on a hydraulic model run, or overtopping rate based on an empirical relationship. Overall estimation of uncertainty needs to include all these contributory factors.

6.4.2 The challenge of structure function discontinuity

If the overall water level due to sea level, flow and waves (or any combination thereof) is thought of as a structure function, then that function has an obvious discontinuity in flood conditions when the river changes from in-bank flow to out-of-bank flow. The overall water level cannot continue to increase indefinitely when the bank crest elevation is exceeded. For the unwary, this poses a potential problem in the use of joint probability analysis. However, with a careful mix of statistical modelling of the input variables (which usually have continuous distributions) and hydraulic modelling of overall water levels (the structure variable, which may have a discontinuous distribution) realistic extreme scenarios can be predicted. Alternatively, if out-of-bank flow occurs in extreme conditions, the discharge over the bank could be considered as the end variable of interest.

6.4.3 The benefits of a joint probability approach

A common and perfectly reasonable approach to estimating extreme overall water levels in rivers is to apply, say, the 200 year return period flow hydrograph at the upstream boundary of a river model and the 200 year return period sea level at the downstream boundary. (Waves and climate change might also be considered.) However, without further work, there is no way of telling whether this scenario might occur roughly once, on average, in every two hundred years, or once, on average, every million years. (Note that 200x200 is not an upper limit to the uncertainty in the probability of this event, as that would only address the probability of both extremes occurring during the same year.) This design scenario could continue to be used in design but it would nevertheless be of interest to have some understanding of the degree of conservatism implicit in this approach.

Joint probability analysis is relevant only in estimating the likelihood of occurrence of flood risk scenarios dependent on more than one loading variable, and in this situation, it is hard to see how a study could be undertaken without joint probability analysis. However, if there is no need to estimate probability of occurrence, or if flood risk depends solely on sea level or solely on river flow, then there would be little point in using joint probability analysis.

7. ACKNOWLEDGEMENTS

In addition to the individuals named on the contract page, the authors are grateful to the following individuals and organisations for ideas, guidance and/or data. Professor Jonathan Tawn, Mr Michael Owen, Mr Ian Meadowcroft and CEH Wallingford for ideas and guidance; the Proudman Laboratory and the Environment Agency for provision of data; and colleagues Andrew Nex, Julian Hatchwell, Rob Cheetham and Paul Samuels for help with river modelling and physical appreciation of the issues.

8. REFERENCES

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

Defra / Environment Agency (2002). Risk, performance and uncertainty in flood and coastal defence: A review. Defra / Environment Agency R&D Technical Report FD2302/TR1 (also referenced as HR Report SR 587).

Defra / Environment Agency (2003a). Joint probability: Dependence mapping and best practice: Technical report on dependence mapping. Defra / Environment Agency R&D Technical Report FD2308/TR1 (also referenced as HR Wallingford Report SR 623).

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

Environment Agency (1999). Forecasting extreme water levels in estuaries for flood warning: Stage 1: [Review of current practice]. Environment Agency R&D Project Record W5/010/1, 1999.

Environment Agency (2000). Forecasting extreme water levels in estuaries for flood warning: Stage 2: Review of external forecasts and numerical modelling techniques. Environment Agency R&D Project Record W5/010/2, December 2000.

P J Hawkes, B P Gouldby, J A Tawn and M W Owen (2002). The joint probability of waves and water levels in coastal defence design. Journal of Hydraulic Research, Issue 3, Volume 40, 2002

HR Wallingford (1981). A computational hydraulic model of the River Severn: Avonmouth to Worcester flood alleviation study: Volume 3 (revised analysis of Avonmouth tidal data). HR Report EX 945, March 1981.

HR Wallingford (1993). Flood water levels in the River Severn: Gloucester to Avonmouth. HR Report EX 2907, November 1993.

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

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

HR Wallingford (2001). Joint probability: Dissemination, industry beta-testing and alternative applications. Final project report (MAFF form CSG15) on Project FD1704.

Institute of Hydrology (1998). Hydrometric Register and statistics 1991 – 1995, Institute of Hydrology.

Institute of Hydrology (1999). Flood Estimation Handbook. Institute of Hydrology report, ISBN 094854094 X.

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

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

Proudman Oceanographic Laboratory (1997). Estimates of extreme sea conditions - Final report: Spatial analyses for the UK coast. POL Internal Document No 112, June 1997.

P G Samuels and T N Burt (2002). A new joint probability appraisal of flood risk. Special edition on the Cardiff Bay Barrage, Water, Maritime and Energy 154, Proceedings of the Institution of Civil Engineers, June 2002, pp109-115.

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

C Svensson and D A Jones (2002). Dependence between extreme sea surge, river flow and precipitation in eastern Britain. International Journal of Climatology, 22 (10), pp1149-1168.

Appendix 1 The Severn Estuary Case Study

1 Introduction to the Severn

The River Severn is the longest river in Britain, with a length of approximately 350km. The source of the Severn is in the Cambrian Mountains in mid-Wales, from here the river flows north-east until the confluence with the Afon Vrynwy where it turns east towards Shrewsbury. From Shrewsbury the river flows in a south-westerly direction towards Wolverhampton, whereupon there is a transition to a more southerly flow through Bewdley, Worcester and Cheltenham. From Cheltenham, the direction of flow is south-westerly through Gloucester, continuing into the Bristol Channel. The tidal range at the estuary mouth is the second largest in the world at approximately 15m.

The purpose of this study was to investigate the influence of sea levels and river flows on the water levels within the estuary. Waves were potentially of importance in flood risk estimation at the downstream end of the study area. However, it was anticipated they could be included using joint probability methods that had already been developed (TVN and HR Wallingford with Lancaster University, 2000a) as a subsequent procedure. The primary focus of this exercise was therefore on the tide-surge-flow interaction.

2 Available data and computational hydraulic model

HR Wallingford had conducted a number of modelling investigations into the River Severn and its estuary, see for example HR Wallingford (1981, 1990 and 1991). An adapted 1d computational model described in HR Wallingford (1991) was used in this study to model the effects of the interaction between sea level and river flow.

There are a number of gauging stations throughout the estuary and measured data were obtained where possible. A summary of the data obtained for this study is provided in Table A1.

Table A1 Summary of measured data available for analysis

Location (downstream to upstream)	Tide and surge	Overall water level	River flow
Avonmouth	87-98	87-98	-
Sharpness	-	93-98	-
Epney	-	93-98	-
Minsterworth	-	93-98	-
Haw Bridge	-	87-98	87-98

For the purposes of this study, an iSIS computational model was adapted from the model that had been created during a previous study (HR Wallingford, 1991). The upstream boundary of the model was at Haw Bridge, approximately 5km north of Gloucester, chosen as it is far enough upstream to be largely unaffected by tides, and also because measured flow and water level data were available here (Table A1). The downstream boundary of the model was at Avonmouth where the effects of high river flows on the overall water level are negligible. An ‘A’ Class tide gauge is located at Avonmouth and extensive analysis of tides and surges has been carried out in the past.

The iSIS model provides a 1d representation of the interaction between tidal and river flows. Out-of-bank flows are ‘contained’ in flood plain reservoir units. The model was constructed by dividing the modelled area into a series of cross sections. A plan view of the model, showing the locations of the cross sections, reservoir units and gauging instruments is shown in Figure A1.

To assess the performance and suitability of the adapted model, a period of measured boundary data was run through the iSIS model and the results compared with the measured data at the intermediate gauging sites. The sample data period was of six month duration from September 1997 to February 1998, covering a time of year when relatively high flows and surges were anticipated. The results of this analysis are shown in Figures A2-A4.

The results for Sharpness (Figure A2) show the water level at this site to depend primarily on sea level. The model and measured data show good agreement at high water levels. This is perhaps not surprising due to the proximity of Sharpness to Avonmouth (model boundary) and the dominant role of tidal flows in this area. The discrepancy at low water levels is most likely due to drying at the location of the gauging instrument.

The influence of the river flow on the overall water level becomes apparent at Epney (Figure A3), with an underlying variation at low water levels. The tide, however, is still the dominating variable. The comparison of the data at high tide levels is reasonable, with the model tending to over predict slightly. There is a noticeable discrepancy at lower water levels during low flows, the cause of which is unknown. However, as the flow increases (at around 3000 hours in Figure A3) the model data show good agreement with the measurements.

The water level at Minsterworth (Figure A4) is obviously dependent on both sea level and river flow. The peak levels, however, remain dominated by the sea level during this period of data. The model shows a reasonable comparison with the measured data at peak levels, although it is noticeable that the model is over predicting in general terms. At lower levels the model reproduces the general trend of the water level, although with an error of over one metre in some periods, but at high flows (around 3000 hours in Figure A4) the error is relatively low. The cause of the discrepancy at lower levels is unknown.

In general terms the model reproduces the behaviour of the estuary flows in a reasonable manner. For consultancy purposes, it is quite likely that more extensive calibration would be carried out to ensure a closer fit and more realistic representation of water levels throughout the study area and throughout the tidal range. However, for the purposes of this research study, the model was sufficiently accurate to allow an unbiased comparison of the various joint probability methods.

3 Preparation of data for the joint density extrapolation approach

3.1 Representation and selection of an event

Joint density extrapolation (Method 3 in Section 4.2 of the main text) requires an event to be represented by summary statistics of the source variables that are temporally independent. The most obvious choice of summary statistics, for estimating extreme estuarine water levels within the Severn using this approach are Avonmouth sea levels and Haw Bridge flows. There are, however, a number of factors that need to be considered, including:

- the response time of the estuary catchment;
- the duration of large flow events;
- the ratio of surge magnitude to tidal range;
- the behaviour of surge events.

3.2 The potential importance of surge as a primary variable

Initial investigations regarding the structure of an event focused on the influence of surge. There was some uncertainty regarding whether surge events needed to be separated from those caused by high astronomical tides, and whether surge needed to be treated as a separate variable.

Initial exploratory analysis focused on the issue of the influence of surge, and sought to clarify whether high surge events had a particularly noticeable effect on water levels within the estuary. In other words, given a peak sea level at Avonmouth; if the contributing features of that sea level were a high surge and moderate astronomical tide or simply a high astronomical tide with little surge, would there be a noticeable difference in observed peak water levels within the estuary? If a noticeable difference had been found, this would have had implications for any subsequent joint probability assessment.

To investigate the influence of surge events, measured water level data from Minsterworth were used in conjunction with measured data from the upstream and downstream boundaries, Haw Bridge and Avonmouth respectively. Minsterworth was chosen as the site of interest since water level here is strongly influenced by both tidal and fluvial flows.

Initially, all sea levels at high tide and their respective surge elevations were extracted from the Avonmouth data. The flow rates at Haw Bridge, at the Avonmouth high tide times, were then also extracted. The concurrent Minsterworth water level data were then analysed and the peak water levels occurring within a 3 hour time period following the Avonmouth high tide times were obtained. These data were all combined to create a file consisting of:

- Avonmouth high sea levels and associated surge levels;
- Haw Bridge flows, associated with Avonmouth high tides;
- Minsterworth peak water levels, associated with Avonmouth high tides.

A series of plots, for specified Avonmouth peak sea level bands (0.2m bandwidth was considered sufficiently narrow not to bias the analysis), were then derived, showing Minsterworth water levels plotted against Haw Bridge flows. Highlighted on these plots are occurrences of high surge (>0.5m). An example plot is shown in Figure A5.

Inspection of Figure A5 shows the highlighted events with high surge values to fall within the general spread of data, with no obviously different pattern for the high surge events. This observation is typical of the various Avonmouth sea level bands investigated. Although not conclusive, this type of analysis gives confidence that subsequent joint probability analysis that considers the Avonmouth sea level as a single variable (i.e. does not consider surge as a separate variable), will not unduly bias estimation of extreme water levels at Minsterworth.

3.3 Record discretisation for temporal independence

A number of statistically based methods, such as auto-correlation analysis, were considered when determining the most appropriate way to obtain temporally independent records. However, the most practical and revealing approach was found to be visual inspection of the time series water level data at Minsterworth (Figure A4). The influence of the flow is clearly observed, although there are distinct peaks in water level that correspond with times of spring tides that occur in 14 day cycles.

Tidal/surge events are such that 14 days between peak tidal events is more than sufficient to be considered as temporally independent. However, flow events in the Severn typically last for 7-10 days, with some exceptional events lasting as long as 20 days (e.g. the flow event at approximately 3000 hours in Figure A6). Given these long duration river flow events occur only rarely, it was considered that the spring tide cycle of 14 days was appropriate for the determination of temporally independent records, with a post analysis manual check to ensure long duration flow events were not ‘double counted’.

4 Event reconstruction and structure function derivation

4.1 Event reconstruction from records produced during data preparation

The initial analysis gave confidence in the use of sea level at Avonmouth as an appropriate variable and also that a time period of 14 days was appropriate for record preparation. There remained uncertainty regarding how, from the summary statistics of peak Haw Bridge flow and peak Avonmouth sea level, the water level at Minsterworth could be determined (i.e. how could the structure function be derived?). The approach adopted involved assumptions regarding the duration and shape of the river flow hydrograph and the tidal cycle.

A unit hydrograph shape and duration were determined using a method involving specification of the peak flow and an estimation of the time to peak described in Institute of Hydrology (1999). An example of the resulting triangular hydrograph shape is shown in Figure A7 (NB: the hydrograph duration is longer than the specified event duration of 14 days, the additional time being required to ‘run the model in’).

The specification of the downstream boundary was based on measured sea level data from Avonmouth. A standard spring tide cycle was extracted from the measured data (with little or no surge present). The required sea levels were derived by specifying a peak level and scaling the extracted sea level sequence accordingly. An example of the tidal boundary conditions for a peak sea level of 9.0mODN is shown in Figure A8. Having defined the boundary conditions, the iSIS model was then used to provide predictions of water levels within the estuary.

The representation of an event, described above, involves a number of simplifying assumptions that could potentially lead to the inaccurate estimation of water levels within the estuary. To gain an insight into the performance of the simplified representation of an event, further analysis was carried out.

This analysis used the Avonmouth high tides and associated flow data that had already been extracted (used in the surge investigation described above) and also made use of data from the continuous simulation (described below). The Avonmouth sea level at high tide data were separated into 0.2m bands and the modelled water levels (obtained from the continuous simulation) at Minsterworth were plotted against Haw Bridge flows (see the small dots in Figure A9). The Minsterworth water levels that were calculated based on the simplified event representation (involving specification of an Avonmouth peak sea level and a peak Haw Bridge flow rate, described above) were then highlighted (larger blue dots in Figure A9).

Figure A9 shows the Minsterworth water level obtained through use of the simplified event representation (i.e. specification of peak sea level and flow rate, and hydrograph assumptions) to fall within the general scatter of the Minsterworth water levels obtained from the continuous simulation (i.e. results that contain no underlying simplifications regarding event duration or peak levels). Figure A9 is typical of the results observed for alternative Avonmouth sea level bands.

To investigate the effects of the discretisation of Avonmouth sea levels into 0.2m bands, the data were separated further into 0.1m bands and displayed in Figure A9 (red and black dots). This analysis shows the data from both the 0.1m bands to be evenly scattered, which indicates the 0.2m discretisation to be appropriate without biasing the results.

This analysis shows that, despite the underlying simplifications, the simplified event representation is sufficiently accurate to estimate water levels at Minsterworth comparable with data obtained without the underlying simplified assumptions.

4.2 Application of the structure function

A method of predicting water levels at Minsterworth from the peak sea level at Avonmouth and the peak flow at Haw Bridge is described above (i.e. the structure function implicit in use of the iSIS model). This section describes how the method of determining the structure function was applied to the joint density extrapolation approach (Method 3 in Section 4.2 of the main text).

The joint density extrapolation approach used in JOIN-SEA involves the simulation of thousands of ‘events’ (representing thousands of years) of the input loading variables (in this study, peak sea levels at Avonmouth and peak flows at Haw Bridge). To determine return periods of the variable of interest (water level at Minsterworth), each event needs to be transformed, via the structure function, to the variable of interest. It is impractical to run the iSIS model for each of the many thousands of simulated events. Therefore an approximation of the structure function was derived through a contour fitting procedure.

The first step of the contouring procedure was to define a representative set of events to run through iSIS. The representative events were chosen to cover a wide range of return periods that consisted of a total of 49 events:

- Haw Bridge flows: 100-700 cumecs at 100 cumec intervals;
- Avonmouth peak sea levels: 6.5-9.5mODN at 0.5m intervals.

The results of the model runs provided predictions of water levels along the entire estuary length, from which the results for Minsterworth were extracted. A contouring procedure that fits a series of bi-cubic splines was then applied to these results. The generated contours are shown in Figure A10. These contours represent the structure function for Minsterworth water levels.

4.3 Use of the Monte Carlo simulated data

The concurrent data on Avonmouth high sea levels and Haw Bridge flows were used as input to the JOIN-SEA software. The fitting of the marginal and dependence distributions was carried out as described in Section 3.4 in the main text of this report, and the Monte Carlo simulation produced a data set representative of many thousands of years. This simulated data set was combined with the contouring analysis to transform each of the simulated records with reference to the structure function contours, and subsequently to determine return period estimates of water levels at Minsterworth. In view of the large amount of data available through the Monte Carlo simulation, return period water levels at Minsterworth were obtained simply by counting back through the structure function values arranged in descending order. (For example, in a 1000 year simulation, the 100 year return period prediction would be taken as the average of the tenth and eleventh largest values.)

4.4 Alternative extreme values derived from continuous simulation

For comparative purposes, the continuous simulation or ‘structure function’ approach (Method 2 in Section 4.2 of the main text) was also used to estimate extreme water levels at Minsterworth. This method used the concurrent measured data at Avonmouth and Haw Bridge to simulate a continuous time series of data at Minsterworth (NB: for practical purposes the ‘continuous’ model run was separated into approximately 6 month batches). The simulated time series data were then subject to a univariate extremes procedure that involved POT analysis, prior to fitting a GPD distribution.

5 Summary and results

The results of the joint density extrapolation approach and the continuous simulation approach (Methods 3 and 2 in Section 4.2 of the main text) are contrasted in Figure A11. Both predictions of Minsterworth water level against return period follow a similar pattern of a limited rate of increase in water level at high return periods. This feature is also apparent in the structure function contours (Figure A10) which show little increase in water level for Avonmouth sea levels above 8m and Haw Bridge flows greater than 500 cumecs. The cause of this is out-of-bank flows filling flood plain storage areas and thus limiting the water levels occurring at Minsterworth. This is obviously a natural feature that the model is reproducing, but it does, however, limit the value of comparison between the two methods at this site.

One of the main benefits of the joint density approach, over the continuous simulation approach, relates to the hydraulic modelling of high return period (more extreme than occurring within the measured data) events that has to be carried out. If any differences in behaviour occur at higher return period events, this method will incorporate these effects. As the continuous simulation method involves the extrapolation of data to high return periods, any differences in hydraulic behaviour outside the range of the source data are not accounted for. In the case of the Severn, it is apparent that the measured data contain a sufficient number of events that incorporate the out-of-bank flow. The extremes distribution resulting from the continuous simulation approach was therefore able to account for the change in hydraulic behaviour and can therefore be considered a good approach to use at this location.

The small differences in water level estimates at high return periods provide confidence in the methods employed in the joint density approach, particularly that of the event representation. The underlying assumptions regarding hydrograph shape and duration appear to provide a sensible and practical approach to use when applying the joint density extrapolation method.

6 References

HR Wallingford (1981). River Severn, Avonmouth to Worcester flood alleviation study, Severn Trent Water Authority. HR Wallingford Report EX 945.

HR Wallingford (1990). Flood alleviation at Gloucester. Gauge data requirements for proposed physical model. HR Wallingford Report EX 2154.

HR Wallingford (1991). River Severn Flood Plain, Avonmouth to Worcester: Computational model study. HR Wallingford Report EX 2252.

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

Institute of Hydrology (1999). Flood Estimation Handbook. Institute of Hydrology report, ISBN 094854094 X.

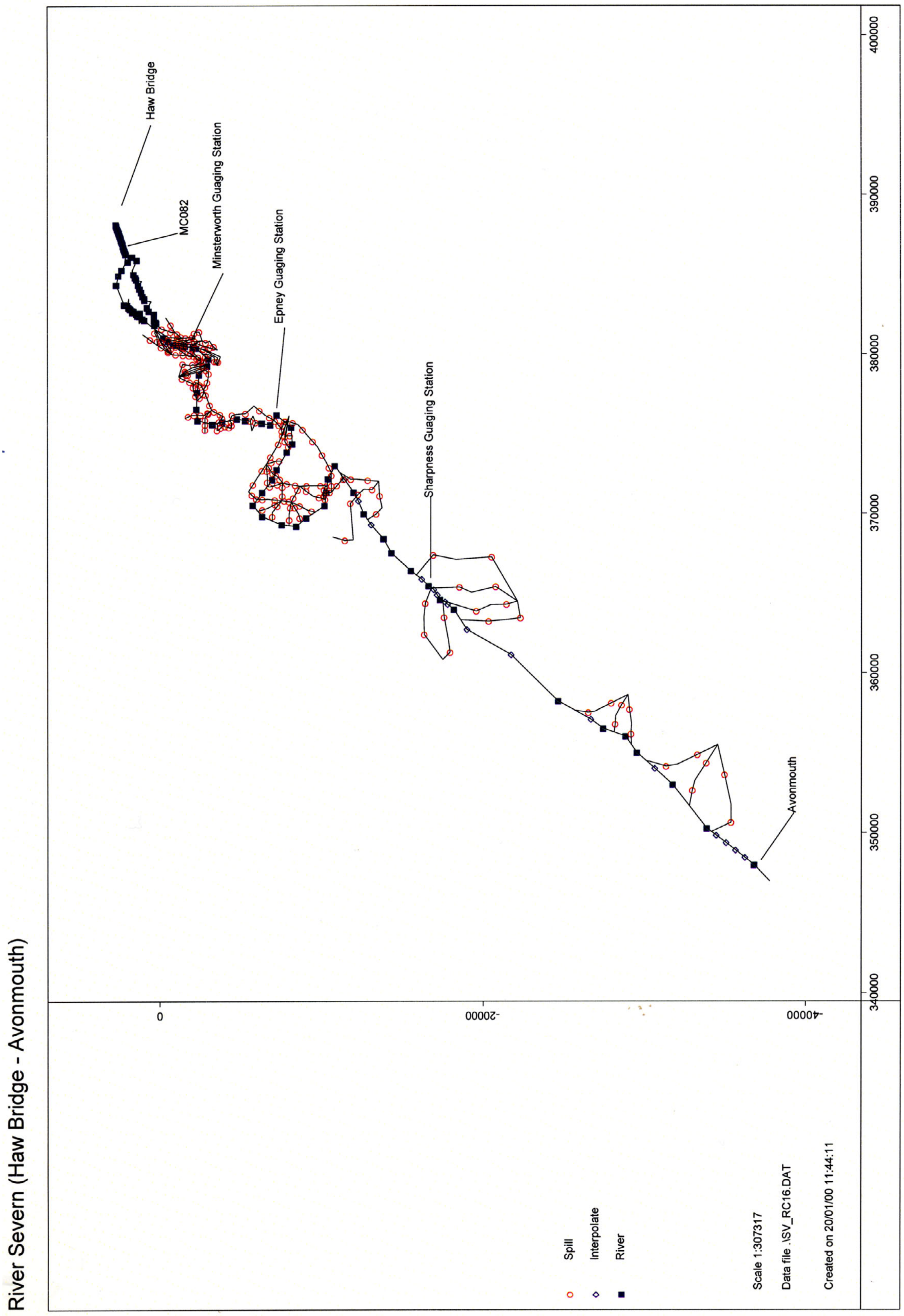


Figure A1 Schematic representation of the iSIS model of the Severn

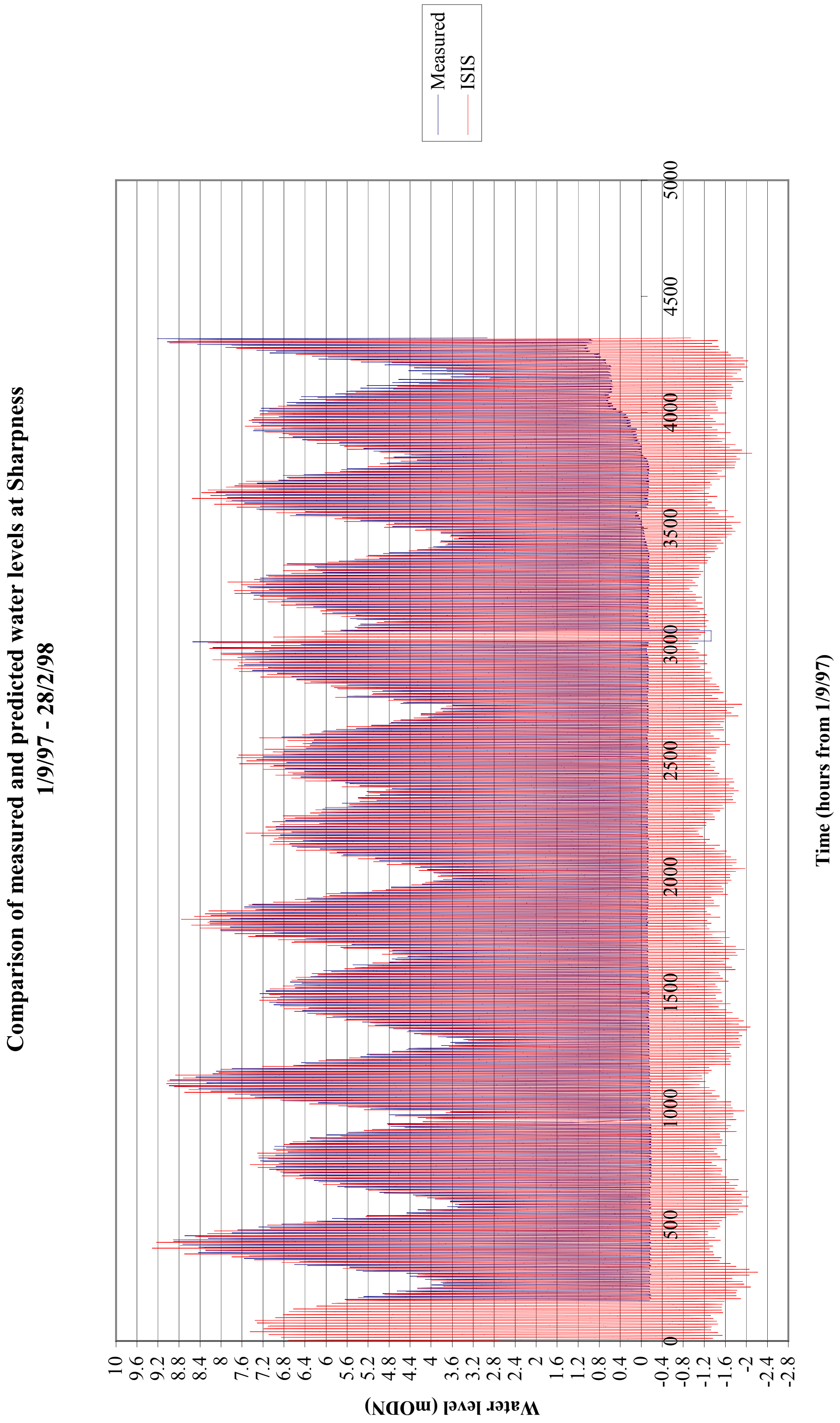


Figure A2 Comparison of measured and predicted water levels at Sharpness 1/9/97 – 28/2/98

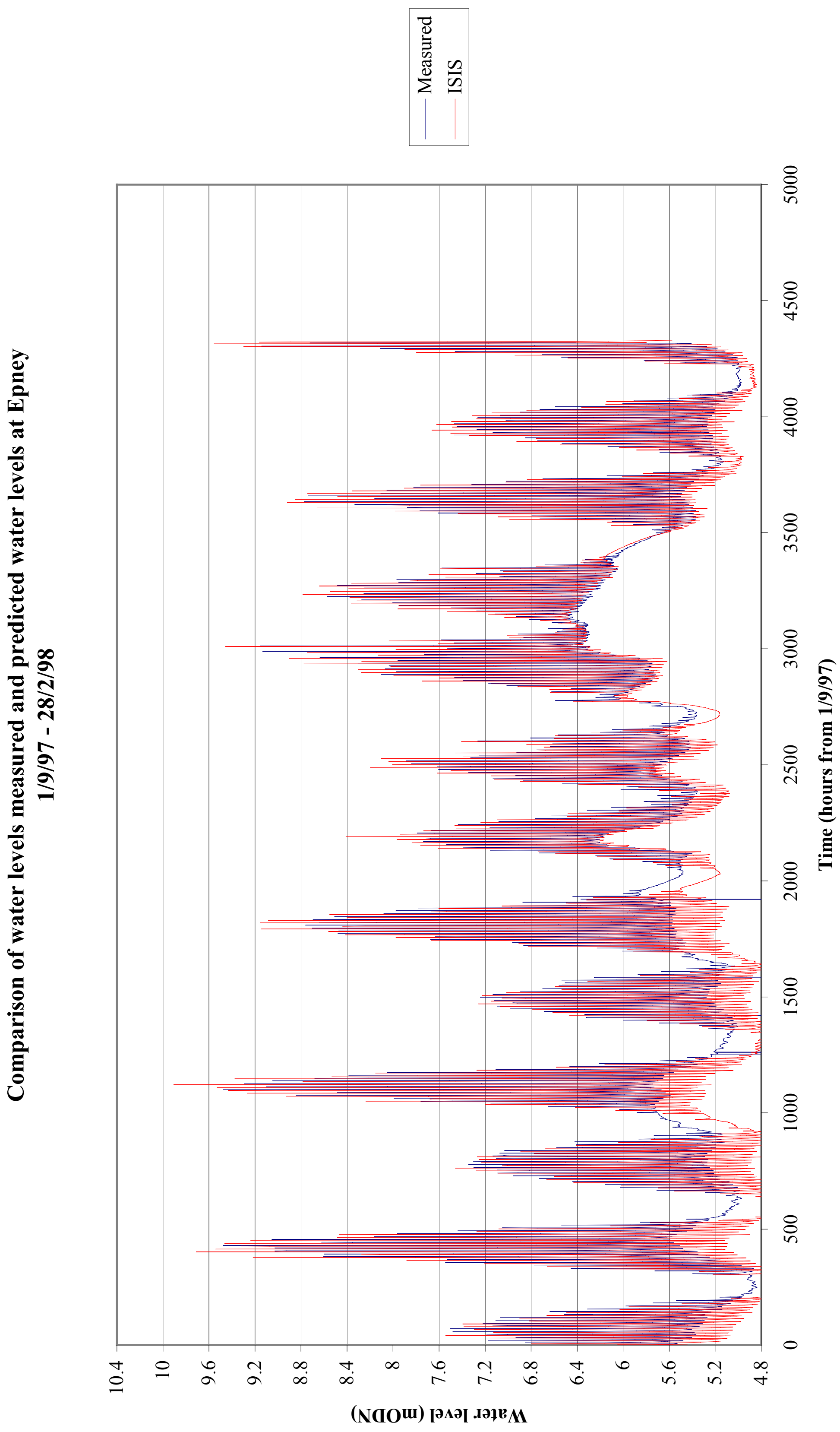


Figure A3 Comparison of water levels measured and predicted water levels at Epney 1/9/97 – 28/2/98

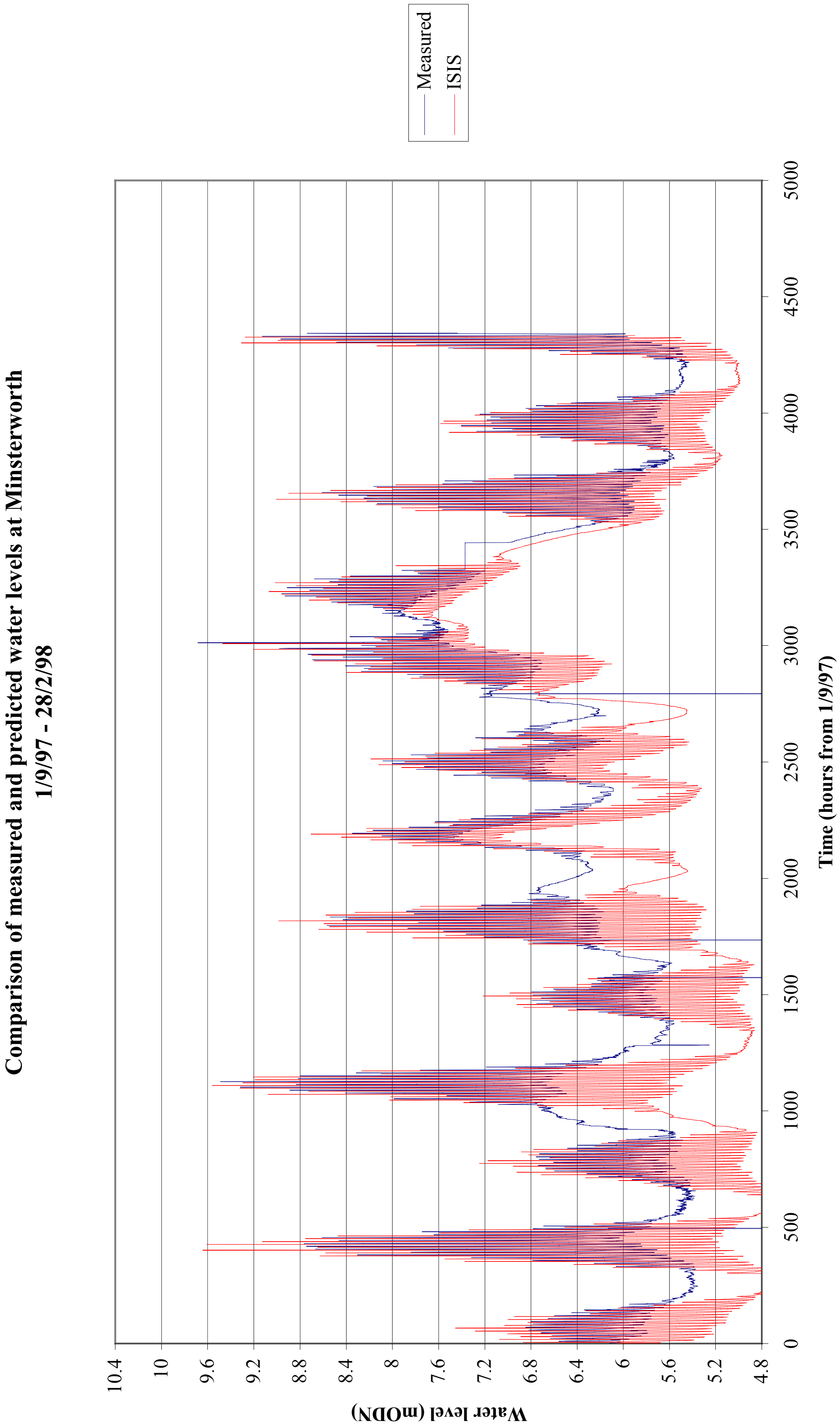


Figure A4 Comparison of measured and predicted water levels at Minsterworth 1/9/97 – 28/2/98

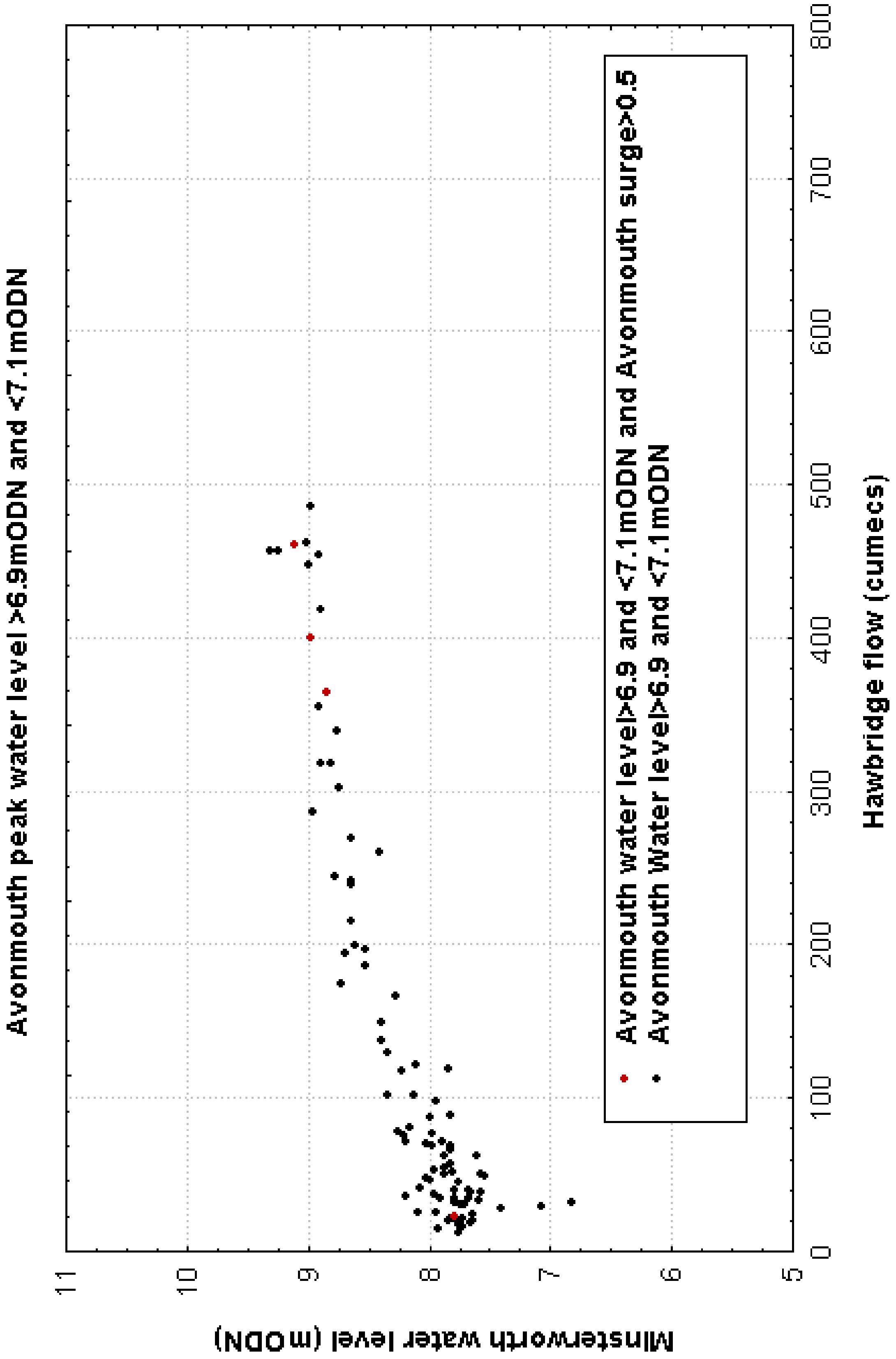


Figure A5 Measured water level at Minsterworth against measured flow at Haw Bridge for a given sea level measured at Avonmouth

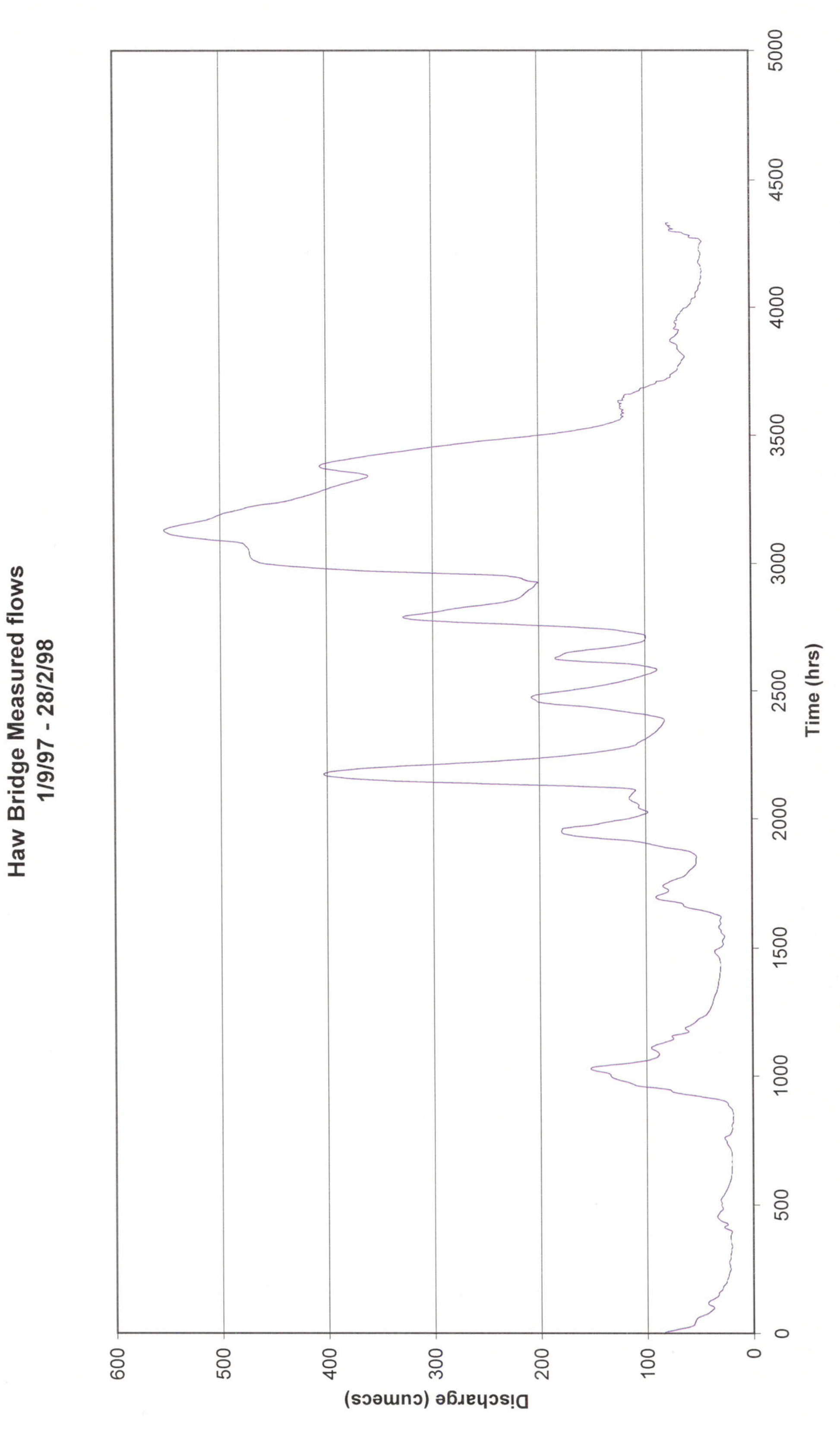


Figure A6 Haw Bridge Measured flows 1/9/97 – 28/2/98

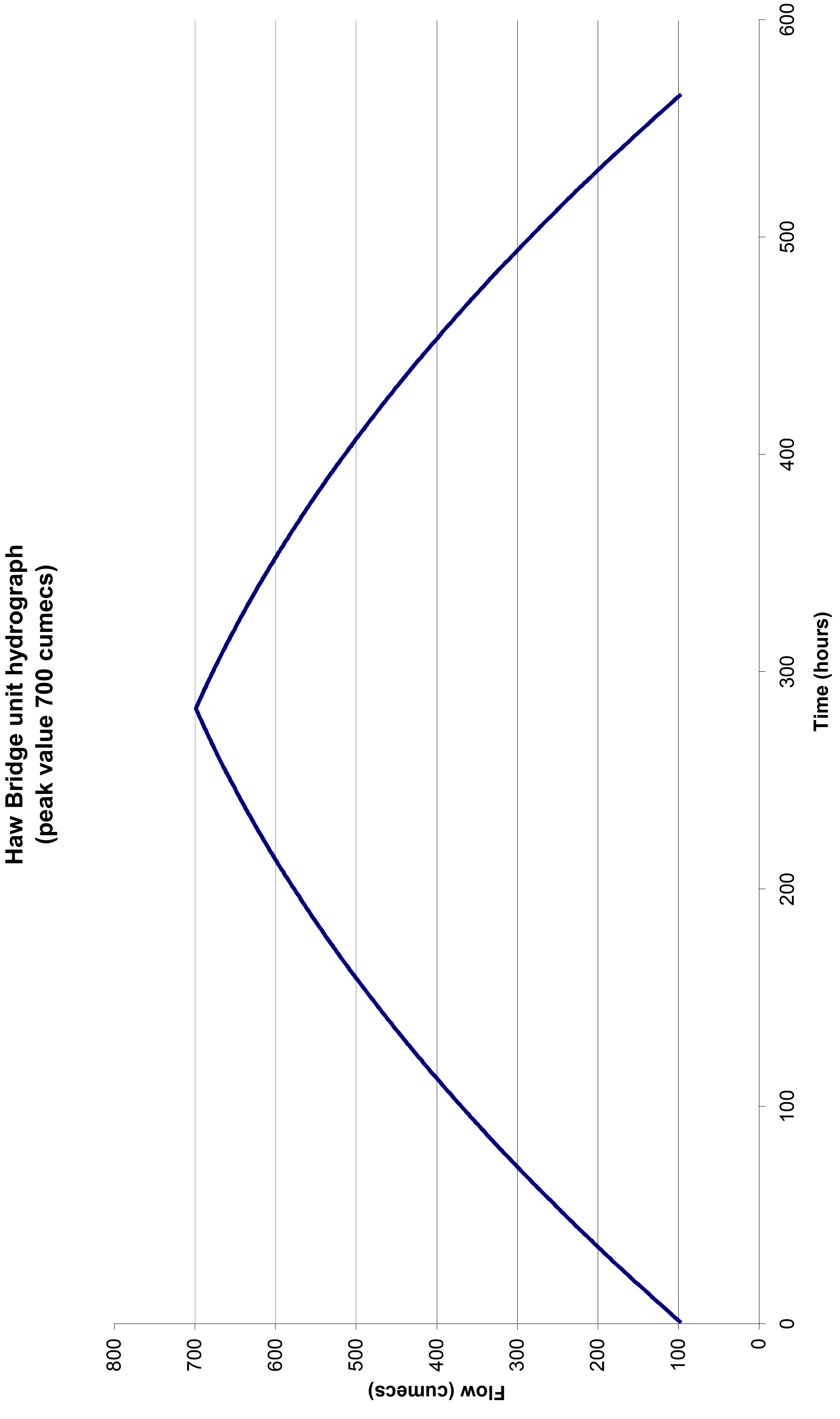


Figure A7 Hawbridge unit hydrograph (peak value 700 cumecs)

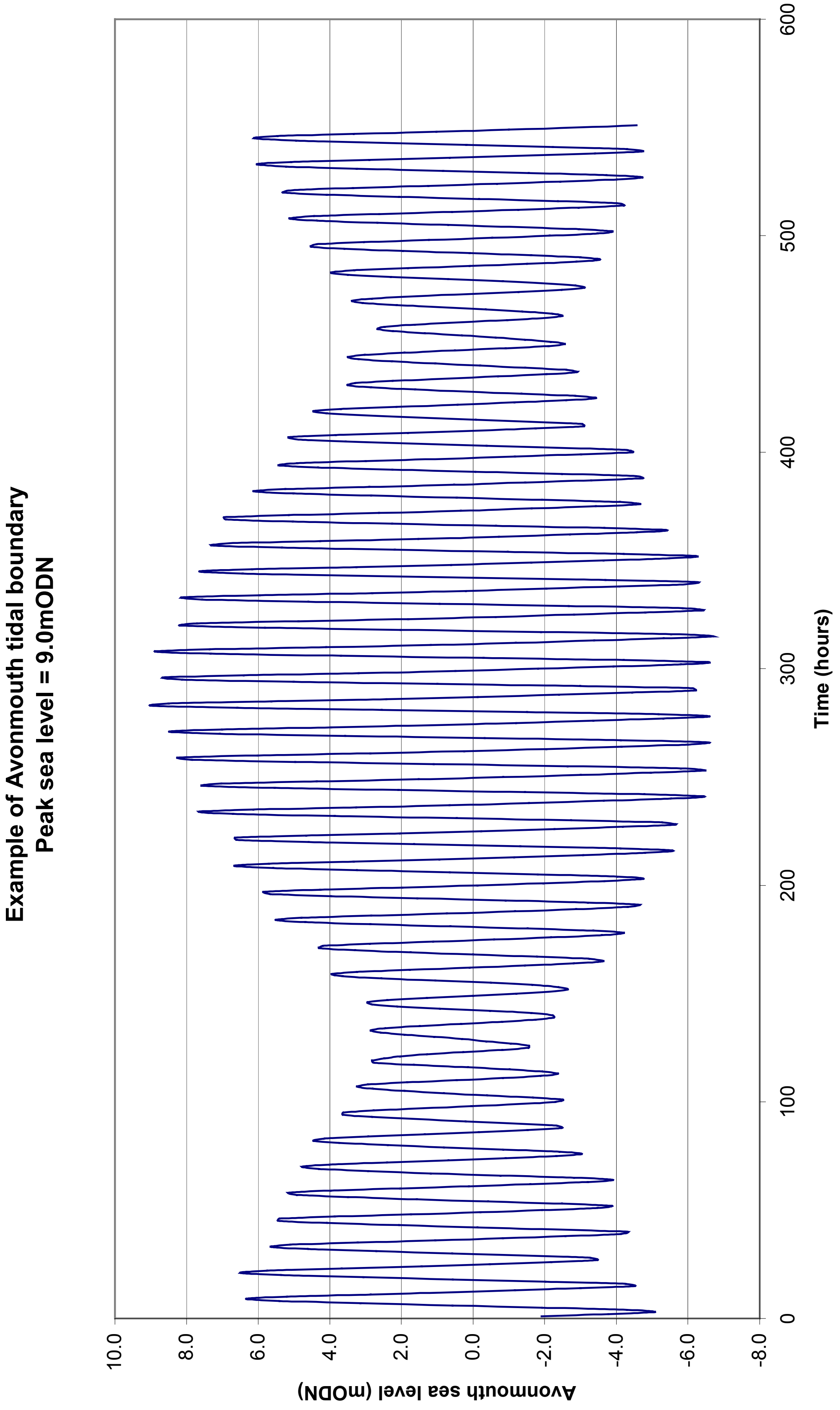


Figure A8 Example of Avonmouth tidal boundary Peak sea level = 9.0mODN

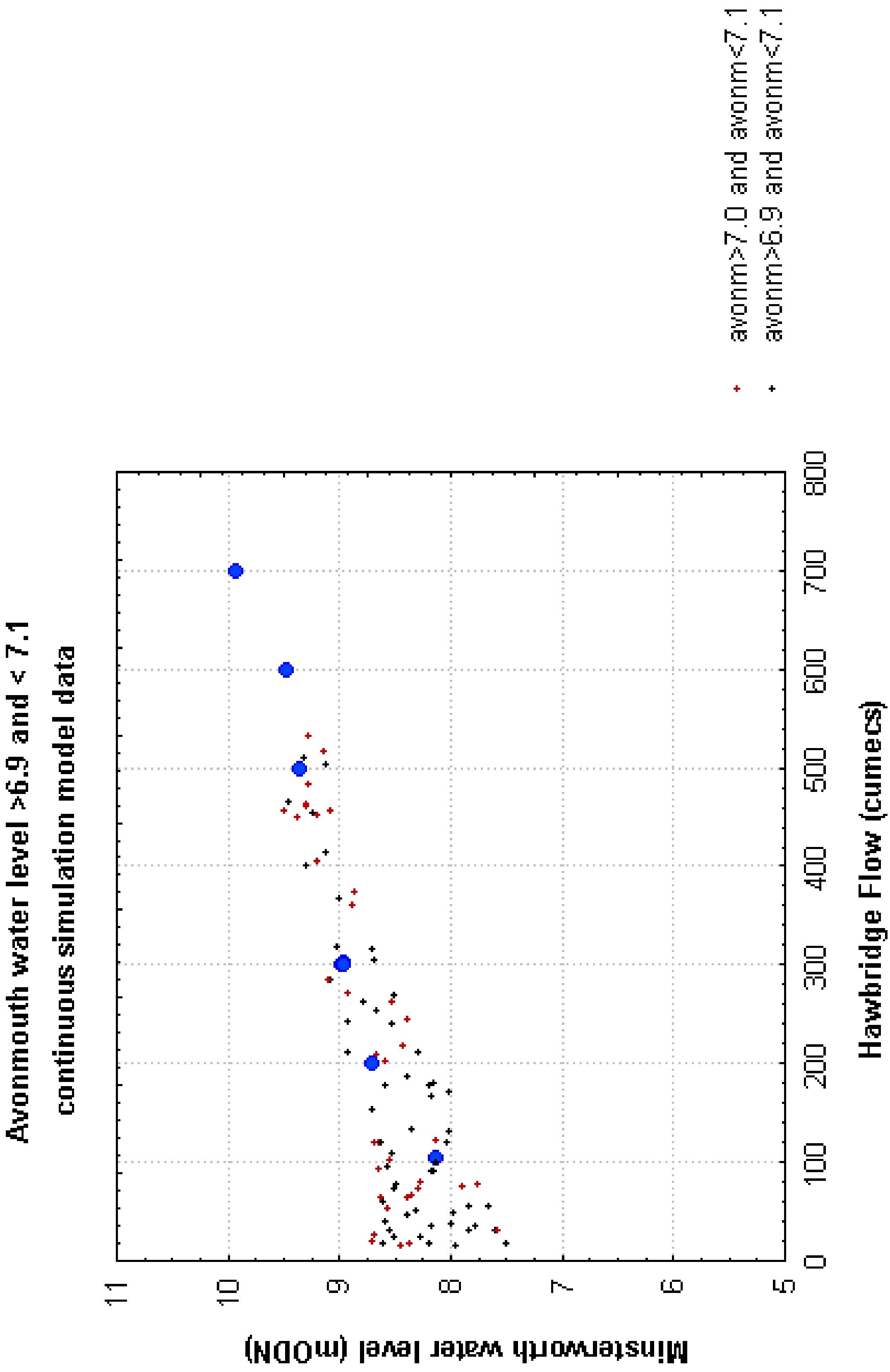


Figure A9 Minsterworth water level from continuous simulation against measured flow at Haw Bridge for a given sea level measured at Avonmouth

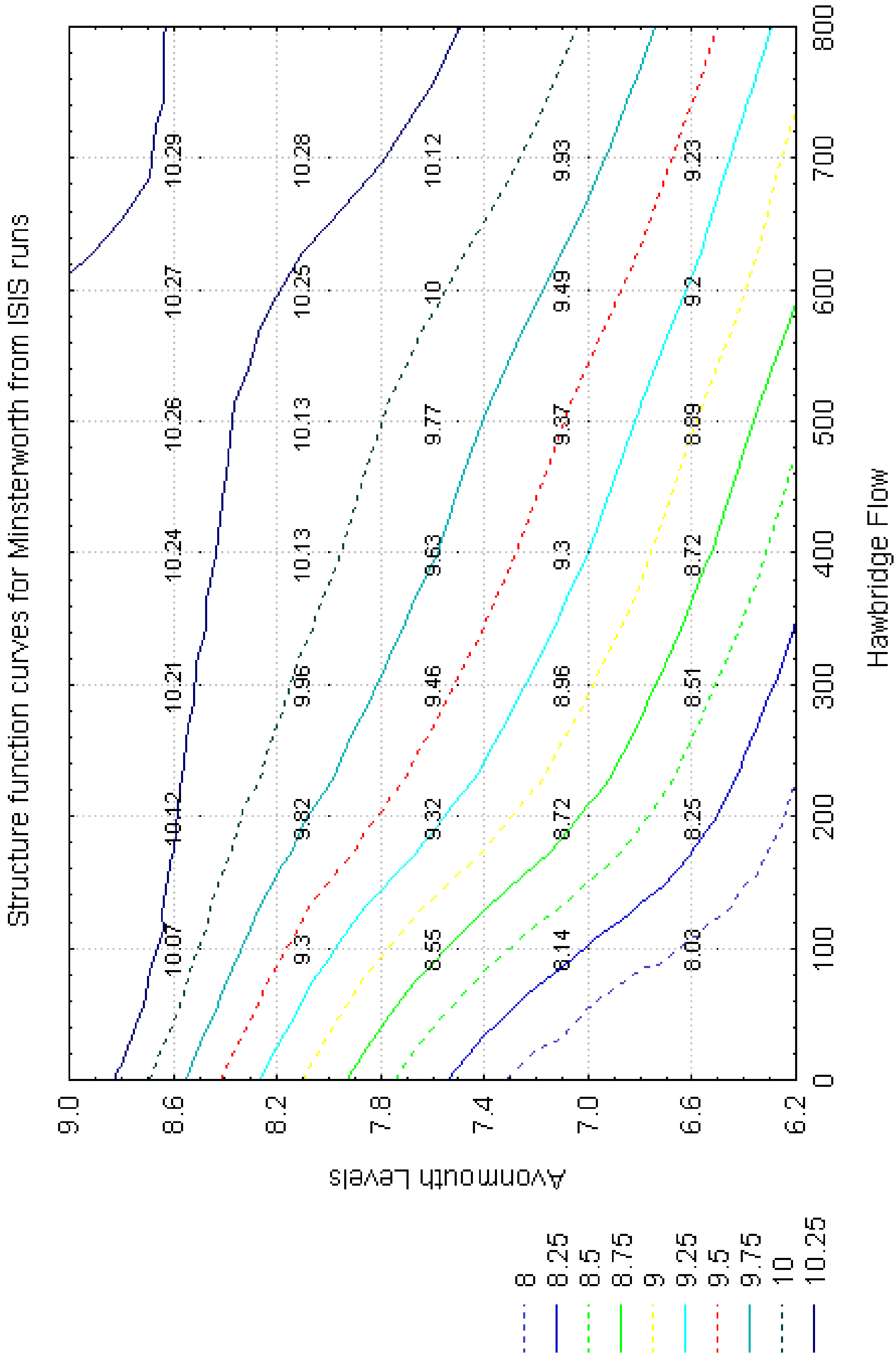


Figure A10 Contours of Minsterworth water level (mODN) predicted by iSIS as a function of Avonmouth sea level (mODN) and Haw Bridge flow (cumecs)

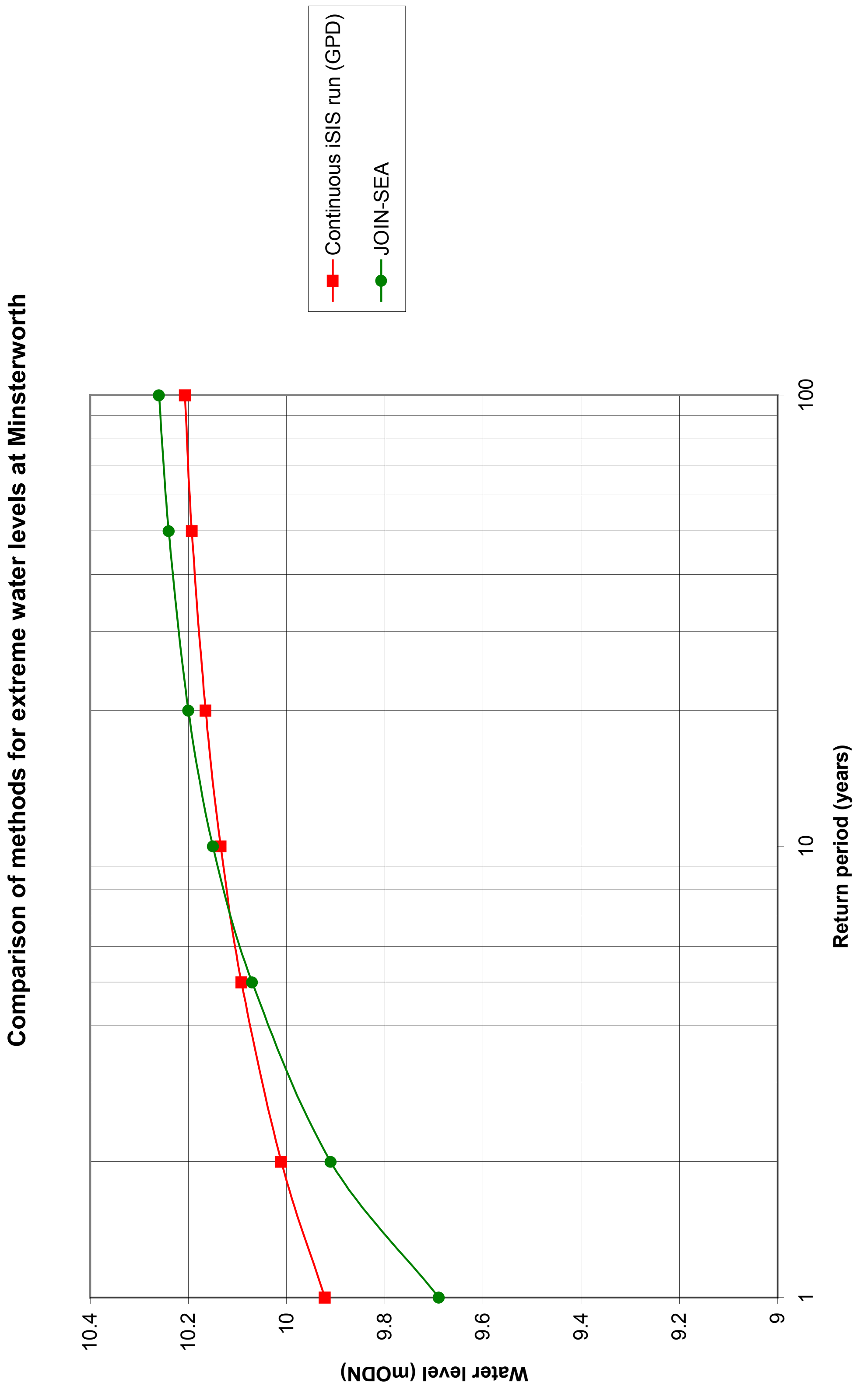


Figure A11 Prediction of extreme water levels at Minsterworth using two alternative joint probability analysis methods