



University  
of Glasgow

Reungoat, Anne Françoise Jeanne (2004) *Classification of river networks for prediction in ungauged basins*. PhD thesis.

<http://theses.gla.ac.uk/6331/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

# Classification of River Networks for Prediction in Ungauged Basins

by

Anne Françoise Jeanne Reungoat

**Thesis Submitted in Fulfilment of the Requirements  
for a Degree of  
Doctor of Philosophy**

to

**The University of Glasgow**

**Faculty of Engineering**

**Department of Civil Engineering**

**Glasgow, UK**

**November, 2004**

## Abstract

The majority of the world's river basins remain ungauged and, therefore, the tried-and-tested empirical techniques for predicting floods and droughts cannot be applied. An alternative approach, which is currently receiving a great deal of attention from research hydrologists, is to develop continuous simulation models whose parameters pertain to physical or hydrological properties of the river basins. However, difficulties related to scale, heterogeneity and complexity of real river basins have made *a priori* estimation of such parameters impossible: their estimation has always required calibration using river flow data. Therefore, estimating hydrological model parameters in ungauged river basins is one of the greatest challenges currently facing research hydrologists. In this thesis research advances towards this goal have been made at three different levels.

First, at a conceptual level, a novel method for classifying river basins according to their physical properties is proposed. It is specifically designed for transferring hydrological model parameters from gauged river basins, where calibration is possible, to ungauged river basins. This approach relies on recognising that river basins can be similar in parts of their hydrological cycle but not in others. Thus, basins go through three independent classifications, one relative to each of the major components of the land phase hydrological cycle: interaction of soil water / vegetation and atmosphere; surface flow; and groundwater flow. This requires the ability to characterise the response of the components of the hydrological cycle independently, which leads to a second conceptual advance; rather than relying entirely on measured river flow data, from which it is difficult to separate out the effects of the three components, classification rules are devised on the basis of synthetic data produced by comprehensive, distributed, physically-based models.

This thesis focuses on the surface flow component, applying the methodology to the identification of the best classifiers for surface flow through river networks. This required simulating river flow through a large number of Scottish river basins, which led to more practical research advances; all available commercial flow routing models were too cumbersome and required an impractical level of

detail to be applied in such a large study. Therefore, a new flow routing modelling system was developed that extracts river network detail from digital databases and numerically solves a distributed flow routing model.

Finally, on a detailed scientific level, significant insights have been made into the relationship between river network geomorphologic structure and stream flow response. In particular, it is shown that: a downstream hydraulic geometry relationship exists for Scottish rivers; although channel conveyance is a key factor in dictating network response, the features of the response hydrograph – namely the percentage attenuation of the flood peak and the lag in time to peak – scale linearly with both roughness and hydraulic geometry coefficients; much publicised invariant power law scaling rules for flood peaks in fact vary as a function of storm duration; statistical multivariate analysis of the simulated network flow responses demonstrated the low capacity of the network descriptors commonly used in regionalisation studies for characterising flow response. Four variables are shown to have significantly higher classifying power than the majority of the commonly used classifiers. Of these, two are entirely new to this thesis.

## **Acknowledgements**

I would like to thank Dr. Bill Sloan for providing me with an inspiring research project and vision. His accessibility, enthusiasm and pertinent advice have been of great help throughout this research.

Moreover, I wish to express my thanks to Dr. Rebecca Lunn and Andrew Aitken at Heriot Watt University, for their technical assistance with GIS and other computer system issues.

I am also grateful to all at the University of Glasgow's Department of Civil Engineering for their kind help and support.

Finally, I wish to warmly thank my partner, Bryan G. Little, my parents and sisters, and my friends for their precious advice and continuous support.

# Table of Contents

<b>Chapter 1</b>	<b>1</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Estimating hydrological model parameters in ungauged basins	2
1.2 A new approach to classifying river basins for regionalisation of hydrological model parameters	4
1.3 Summary	4
<b>Chapter 2</b>	<b>6</b>
<b>2 Literature review: River basin classification</b>	<b>6</b>
2.1 Regional flood frequency analysis	6
2.2 Flood statistics in ungauged river basins	8
2.3 Estimating hydrological model parameters for ungauged river basins	9
2.4 Other methods used for river basin classification	12
2.5 Summary	13
<b>Chapter 3</b>	<b>14</b>
<b>3 Framework</b>	<b>14</b>
3.1 A new scheme for classifying river basins - concepts	14
3.2 Deriving the new classification scheme - method	15
3.2.1 Generation of river network data	16
3.2.2 Flow routing simulation	17
3.2.3 Multivariate statistical analysis	18
3.3 Summary	19
<b>Chapter 4</b>	<b>21</b>
<b>4 Flow routing model</b>	<b>21</b>
4.1 Channel flow routing	24
4.1.1 Governing equations: the kinematic wave model	24
4.1.2 Numerical solution to the kinematic wave equations	26
4.1.3 Stability and convergence of the numerical scheme	29
4.2 From channel routing to network routing	38
4.3 Summary	40
<b>Chapter 5</b>	<b>41</b>
<b>5 Network Structure Data</b>	<b>41</b>
5.1 Derivation of a GIS database for the physical structure of Scottish rivers	42
5.2 Automated extraction of catchment network data	45
5.2.1 Catchment network extraction	45
5.2.2 Generating 'hydrologically consistent' network data based on digital terrain data	46
5.2.3 Automation of the catchment network data extraction	51
5.3 Summary	53
<b>Chapter 6</b>	<b>54</b>
<b>6 Estimation of conveyance parameters</b>	<b>54</b>
6.1 A brief review of generic patterns in river cross-section geometry	55

6.2 Validation of a downstream hydraulic geometry model for Scottish rivers	58
6.3 Application of the top width model to all the reaches in a network	61
6.4 Roughness coefficient	62
6.5 Summary	63
<b>Chapter 7</b>	<b>64</b>
<b>7 Simulation</b>	<b>64</b>
7.1 Input hyetographs and response hydrographs	64
7.1.1 Hyetographs	65
7.1.2 Lateral inflow	68
7.1.3 Response hydrograph variables	68
7.2 Preliminary simulations	69
7.2.1 Do different shaped networks respond differently?	69
7.2.2 Importance of the conveyance parameters	74
7.2.3 The importance of scale	77
7.3 'At a scale' simulations	85
7.3.1 Definition of the scale criteria	85
7.3.2 Isolation of networks with same total length	86
7.3.3 Synthetic channel top widths	87
7.3.4 Simulation results	87
7.3.5 Results of simulation for another conveyancing value	88
7.3.6 Results of simulation with real channel top widths	89
7.3.7 Results of simulation at another scale	90
7.4 Summary	91
<b>Chapter 8</b>	<b>93</b>
<b>8 Statistical multivariate analysis</b>	<b>93</b>
8.1 Cluster analysis of simulated network responses	93
8.1.1 Clustering method	93
8.1.2 Network response clusters at 300-kilometre total length scale	96
8.1.3 Network response clusters at 600-kilometre total length scale	99
8.2 Discriminant analysis	101
8.2.1 Method	101
8.2.2 Network descriptors	104
8.2.3 Discriminant analysis at 300-kilometre scale	110
8.2.4 Discriminant analysis at 600-kilometre scale	119
8.3 Best Network Classifiers	124
8.4 Summary	126
<b>Chapter 9</b>	<b>128</b>
<b>9 Conclusions and Future Research</b>	<b>128</b>
9.1 Conclusions	128
9.2 Future Research	131
<b>References</b>	<b>133</b>
<b>Appendix A – List of catchment outlets</b>	<b>141</b>
<b>Appendix B – Network descriptor values</b>	<b>144</b>

## List of Figures

<b>Figure 3.1</b>	Schematic diagram of the best classifiers objective test	19
<b>Figure 4.1</b>	Schematic diagram of channel	30
<b>Figure 4.2</b>	The three-part analytical solution	31
<b>Figure 4.3</b>	Comparison of analytical and numerical results	38
<b>Figure 4.4</b>	Network channels ordering system	39
<b>Figure 5.1</b>	Raster representations of terrain elevation and flow accumulation variables: (a) Elevation; (b) Flow accumulation.	43
<b>Figure 5.2</b>	Digitised river network (blue lines) and DTM-generated flow paths derived for 3 decreasing flow accumulation threshold values: (a) Threshold $0.250\text{km}^2$ ; (b) Threshold $0.215\text{km}^2$ ; (c) Threshold $0.025\text{ km}^2$ .	48
<b>Figure 5.3</b>	Watershed delineation, source-point localisation and drainage network derivation	49
<b>Figure 5.4</b>	Sampled examples of extracted DTM & sources – generated catchment drainage networks (red lines) compared with digitised rivers (blue lines)	50
<b>Figure 5.5</b>	Catchment positions on the map of Scotland	53
<b>Figure 6.1</b>	Derivation of at a station hydraulic geometry	59
<b>Figure 6.2</b>	Derivation of a downstream hydraulic geometry model for Scottish rivers	60
<b>Figure 6.3</b>	Downstream hydraulic geometry model; 95% confidence regression analysis.	61
<b>Figure 7.1</b>	Profile of the 1-hour design storm used for simulation	66
<b>Figure 7.2</b>	Profile of the 1-day design storm used for simulation	67
<b>Figure 7.3</b>	Profile of the 6-hour design storm scenario	67
<b>Figure 7.4</b>	Profile of the 12-hour design storm scenario	68
<b>Figure 7.5</b>	The network extracted from GIS databases which is used as 'model' for generation of synthetic networks. It has code w3001 (see Appendix A).	70
<b>Figure 7.6</b>	Diagram of 2 synthetic networks with same total length: (a) Synthetic Network 1; (b) Synthetic Network 2 (partial representation).	71
<b>Figure 7.7</b>	Response points for 3 different shaped networks; 2 synthetic networks and an extracted network which have the same drainage area, total length and average slope.	72
<b>Figure 7.8</b>	Multiple realisations of flow simulated for 48 catchments	74
<b>Figure 7.9</b>	Variability of network hydrograph responses with conveyance parameters: (a) Percentage attenuation of peak flow; (b) Lag in time to peak.	75
<b>Figure 7.10</b>	Simulated peak flows along 8 networks using the 24-hour duration storm	80

<b>Figure 7.11</b>	Simulated peak flows along 2 synthetic networks and an extracted network which have the same drainage area, total length and average slope	81
<b>Figure 7.12</b>	Logarithmic plots of simulated peak flows along 8 networks for 5 storm durations: (a) 24-hour storm; (b) 12-hour storm; (c) 6-hour storm; (d) 1-hour storm.	82
<b>Figure 7.13</b>	Simulated responses for 46 networks of 300km total length	88
<b>Figure 7.14</b>	Simulated responses for conveyance parameters (a,n) equal to (3.64,0.067)	89
<b>Figure 7.15</b>	Simulated responses with 'real' channel top widths	90
<b>Figure 7.16</b>	Simulated responses for 24 networks of 600km total length	91
<b>Figure 8.1</b>	Dendrogram suggesting an optimised partition of 3 groups	95
<b>Figure 8.2</b>	Hierarchical agglomerative clustering of the hydrographs for the 46 networks of 300km total length: (a) Dendrogram using Ward's linkage method; (b) Initial partition.	98
<b>Figure 8.3</b>	Optimised partition; final 300km network response clusters.	99
<b>Figure 8.4</b>	Hierarchical agglomerative clustering of the hydrographs for the 24 networks of 600km total length: (a) Dendrogram using Ward's linkage method; (b) Initial partition.	100
<b>Figure 8.5</b>	Optimised partition; final 600km network response clusters.	101
<b>Figure 8.6</b>	300km network grouping using the identified best classifiers	118

## List of Tables

<b>Table 5.1</b>	Sample of an extracted catchment network data file	52
<b>Table 8.1</b>	Correlation matrix for the 300km network descriptor variables (upper number = correlation coefficient; lower number = p-value)	112
<b>Table 8.2</b>	Results of discriminant analysis at 300km scale. The variables are listed in decreasing order of discriminating contribution.	116
<b>Table 8.3</b>	Linear discriminant function coefficients and derived potential classifier's discriminating powers at 300km scale	118
<b>Table 8.4</b>	Correlation matrix for the 600km network descriptor variables (upper number = correlation coefficient; lower number = p-value)	120
<b>Table 8.5</b>	Results of discriminant analysis at 600km scale. The variables are listed in decreasing order of discriminating contribution.	123
<b>Table 8.6</b>	Linear discriminant function coefficients and derived potential classifier's discriminating powers at 600km scale	124

## *Chapter 1*

### **1 Introduction**

Hydrological models are currently employed to assess a wide variety of water resource and environmental issues such as regional water resource planning, flood risk assessment or assessing the effects of climate and land use changes on the hydrology of river basins. Ideally, all hydrological modelling approaches should be conducted using distributed physically based models that partition the region into small control volumes and describe the hydrological processes using the underlying physical equations. Theoretically, the parameters of such models could be directly measured in the field. However, the large amount of computer processing and data collection required for application of these types of models limit their practicality (Dooge, 1982). The alternative more pragmatic modelling approaches (e.g. Beven, 1997; Maidment et al, 1996) require some element of calibration using hydrological data. Commonly, these hydrological data are river flows, which are expensive to collect and usually sparse for large river basins. In many parts of the world, river basins are ungauged or poorly gauged, and in some cases, existing measurement networks are declining (Sivapalan, 2003). As a consequence, for the majority of river basins, model parameters cannot be calibrated. This leaves the problem of estimating model parameters for ungauged river basins, which has recently become a keenly researched topic in hydrology. The International Association of Hydrological Sciences (IAHS) launched the IAHS decade on Prediction in Ungauged Basins (PUB) in 2003. This is a new initiative aimed at engaging the scientific community in formulating and implementing appropriate science programmes to make major advances in the capacity to predict the hydrology of ungauged basins (Sivapalan et al, 2003). This initiative recognises that whichever of the currently available approaches is implemented – including extrapolation of hydrological response from gauged to ungauged basins; measurements by remote sensing; application of process-based

hydrological models – a key difficulty is that the model predictions cannot be conditioned or validated by observations in the ungauged basin of interest. Therefore, predictions in ungauged basins involve extrapolation of some kind from gauged to ungauged basins. The IAHS initiative focuses on reducing the uncertainty of predictions in ungauged basins through reducing uncertainties in the three components constituting a general hydrological prediction system: the model structure; the model parameters; the climatic inputs. This research participates actively in the realisation of these goals by aiming at improving methods for estimation of hydrological model parameters in ungauged basins.

## **1.1 Estimating hydrological model parameters in ungauged basins**

The approaches previously used to estimate hydrological model parameters in ungauged river basins all derive from techniques used in regional flood frequency analysis where the objective is to infer river flow statistics for ungauged river basins, from those for gauged river basins. These techniques, generically called regionalisation techniques, involve 3 stages: the first is a classification stage and consists of grouping together the river basins whose river flow regimes are likely to be similar; the second stage uses the flow data from gauged river basins of each group (or region) to determine regression equations relating flow statistics and the physical / climatic river basin characteristics; the final stage consists of applying the regression equations to infer river flow statistics in the ungauged river basins of each region.

When applied to estimate model parameters in ungauged basins, these techniques are modified so that flow statistics are replaced by model parameters. However, the success of these modified regionalisation techniques relies heavily on the reliability of the initial classification stage, which, in most published studies, is not tested or has been completely neglected. Since even the simplest hydrological models seek to simulate not just a particular flow statistic, but three major components of the land phase of the hydrological cycle – interaction of soil water / vegetation and atmosphere; surface flow; groundwater flow – classification

schemes for estimating model parameters have sought groups of basins that are similar across all components of the land phase of their hydrological regime. The main difficulty in this arises from the fact that two river basins can exhibit strong similarities in parts of their hydrological regime whilst being very dissimilar in others. For example, it is possible that the topography and land uses of two adjacent river basins are sufficiently alike for their surface water flow regimes to be similar, but that they are underlain by very different geology, which makes their groundwater flow regimes likely to be dissimilar. Seeking groups of river basins that are homogeneous across all parts of their hydrological regime leads to a large number of groups, each of which contains only a few basins. It is therefore difficult to have confidence in the regression equations derived for inferring model parameters. An alternative approach, that has been suggested by some researchers, is to relax the degree of homogeneity in the classification stage, then create larger but less homogeneous groups. Again, it is difficult to have confidence in the regression equations derived from such groups since they have to explain more of the variability in model parameter values using basin characteristics. Studies such as Abdulla and Lettenmaier (1997) and Sloan (1999) report significant errors in regional modelling studies of the same area in central USA, which they both attribute to the fact that regionalisation of model parameters had been conducted within heterogeneous groups of sub-basins. If regionalisation techniques are to be improved for prediction in ungauged river basins, there is a need for a reliable method of classifying river basins into hydrologically homogeneous groups. The overall aim of this thesis is to develop a new classification scheme specifically for regionalising hydrological model parameters.

## **1.2 A new approach to classifying river basins for regionalisation of hydrological model parameters**

The classification scheme proposed in this thesis differs from the previous approaches alluded to in section 1.1 and detailed in Chapter 2, in two significant ways. The first is to discard the constraint that each basin only belongs to one group, and propose that it undergoes three independent classifications, one for each of the major components of the land phase hydrological cycle introduced earlier: interaction of soil water / vegetation and atmosphere; surface flow; groundwater flow. The advantage of this approach is that group sizes can remain fairly large without compromising homogeneity. Consequently, there will be increased confidence in the regression equations derived for each group. The second difference is in the methodology for testing which of the basin characteristics, among the ones that could potentially be used as classifiers, are the most appropriate. Having partitioned sub-basins into their major components, it is no longer possible to identify the best classifiers based on river flow gauging data alone. Therefore, an alternative scheme has been derived that relies on the output of detailed, physically based, distributed models.

## **1.3 Summary**

The ability to classify river basins into relatively homogeneous groups is central to the successful application of hydrological models in areas where hydrological data are sparse. It allows regression relationships to be derived for inferring model parameters in ungauged river basins where no data exist for calibration. This research proposes a novel approach to river basin classification. It is anticipated that this new approach will significantly improve the reliability of regression equations used to infer model parameters in ungauged catchments.

The development of a strategy for deriving this new classification scheme forms a major part of the PhD thesis. It is detailed in Chapter 3 and also in a peer reviewed British Hydrological Society (BHS) occasional paper (Reungoat and Sloan, 2002).

In this thesis, the focus is placed on the derivation of the ‘surface flow-component’ classification and a methodology for objectively selecting the variables that best characterise surface flow through the river network is developed. Chapters 4, 5 and 6 describe the tools developed in order to carry out the study. The results are presented and analysed in Chapters 7 and 8.

## *Chapter 2*

### **2 Literature review: River basin classification**

As outlined in the introduction, the ultimate aim of this research is to devise a river basin classification method specifically designed for regionalisation of hydrological model parameters. There have only been a few studies, mainly conducted during the last two decades, with this same aim (e.g.: Burn and Boorman, 1993; Post and Jakeman, 1996; Abdulla and Lettenmaier, 1997; Fernandez et al, 2000; Merz and Bloschl, 2004). However, the classification of river basins has a long history in the context of regional flood frequency analysis, and most of the studies aimed at hydrological model parameter estimation have borrowed the classification techniques first exposed in this context. Therefore, this chapter starts with a review of the classification techniques used for regional flood frequency analysis, prior to embarking on a critical review of how those classification techniques have been employed for hydrological model parameter estimation.

#### **2.1 Regional flood frequency analysis**

The aim of flood frequency analysis is to estimate the magnitude of extreme high flow events that have a low probability of occurring, on the basis of historic flow data. The procedures used for frequency analysis of flood data recorded at a gauged river basin are well established. Statistics of the measured flood data are used to fit a continuous probability distribution, which describes the probability that a flow event of any given magnitude is exceeded. It is then assumed that the distribution is capable of modelling the exceedance probability of floods outwith the range of magnitudes in the original flood data sample. As a consequence, estimates of the magnitude of events with a low probability of being exceeded can

be extrapolated. The reliability of these estimates is largely dependent on the size of the flood data sample. It is generally held that the magnitude of an event that has the probability  $1/T$  of being exceeded, in any year ( $T$  is known as the return period), can only be reliably estimated from a data record of  $n$  years length if  $n$  is greater than  $T$ . Flow records are rarely longer than 50 years. Therefore, in typical flood studies where the magnitude of floods with a return period of 100 years or more is required, but the flow record is only of 50 year duration, the sample size is not sufficient to ensure reliable estimations.

One approach, originally developed to overcome this problem of short data records at a gauged river basin, is called 'regional flood frequency analysis'. It relies on pooling data samples from river basins that are deemed to have similar flood frequency distribution to the river basin of interest, then conducting a traditional frequency analysis on the pooled data. This larger data sample decreases the uncertainty in the flood frequency estimates. The river basins that contribute to this pooled data set are said to belong to a 'homogeneous region' (NERC, 1975). However, the increased reliability is gained at the expense of having to identify data samples that have a similar frequency distribution. In other words, the first step in regional frequency analysis is to identify homogeneous regions.

Despite the fact that ideally, for flood frequency analysis, homogeneous regions were defined according to similarities in flood statistics, initial attempts at identifying such regions paid little attention to flood frequency data. Fairly arbitrary methods were used to group river basins based on, for example, administrative boundaries (NERC, 1975). The assumption was that because river basins within these groups were physically close to one another, they would have similar frequency distributions. It rapidly became apparent that this assumption was flawed (Hosking et al, 1985) and researchers sought alternative methods. In all of these methods, the similarity in flood frequency distributions was defined on the basis of arbitrarily chosen flood statistics. The methods differ in the choice of statistics (or group of statistics) and subsequent homogeneity / heterogeneity tests used. Reed et al (1995) proposed a classification of 703 UK river basins grouping them according to date-based indices defined by Peak Over Threshold (POT)

series of flood and rainfall regimes. The Flood Estimation Handbook (NERC, 1999) lays out the regionalisation procedures, for extending flood data samples that are currently considered to be best practice in the UK. These are based on methods proposed by Hosking and Wallis (1997) in which the L-moments of flood frequency distributions are used to determine homogeneity. This method of pooling data to extend records is now the industry standard for practising hydrologists.

## **2.2 Flood statistics in ungauged river basins**

The desire to improve the reliability of flood frequency statistics at gauged river basins is still one of the main motivations for research in regionalisation. However, the general ideas and some of the methods first exposed in this context have been borrowed by researchers interested in estimating flood statistics in ungauged river basins where no flood data are available. In this case, a grouping method that defines regions according to similarities in flood statistics cannot be used. The alternative preferred by most researchers is to use a grouping method, such as cluster analysis, that defines regions based on similarities in the physical characteristics of river basins instead of flood statistics (e.g. Acreman and Sinclair, 1986). There are a number of difficulties associated with this approach. Two in particular have received considerable attention from researchers. First, cluster analysis conducted using the available physical characteristics might generate regions homogeneous in terms of physical characteristics but not in terms of flood frequency distribution. Second, the available physical characteristics might be too numerous or too highly correlated to successfully apply a cluster analysis. In order to overcome these problems, procedures allowing the selection of the most important variables associated with the flood statistic of interest have been included in the grouping methods. Multiple regression and principal component analysis are typical examples of such procedures. The methodology presented by Nathan and McMahon (1990) to classify 184 catchments located in southeastern Australia uses multiple regressions to select and weight the most appropriate variables, cluster analysis to derive preliminary groupings and finally

multi-dimensional plotting to minimise group heterogeneity. Other multivariate statistical analysis including canonical variates and canonical correlation analysis have been tested to identify the variables that contribute most to the association between flood statistics and basin characteristics (Bates et al, 1998).

The 'region of influence' (ROI) approach, suggested by Acreman and Wiltshire (1987) and subsequently implemented by Burn (1990a,b), is another grouping approach used to estimate flood statistics at ungauged river basins: each river basin has a potentially unique set of associated gauged river basins, which constitutes its ROI. When dealing with ungauged river basins, catchment characteristics are used as the attributes that define their similarity to the available gauged river basins (Zrinji and Burn, 1994). Gauged river basins are successively added to the ROI of the ungauged river basin, starting with the most similar. This process terminates when including an additional gauged river basin leads to a lack of homogeneity within the ROI. The homogeneity test, which controls the termination of the process, uses flood statistics like, for example, the sample L-moment ratios proposed by Hosking and Wallis (1997).

### **2.3 Estimating hydrological model parameters for ungauged river basins**

The previous two sections have been concerned with estimating flood statistics for river basins: section 2.1 dealt with improving reliable estimation of these in gauged river basins, by pooling flood data samples; section 2.2 dealt with their estimation in ungauged river basins. However, these flood frequency analysis type models are not suitable when prediction of more than a single flood statistic is needed, as is the case when assessing most water resources and flooding issues. For example, in many flooding problems the transient behaviour of high water levels in a vulnerable area such as a low lying settlement is required. This is usually achieved by implementing a comprehensive hydraulic model in a restricted reach of the river in the immediate locale of the settlement. Traditionally the boundary condition for flow at the upstream end is predicted using a flood

frequency analysis. However, this approach only produces a single extreme flow value rather than a hydrograph, thus, in one had to assume the river flow was in steady state. Clearly this is a gross simplification of the passage of a flood wave. Therefore there is an increasing demand for hydrological models that can simulate the entire behaviour of a river basin and produce hydrographs. In engineering faculties in UK universities a distinction has evolved in which the detailed simulation of in-stream hydrological processes in restricted reaches is called 'hydraulic modelling' and the prediction of hydrographs is termed 'hydrological modelling'. This restriction is not apparent in other disciplines, such as geography or environmental sciences, or in other countries (Hornberger et al, 1998). The introduction to this thesis outlined the unpracticability of 'distributed' physically based modelling approaches, which partition the catchment into small control volumes. 'Lumped' models are an alternative pragmatic approach: they partition the region into sub-basins and simulate spatially averaged hydrological variables in each sub-basin. Therefore, 'lumped' models offer many advantages. However, they still present difficulties related to the fact that there is no guarantee that the hydrological process equations shown to hold at the small scale of the control volume will hold at large sub-basin scales. Moreover, although the parameters have physical meaning, they do not directly represent physical properties of the sub-basin. It has indeed been shown that they are not just an average of the small-scale parameters (Beven, 1995). Identification of these parameters is commonly obtained by calibration of the model against river flow data. This leads to the difficulties, highlighted in the introduction to this thesis, in applying such models to ungauged river basins where no flow data are available. As a consequence, the estimation of hydrological model parameters, rather than flood statistics, for prediction in ungauged basins, has become a significant research topic during the last decade.

The methods proposed in the literature for estimation of hydrological model parameters in ungauged basins borrow heavily on those discussed in section 2.2. The main difference is that, in gauged river basins where data exist, a hydrologic model is applied and calibrated. The set of calibrated model parameters then replaces the flood statistics. This essentially means that the grouping method has to be able to create regions homogeneous according to the river basin's entire

hydrological behaviour, which includes surface flow response but also other responses like groundwater flow and evapotranspiration. Some of the problems associated with grouping basins for estimating flood statistics are made worse as a consequence of this. For example, it was pointed out that there is no guarantee that similarity in terms of river basin characteristics implies similarity in terms of flood statistics, there is even less guarantee that similarity of river basin characteristics implies similarity of all aspects of the river basin's hydrological behaviour. In order to overcome this difficulty, particular attention has been placed on the initial selection of the most appropriate river basin characteristics. With all the published methods, an initial set of characteristics that can be used as classifiers is selected subjectively (e.g. Post and Jakeman, 1996). Then, some objective technique is used to reduce the size of the set. These techniques tend to fall into two categories: those that attempt to discard variables that are deemed to be poor classifiers, such as stepwise regression; and those that attempt to combine variables, such as principal component analysis (Nathan and MacMahon, 1990).

However, published regionalisation studies based on such classification methods have had limited success. Studies such as Abdulla and Lettenmaier (1997) and Sloan (1999) report significant errors in regional modelling studies of the same area in central USA, which they both attribute to the fact that regionalisation of model parameters had been conducted within heterogeneous groups of sub-basins. This difficulty arises from the fact that seeking groups of river basins that are similar in their entire hydrological cycle leads to a large number of groups, each of which contains only a few basins. It is therefore difficult to have confidence in the regression equations derived for inferring model parameters. An alternative approach, that has been suggested by some researchers, is to relax the degree of homogeneity in the classification stage, then create groups that contain a larger number of basins but are less homogeneous. Again, it is difficult to have confidence in the regression equations derived from such groups since they have to explain more of the variability in model parameter values using basin characteristics.

## 2.4 Other methods used for river basin classification

Two recent approaches used for catchment grouping can be highlighted:

Considering that river basin grouping could be regarded as an example of the wider problem of classification of datasets, Hall and Minns (1999) applied artificial neural networks and fuzzy set techniques to flood and catchment characteristics datasets for Wales and the Southwest of England. Artificial neural networks (ANNs) are either able to 'learn' the relationship between a set of inputs and outputs (supervised learning process) or to 'recognise' patterns in input data (unsupervised learning process). In fuzzy classification, each site is allowed a degree of membership in more than one group. The study showed that combinations of catchment characteristics were a more logical basis for regionalisation than geographical proximity.

Recently, Wolock et al (2003) used Geographical Information System (GIS) tools combined with principal component and cluster analysis to delineate 'hydrologic setting regions' in the US. Their grouping method is based on the concept that the basic building block of all landscapes (the 'hydrological landscape unit') includes an upland adjacent to a lowland separated by an intervening steeper slope. The attributes used to describe the hydrologic landscapes include 4 attributes relevant to the land surface form, 2 relevant to the geologic texture and 2 relevant to the climate. These 8 variables were subjected to a correlation analysis. Then, a cluster analysis using principal component results was conducted. The results showed that the hydrologic setting regions were effective in dividing the country into distinct areas as measured by differences in the landscape form, geologic texture and climate characteristics. However, the homogeneity of these regions in term of hydrological behaviour was not demonstrated.

## 2.5 Summary

The classification of river basins has a long history, which has its roots in regional flood frequency analysis. The classification methods developed in this context have then been adapted to be applicable to the regionalisation of hydrological model parameters. In addition, some new classification methods have also been developed. However, published regionalisation studies based on such classification methods have had limited success. Two major unresolved difficulties have been identified in this literature review; the first consists in obtaining groups of basins that are homogeneous enough and contain a number of basins large enough for regression techniques to be reliably applicable; the second is in the selection of the variables that are the most appropriate classifiers.

## *Chapter 3*

### **3 Framework**

The introduction and subsequent literature review highlighted the need for a pragmatic method for classifying river basins specifically designed for regionalisation of hydrological model parameters. In this chapter, the novel concepts on which the proposed classification scheme relies are presented and a method for deriving the new scheme is developed. Devising the philosophy outlined below has been a significant part of the research in this thesis and has been presented in a publication (Reungoat and Sloan, 2002). In addition, a poster outlining the framework was awarded 'Outstanding Student Paper' at the American Geophysical Union fall meeting (2002) in San Francisco.

#### **3.1 A new scheme for classifying river basins - concepts**

The literature review identified two major difficulties when attempting to classify river basins into hydrologically homogeneous groups; the first consists in obtaining groups of basins that are homogeneous enough and contain a sufficiently large number of basins for regression techniques to be reliably applied; the second is in the selection of the variables that are the most appropriate classifiers. The proposed river basin classification scheme differs from previously used schemes (e.g. Burn and Boorman, 1993; Post and Jakeman, 1996; Abdulla and Lettenmaier, 1997; Fernandez et al, 2000; Merz and Blöschl, 2004) in two significant ways, each attempting to address one of the two identified difficulties. The first, is to discard the constraint that each basin only belongs to one group, and propose that each basin undergoes three independent classifications, one relative to each of the major components of the land phase hydrological cycle: interaction of soil water / vegetation and atmosphere; surface flow; and

groundwater flow. The advantage of this approach is that ‘component-related’ groups can contain a fairly large number of basins without compromising their homogeneity. Consequently, there will be increased confidence in the regression equations derived for each group. The second difference is in the methodology used for objectively testing which of the basin characteristics, from a set of potential classifiers, are the most appropriate. The set of classifiers is constrained by excluding climatic characteristics, which are used as hydrological model input data. It is further constrained by the fact that ‘component-related’ classifications are conducted, so that less characteristics are of importance in describing the hydrology of one river basin component (e.g. surface water flow), than the hydrology of the basin as a whole. However, there still remain a large number of potential classifiers and it is necessary to objectively test the relative importance of each of them. Such a test would normally require measured hydrological data. Herein lies the major difficulty and, consequently, the major research issue in partitioning river basins into component parts for classification. It is almost impossible to measure hydrological variables that can characterise the components (e.g. surface water) in isolation. However, sufficiently comprehensive hydraulic and hydrological models describing individually the river basin components do exist (e.g. open channel flow routing models) to allow synthetic data to be generated. Therefore, synthetic data are used to test the relative importance of the various pre-selected classifiers. The proposed objective test consists of simulations whose results are analysed using standard statistical multivariate data analysis methods.

### **3.2 Deriving the new classification scheme - method**

The proposed method includes classification schemes for the three major components described by the standard hydrological models. Each ‘component-related’ classification will lead to the estimation of the hydrological model parameters for that component alone. Thus the three ‘component-related’ classifications have to be derived for all the model parameters to be estimated.

This thesis focuses on the derivation of the ‘surface flow-related’ classification which will be used for estimating the model parameters relevant to the surface flow component. In particular, the thesis will concentrate on flow through river networks. A practical method for objectively identifying the variables that characterise this flow is developed. This method involves three stages: first, is the generation of a large set of river network models suitable for flow routing simulation; second, is the derivation of synthetic flow data, using a flow routing model (sometimes called hydraulic model) to simulate the hydrological response at the outlet of each river network to a pulse of runoff distributed uniformly throughout the network; third, a statistical multivariate analysis of the synthetic data derived by simulation is conducted in order to identify the variables describing the network which best characterise the network hydrological response.

### 3.2.1 Generation of river network data

The application of common network flow routing models requires network structure and conveyance data. Network structure data are increasingly becoming available on digital databases. However, determining the conveyance of a network normally requires detailed surveys of river cross-sections. Collecting such data in the field is a difficult and expensive task. The construction, in such a way, of a large set of river network models is therefore unfeasible. To overcome this problem, synthetic model parameters are sampled from distributions of physically reasonable values. However, the aim is to keep the synthetic data to a minimum. Therefore, only the river model parameters that determine conveyance are sampled from statistical distributions. These distributions are constrained by using the rules of hydraulic geometry for stable channels. The synthetic conveyance data comprise channel top width and Manning’s  $n$ . Determining approximate values for top width involves using the Leopold and Maddock power law (1953),

$$W = a Q^b, \quad (3.1)$$

where  $W$  is the top width,  $Q$  is the bankfull flow and  $a$  and  $b$  are constants. This is an empirical formula and, in practice, the value of  $a$  varies within bounds for UK rivers (Hey, 1982). Therefore, rather than defining only one possible cross-sectional geometry for each reach, a range of possible cross-sections are derived.

Similarly Manning's  $n$  values are selected. Therefore for each reach of the network, a range of possible conveyance data is derived. All the other parameters necessary for the application of the flow routing model are derived from real network structure information, taken from detailed maps.

Digital terrain and river maps of Scotland have been imported into a GIS database. A suite of programs automatically generates all the topographic information on the river network upstream of a point of known coordinates. These programs first delineate the river basin draining through the selected point, then isolate its associated river network, and finally export a table containing a description of the nodes and reaches necessary for the flow routing. The table of network structure information is augmented by synthetic conveyance data for each reach of the network. Therefore, for each river network an ensemble of tables is generated, each one corresponding to a different set of 'possible' conveyance characteristics. Each flow routing simulation is repeated for each 'possible' network model, thus multiple physically reasonable realisations of each network are simulated.

### **3.2.2 Flow routing simulation**

The success of the method is partly dependent on the choice of the flow routing model for simulating the response of a network to a pulse of runoff. This model must be capable of dealing realistically with spatially distributed information of physical properties, and have a proven track record of being able to accurately reproduce a network outlet flow when appropriately parameterised. In this thesis, Scottish rivers are analysed which are dominated by upland river networks, where channels are steep. Therefore, the simulations are conducted using a kinematic wave model. In order to have complete flexibility in the automation of the simulations, a new model is developed. This model numerically solves the kinematic wave equation including lateral inflow. The runoff into each reach of the network is assumed to be directly proportional to the area of the catchment draining into it, the input rainfall hydrograph and a runoff ratio. The same input rainfall hydrograph is used for all simulations. This hyetograph corresponds to a design extreme rainfall event, generated using the FEH storm design procedures

(NERC, 1999). In constructing such a runoff hydrograph, the aim is not to reproduce any one particular storm event but rather to have a hydrograph with standardised shape and typical flow rates with which to compare the hydrological responses at the outlet of many river networks.

### **3.2.3 Multivariate statistical analysis**

Ultimately, the aim is to identify variables describing the physical properties of the catchment network, which best characterise the hydrological response of the network. This is achieved conducting a multivariate statistical analysis. The first step in this analysis consists in applying standard clustering techniques to identify clusters within the network hydrological responses. Each network hydrological response is summarised by two simple variables measuring the peak flow attenuation and time to peak lag of the flood wave at the outlet of the network. The second step of the analysis consists in identifying the network descriptors that produce the clusters that best match the network response clusters. This is achieved by applying discriminant analysis to the clustered networks using variables describing the physical properties of these networks. Network descriptor variables traditionally and more recently used in regionalisation studies as well as original variables are investigated. A schematic diagram of the method is given in Figure 3.1.

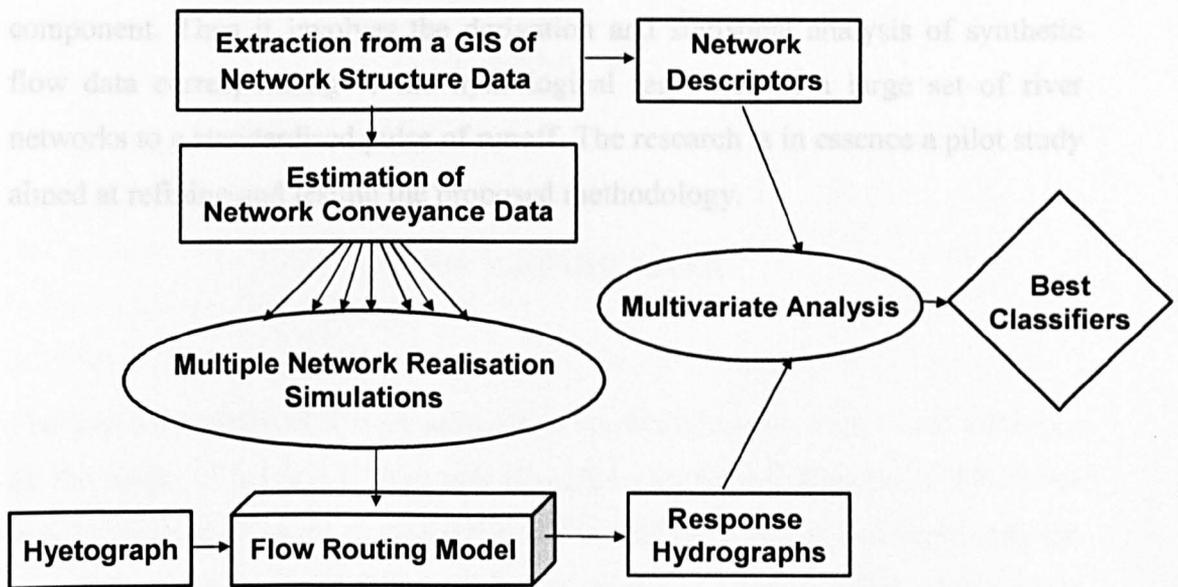


Figure 3.1 Schematic diagram of the best classifiers objective test

### 3.3 Summary

This research proposes a novel approach to river basin classification specially designed for regionalisation of hydrological model parameters. This approach differs from other currently available methods in two significant ways. The first is to discard the constraint that each basin only belongs to one group, and propose that each basin undergo three independent classifications, one relative to each of the major components of the land phase hydrological cycle: interaction of soil water / vegetation and atmosphere; surface flow; groundwater flow. The second difference is in the methodology for objectively selecting which of the basin characteristics are the most appropriate classifiers. Synthetic data are used to test the relative importance of the various pre-selected classifiers. The proposed objective test consists of simulations whose results are analysed using standard statistical multivariate analysis techniques.

Although the research outlined above describes an approach to identifying the most appropriate classifiers and classification schemes for three major components of the hydrological cycle, this thesis focuses on the surface flow

component. Thus it involves the derivation and statistical analysis of synthetic flow data corresponding to the hydrological responses of a large set of river networks to a standardised pulse of runoff. The research is in essence a pilot study aimed at refining and testing the proposed methodology.

## *Chapter 4*

### **4 Flow routing model**

The geomorphology of a river network is known to exert a significant influence on the shape of a flood hydrograph (Rodrigue-Iturbe and Rinaldo, 1997). Thus ideally, every catchment hydrology model would have an explicit representation of the river network and flow within it (Ewen and Parkin, 1996). However in practice, this rarely occurs. One might expect that of all the components of the hydrological cycle open channel flow best lends itself to a fully distributed modelling approach. Much of the data required to parameterise such a model, such as river cross-section and along stream profiles, are visible and accessible via remote sensing, unlike for example ground water flow where characterising the properties of an unseen aquifer is fraught with practical and conceptual difficulties (Fetter, 1994). Nonetheless, most hydrological models adhere to a parsimonious description of the effects of channel routing in line with their representation of the other hydrological components (Beven and Kirkby, 1979). The reason for this is that despite the fact that with a lot of money and labour one can build up a fairly comprehensive picture of channel form, the derived flow model will always require calibration. Therefore why bother with detailed description when a calibrated simple mathematical function will do the job equally well? This argument works well where long time series of hydrological data exist. However, as outlined in the previous chapter, it leaves hydrologists with a dilemma when it comes to ungauged basins. Two options then present themselves for investigation: fully distributed modelling or the approach advocated in this thesis of characterising the physical properties of the network that exert the strongest influence on the hydrograph, such that model parameters might be transposed from gauged to ungauged basins. The reason for adopting the latter approach is purely pragmatic: calibrating a model on a hydrologically similar network and translating it to the ungauged basin of interest is significantly quicker than

collecting all the data required to run a fully distributed model. Furthermore, the additional computational expense of the fully distributed approach seems unwarranted if one is ultimately no more confident in the resulting hydrograph simulation than from a prudently simplified lumped model. This then leads to the research problem, which forms the bulk of this thesis, of how best to characterise a river network. This has been an active research field for more than 70 years. Engineers such as Horton (1932, 1945), Strahler (1952) and Shreve (1966) attempted to classify networks according to the way they bifurcate as early as 1932. Much of the research that followed has been driven by characterising the topography and topology of the network. During the latter part of the 20<sup>th</sup> century, several researchers began to try to incorporate key features of the river network into hydrological models, most notably with the Geomorphologic Unit Hydrograph (Rodriguez-Iturbe and Valdes, 1979) and the network width function (Mesa and Mifflin, 1988). While these have been shown to reproduce observed hydrographs when suitably calibrated, it is still unclear whether they yield any understanding of how a river responds to a flood wave. For example, if the network width function for two rivers is similar, will the hydrographs in response to the same rainfall be similar? The problem is that the models represent the whole hydrological cycle and it is impossible to uncouple the hillslope and network responses.

This difficulty in separating out the components of the hydrological cycle has very recently led researchers down a theoretical route when attempting to characterise flow through river networks. During the course of this PhD research, several eminent researchers have also started advocating similar approaches to the one adopted here, that is to use a distributed routing model to simulate flow through the network and then attempt to characterise the features of the network that exert the strongest influence on synthetic hydrograph shape. Much of the research has been abstract in the sense that it doesn't use real river networks (Gupta et al, 1996; Veitzer and Gupta, 2001; Menabde et al, 2001). Furthermore, the distributed models used are extremely simple, for example assuming a constant fixed velocity in every reach (Veitzer and Gupta, 2001; Menabde et al, 2001). Nonetheless, this research is beginning to highlight some general patterns in network responses, for example in the way flood peaks scale with area in networks with different

bifurcation ratios. However, the simplicity of their modelling and the abstract nature of their river networks leave an air of doubt hanging over how their results will translate to real river networks.

In this thesis, an attempt is made to keep the networks and the representation of flows as true to life as possible. Therefore a more standard approach to modelling the hydraulics of flows through a network is adopted. There is a wide variety of commercial software for modelling flow in river reaches. These tend to be aimed at simulating water levels to predict flood extent. These were investigated for use in this project, but all were found to be unsuitable for one or more of the following reasons:

- Here the interest is in simulating hydrographs, which precludes models built on steady state Bernoulli energy conservation equation (e.g. HECRAS).
- One dimension depth averaged hydraulic models based on a solution to the Saint Venant equations (e.g. ISIS) all adopt the hydrostatic assumption which supposes that pressure vertically downward in the water column varies as hydrostatic pressure. This assumption does not hold in steep channels and the numerical solutions to the underlying equations become unstable (Abbott, 1978). Furthermore, none of the commercial packages give access to the source code.
- All commercial codes incorporate a graphical user interface. This makes automated data input impossible. In this thesis, flow through very many large river networks is simulated. Therefore some sort of automated data entry is essential.

In the light of these problems with commercial packages, it became apparent that a new distributed flow routing model was required.

## 4.1 Channel flow routing

### 4.1.1 Governing equations: the kinematic wave model

All transient flow routing models are based on solving equations for the continuity of mass and momentum (Abbott, 1978). The forces that are incorporated into the momentum equation are a function of the channel form. When channels are steep, the two dominant forces are gravity and friction and it is usually assumed that the local acceleration, convective acceleration and pressure terms can be neglected in the momentum balance equation. Equation (4.1) is the Saint Venant continuity equation and equation (4.2) is obtained by simplification of the Saint Venant momentum equation where local acceleration, convective acceleration and pressure terms are neglected. Wind shear and eddy losses are also neglected. These assumptions are warranted when channels are steep. The flow is then described by the kinematic wave model. Chow et al (1988) assumes that the kinematic wave model is suitable for describing downstream wave propagation when channel slope is greater than 0.01%. Li et al (1975) showed that the kinematic wave model successfully simulates hydrographs on slopes as little as 0.1%. The network data used in this thesis for flow simulations correspond to upland Scottish catchments where the large majority of channels are steep. Indeed, a GIS analysis showed that 94% of channels had slopes greater than 0.1%. Therefore the kinematic wave model is appropriate in most of Scotland's rivers and the routing model developed here solves the kinematic wave equation. Clearly, there are flat channels in Scotland, where pressure forces will propagate the flood wave rather than gravity and a more complete momentum equation is warranted. However the aim here is to investigate the characteristics of network shape that influence flood routing. Therefore, rather than significantly complicating the routing model by solving the full momentum equation, the channel slope is altered such that the wave can be propagated by gravity. The small errors introduced by doing this in a few reaches of the network are swamped by uncertainties in some of the key model parameters such as Manning's  $n$ . Furthermore, our kinematic model is significantly more comprehensive than models used by other researchers in characterising network responses.

The kinematic wave equation is given by:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q \quad (4.1)$$

$$S_0 = S_f \quad (4.2)$$

where  $x$  is the distance along the channel,  $t$  the simulation time,  $Q(x,t)$  the flow in the channel,  $A(x,t)$  the channel cross section area,  $q(x,t)$  the lateral inflow per unit of channel length,  $S_0$  the channel bottom slope and  $S_f$  the friction slope.

It is generally accepted that Manning's equation can be used to relate flow to cross-section area,

$$Q = \frac{\sqrt{S_f}}{nP^{2/3}} A^{5/3}, \quad (4.3)$$

where  $n$  is Manning's roughness coefficient and  $P$  the wetted perimeter.

Equation 4.3 can be expressed in a more generic form as,

$$A = \alpha Q^\beta, \quad (4.4)$$

where

$$\alpha = \left( \frac{nP^{2/3}}{\sqrt{S_f}} \right)^{3/5} \quad (4.5)$$

and

$$\beta = 3/5.$$

Assuming that the channel depth is much smaller than its width, then the value of the wetted perimeter,  $P$ , can be approximated by the value of the channel width,  $W$ . The friction and bedslope are equivalent (equation 4.2), therefore,

$$\alpha \approx \left( \frac{nW^{2/3}}{\sqrt{S_0}} \right)^{3/5}. \quad (4.6)$$

$\alpha$  is assumed to be constant along a reach of one channel.

Both the steep and wide channel assumptions have been made by Li et al (1975) when successfully validating the kinematic wave model against measured runoff data at a catchment scale. Furthermore, if mean depth and width hydraulic geometry relationships shown to hold for a wide range of UK river reaches (Hey and Thorne, 1986) are used to calculate a channel mean depth to width ratio, bankfull flows of for example 2 and 50 cubic metres give a depth to width ratio of 0.09 and 0.06 respectively, validating the wide channel assumption.

Equations 4.1 and 4.4 can be solved analytically to describe flow along uniform channels for a limited set of boundary conditions and inflow rates. In practice, where channel geometries change, or for large river networks, it is necessary to solve these equations numerically.

#### 4.1.2 Numerical solution to the kinematic wave equations

A wide range of numerical solutions to the kinematic wave equation have been proposed (Franco and Chaudhry, 1998). Their relative merits reflect the intended use of the model and tend to be a trade off between accuracy and stability. The network model developed here is intended for use within an automated system, where multiple simulations of very many networks will be conducted. Therefore, stability is an imperative. An unstable model has the potential to produce spurious results, which unless immediately obvious, could bias any statistical analysis. Therefore, an unconditionally stable scheme was implemented. It is based on a paper by Li et al (1975). The stability criteria imposed may result in an unacceptable loss of accuracy. The accuracy of the scheme is tested by comparing the numerical scheme with an analytical solution to equations 4.1 and 4.4.

The scheme is initially implemented for a single channel of length  $L$  with a time and space varying lateral inflow  $q(x,t)$ , where  $x$  is distance down the channel and  $t$

is time. It is assumed that flow at the upstream boundary is known for all time and that the initial flow along the length of the channel is also known, so that the system is defined by,

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q,$$

$$A = \alpha Q^\beta,$$

$$Q(x,0) = Q_{ini}(x), \quad 0 \leq x \leq L,$$

$$Q(0,t) = Q_{up}(t), \quad \forall t. \quad (4.7)$$

The channel is divided into  $n$  sections of equal length  $\Delta x$  and the simulation time into  $n$  time steps of equal duration  $\Delta t$ .

Using the following backward finite-difference approximations

$$\frac{\partial Q}{\partial x} \approx \frac{Q_i^j - Q_{i-1}^j}{\Delta x},$$

$$\frac{\partial A}{\partial t} \approx \frac{A_i^j - A_i^{j-1}}{\Delta t},$$

$$q \approx \frac{q_i^j + q_i^{j-1}}{2}, \quad (4.8)$$

where

$$Q_i^j \approx Q(i \cdot \Delta x, j \cdot \Delta t),$$

$$q_i^j \approx q(i \cdot \Delta x, j \cdot \Delta t)$$

and

$$A_i^j \approx A(i \cdot \Delta x, j \cdot \Delta t),$$

a finite-difference approximation of equation 4.1 is,

$$\frac{Q_i^j - Q_{i-1}^j}{\Delta x} + \frac{A_i^j - A_i^{j-1}}{\Delta t} = \frac{q_i^j + q_i^{j-1}}{2}. \quad (4.9)$$

Using equation 4.4,

$$A_i^j = \alpha(Q_i^j)^\beta \quad (4.10)$$

and

$$A_i^{j-1} = \alpha(Q_i^{j-1})^\beta. \quad (4.11)$$

Substituting these into equation 4.9 and rearranging gives,

$$\frac{\Delta t}{\Delta x} Q_i^j + \alpha(Q_i^j)^\beta = \frac{\Delta t}{\Delta x} Q_{i-1}^j + \alpha(Q_{i-1}^j)^\beta + \Delta t \left( \frac{q_i^j + q_i^{j-1}}{2} \right). \quad (4.12)$$

To solve this we note that if, as is the case, the solution propagates forward in time and space, then the only unknown in equation 4.12 is  $Q_i^j$  and the equation can be re-written as follows,

$$\frac{\Delta t}{\Delta x} Q_i^j + \alpha(Q_i^j)^\beta - C = 0, \quad (4.13)$$

where

$$C = \frac{\Delta t}{\Delta x} Q_{i-1}^j + \alpha(Q_{i-1}^j)^\beta + \Delta t \left( \frac{q_i^j + q_i^{j-1}}{2} \right).$$

This non-linear equation is solved using Newton's iterative method. This offers a generic numerical solution to a non-linear equation of the form,

$$F(Y) = 0. \quad (4.14)$$

It can be shown (Press et al, 1993) that,  $Y_k$ , given by

$$Y_{k+1} = Y_k - \frac{F(Y_k)}{F'(Y_k)}, \quad (4.15)$$

where  $F'$  is the first derivative of  $F$ , converges on a solution of equation 4.14 as  $k$  tends to infinity. Thus, starting with an initial guess,  $Y_0$ , iterations are continued until there is no significant change in  $Y_k$ .

In the flow routing model,

$$F(Y) = \frac{\Delta t}{\Delta x} Y + \alpha Y^\beta - C. \quad (4.16)$$

The stability and convergence of the finite-difference scheme leading to equation 4.12 are discussed in the following section.

### 4.1.3 Stability and convergence of the numerical scheme

Proof of the unconditional stability of the finite-difference scheme (equation 4.12) has been given by Li et al (1975).

To test the convergence of the numerical solution, it is compared with a true solution, which has been derived analytically for a particular set of initial and boundary conditions. The numerical solution developed in the previous section is intended for use within a river network model. Therefore each reach of the network will receive flow at its upstream boundary and a lateral inflow. Despite the wide range of analytical solutions to the kinematic wave equation, none could be found for these boundary conditions. The closest model was by Streeter (1966), who derived an analytical solution for the flow caused by rain falling on a sloping plane. This solution has been modified for channels below.

### *Analytical solution to the kinematic wave*

An analytical solution to the kinematic wave equation has been developed for a river reach of length  $L$  with uniform cross-section. It has a continuous constant baseflow,  $Q_0$ , entering at its upstream boundary and experiences a pulse of lateral inflow representing runoff from the catchment area (Figure 4.1).

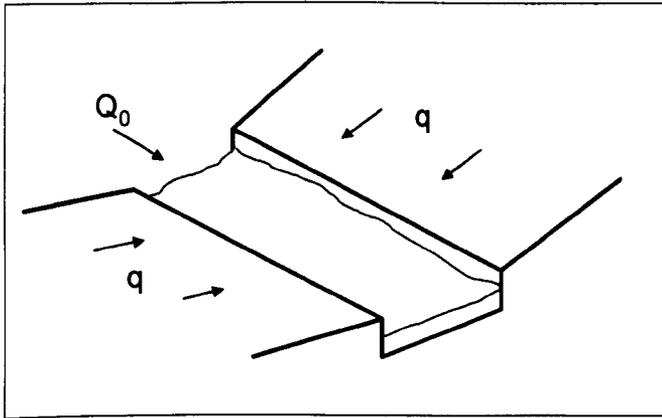


Figure 4.1 Schematic diagram of channel

The pulse starts at a time  $t_0$  with a magnitude  $q_0$ , and drops to zero at a time  $t_1$ . The time  $t_1$  has been selected to be large enough such that the flow reaches steady-state at some time,  $t_s$ , before dropping back down to the baseflow,  $Q_0$  (Figure 4.2).

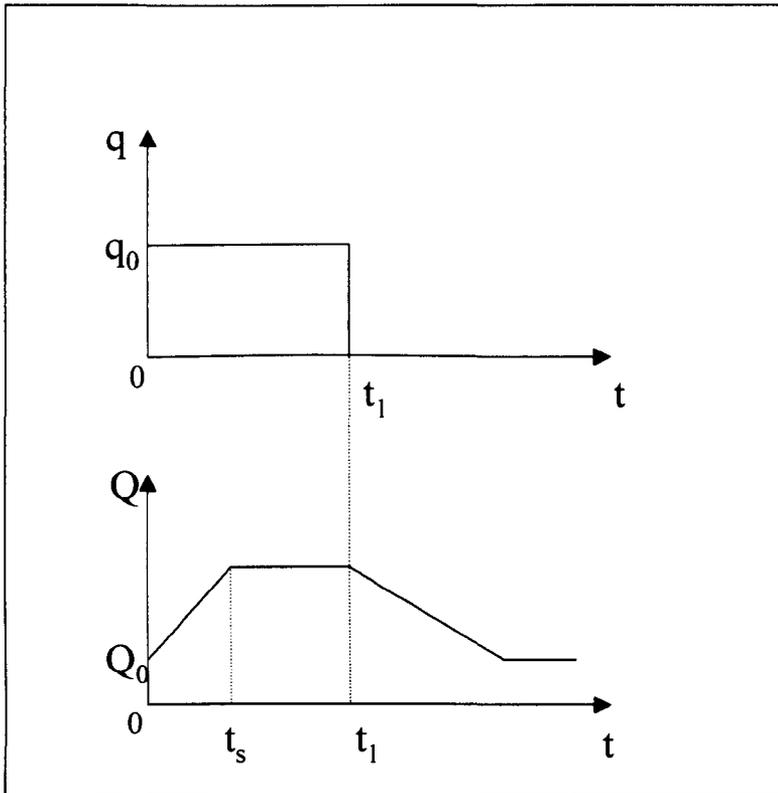


Figure 4.2 The three-part analytical solution

Formally, the model is defined by,

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q,$$

$$A = \alpha Q^\beta,$$

$$Q(x,0) = Q_0, \quad 0 \leq x \leq L,$$

$$Q(0,t) = Q_0, \quad \forall t,$$

$$q(x,t) = \begin{cases} q_0, & 0 \leq t \leq t_1, \\ 0, & t > t_1 \end{cases}. \quad (4.17)$$

The solution is constructed in three parts. The first component describes the flow from the start of the pulse to the time  $t_s$  when the system reaches steady state. The second part is trivial and describes the steady state conditions between  $t_s$  and  $t_1$ , when the pulse drops. The third part describes the recession in flow beyond  $t_1$  after the lateral inflow ceases (Figure 4.2).

The analytical solution was derived using the characteristic equations, which allow the partial differential equation 4.1 to be converted into a system of first-order ordinary differential equations.

First, equation 4.4 is re-arranged to give flow rate as a function of area,

$$Q = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}} A^{\frac{1}{\beta}}. \quad (4.18)$$

For convenience, let  $a = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}}$  and  $b = \frac{1}{\beta}$ , so that,

$$Q = aA^b. \quad (4.19)$$

Substituting this into equation 4.1 and differentiating gives,

$$abA^{b-1} \frac{\partial A}{\partial x} + \frac{\partial A}{\partial t} = q, \quad (4.20)$$

which is the kinematic wave equation expressed in terms of cross-sectional area.

Now consider the total (or material) derivative of  $A$ ,

$$dA = \frac{\partial A}{\partial x} dx + \frac{\partial A}{\partial t} dt, \quad (4.21)$$

or, on dividing through by  $dt$ ,

$$\frac{dx}{dt} \frac{\partial A}{\partial x} + \frac{\partial A}{\partial t} = \frac{dA}{dt}. \quad (4.22)$$

We now note that the material derivative, which holds for any description of flow, and the kinematic wave equation take very similar forms.

Equating the coefficients of the partial derivatives in 4.20 and 4.22 gives,

$$\frac{dA}{dt} = q \quad (4.23)$$

and

$$\frac{dx}{dt} = abA^{b-1}. \quad (4.24)$$

The system of equations 4.23 and 4.24 constitutes the characteristic form of the kinematic wave equation. This pair of ordinary differential equations implicitly incorporates the momentum and mass continuity equations. They can be interpreted as inferring that if you are an observer travelling with the flow at a rate  $abA^{b-1}$ , you will see the cross-section area increase at a rate  $q$ . The only difference between this and standard derivations of the characteristic equations (Chow et al 1988) is that they have been developed in terms of cross-sectional area. The following specific solution, given the boundary and initial conditions (system of equations 4.17), is non-standard.

### ***First part of the solution: rising limb***

This part holds until steady state is reached at an, as yet, unknown time  $t_s$ . First, equation 4.23 is integrated from  $t = 0$ , giving

$$A(x, t) = A_0 + qt \quad (4.25)$$

where  $A_0 = A(x, t = 0)$ . Using Manning's equation,  $A_0$  is obtained from the initial flow,

$$A_0 = \alpha Q_0^\beta. \quad (4.26)$$

Equation 4.25 holds for any specific characteristic. Now substituting 4.25 into 4.24 gives,

$$\frac{dx}{dt} = ab(A_0 + qt)^{b-1}. \quad (4.27)$$

Integrating this gives,

$$x(t) = x_0 + \frac{a}{q} \left[ (A_0 + qt)^b - A_0^b \right], \quad (4.28)$$

where  $x_0$  is the starting point of the characteristic at time  $t = 0$ . Again, this holds for any specific characteristic. Now let's consider one characteristic in particular, starting at the most upstream point in the channel,  $x_0 = 0$ . We note that at steady state,  $Q(x) = Q_0 + qx$ . Substituting  $A$  (equation 4.25) into 4.28 and re-arranging gives,

$$qx + aA_0^b = aA(x, t)^b \quad (4.29)$$

or

$$qx + Q_0 = Q(x). \quad (4.30)$$

Therefore, the area on the characteristic described by 4.20 represents the area at steady state all the way along the channel. Now the time,  $t_s$ , to steady state is simply given by equation 4.28 with  $x = L$  and  $x_0 = 0$ ,

$$t_s = \frac{1}{q} \left[ \left( A_0^b + \frac{qL}{a} \right)^{\frac{1}{b}} - A_0 \right]. \quad (4.31)$$

The flow anywhere in the channel is given by,

$$Q(x) = a(A_0 + qt)^b \quad (4.32)$$

until it reaches steady state. Therefore the flow at the outlet,  $x = L$ , is,

$$Q(x = L) = a(A_0 + qt)^b, \quad t < t < t_s. \quad (4.33)$$

***Second part of the solution: steady state***

The second part of the solution is trivial in that the flow remains at steady state until time  $t_1$  when the pulse of lateral inflow  $q$  drops to zero,

$$Q = Q_s = a(A_0 + qt_s)^b, \quad t_s \leq t \leq t_1. \quad (4.34)$$

***Third part of the solution: falling limb***

For the falling limb of the hydrograph, we return to equations 4.23 and 4.24, which define the characteristic. However, in this case, equation 4.23 becomes,

$$\frac{dA}{dt} = 0. \quad (4.35)$$

Therefore the cross-sectional area along a characteristic is constant,

$$A(x, t) = A(x). \quad (4.36)$$

Substituting this into 4.28 gives,

$$x(t) = x_1 + abA_1(x)^{b-1}(t - t_1), \quad (4.37)$$

where

$$x_1 = x(t = t_1).$$

Now let's consider the starting condition of being at steady state. From the first part of the solution (rising limb), we know that at  $t = t_s$ ,

$$A(x, t = t_s) = A_1 = \frac{a}{q} (A_0 + qt_s)^b. \quad (4.38)$$

We also know (equation 4.28) that

$$x = \frac{a}{q} [(A_0 + qt)^b - A_0^b]. \quad (4.39)$$

Therefore, substituting 4.38 and 4.39 into 4.27 gives,

$$x_1 = \frac{a}{q} (A_1^b - A_0^b). \quad (4.40)$$

Therefore,

$$x = \frac{a}{q} A_1^b + \frac{a}{q} A_0^b + abA_1^{b-1} (t - t_1). \quad (4.41)$$

This determines the cross-sectional area at a distance  $x$  down the channel at time  $t$ . To get an equation for flow we note,

$$A_1 = \left( \frac{Q}{a} \right)^{\frac{1}{b}}, \quad (4.42)$$

and therefore,

$$x = \frac{Q}{q} + ab \left( \frac{Q}{a} \right)^{\frac{b-1}{b}} (t - t_1) - \frac{a}{q} A_0^b. \quad (4.43)$$

At the outlet  $x = L$  and therefore the discharge is described by the non-linear equation,

$$0 = L - \frac{Q}{q} + ab \left( \frac{Q}{a} \right)^{\frac{b-1}{b}} (t - t_1) - \frac{a}{q} A_0^b. \quad (4.44)$$

This can be solved for any  $t$  using Newton's method.

This completes the analytic solution.

### ***Comparison between numerical and analytical solutions***

Both the analytical and numerical solutions were implemented in FORTRAN codes, which take inputs of lateral inflow and upstream flow hydrographs, and return the channel outlet hydrograph.

The results of both solutions are compared for the same boundary conditions, and the sensitivity of the numerical scheme to space and time discretization is tested. The comparison is performed for a 10 kilometre long and 4 metre wide channel reach, which has a continuous constant baseflow of 2.7 cubic metres entering at its upstream boundary. The channel experiences a pulse of lateral inflow representing runoff from an area of 800 square kilometres. It starts at a time  $t_0 = 0$  with a magnitude corresponding to a rainfall of 40 millimetres per hour of intensity, and drops to zero after one hour and 15 minutes. Figure 4.3 shows the analytical results and the numerical results obtained for various values of  $\Delta x$  and  $\Delta t$ .

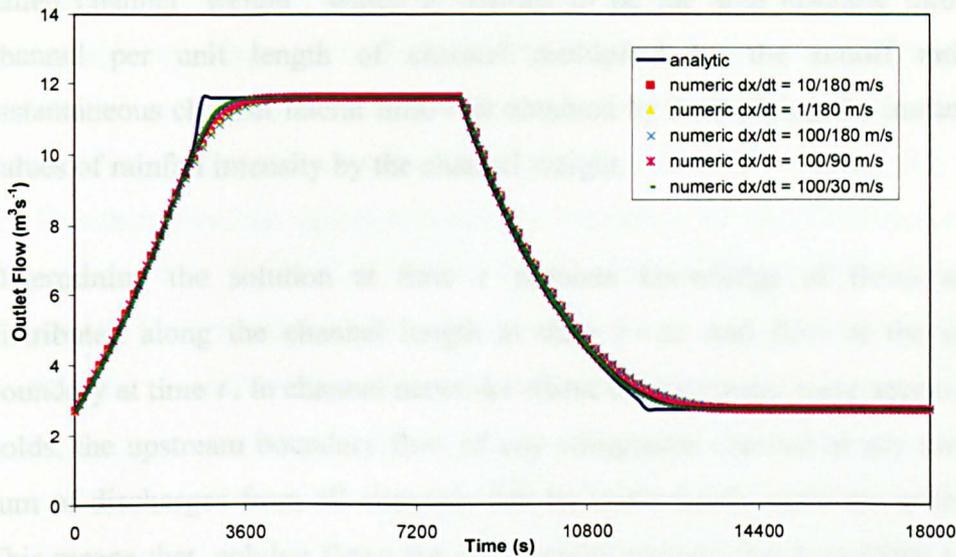


Figure 4.3 Comparison of analytical and numerical results

The analytical and numerical solutions are well matched. The numerical solution is insensitive to the values of  $dx$  and  $dt$  within the range of values that is feasible to use in a large network. Clearly if  $dx/dt$  was selected such that it was much larger than the kinematic wave celerity, we might anticipate errors, but this is easily avoided. The sharp change from increasing to steady flow that is apparent in the analytic solution is not reproduced by the numerical model. This can be attributed to numerical diffusion (Li et al, 1975) and is a price one must pay for having unconditional stability. Since the model is to be used to perform automated simulations of large networks, it was felt that this was a relatively small price that was worth paying. Furthermore, the error is transient and short lived.

## 4.2 From channel routing to network routing

The numerical solution developed in the previous section simulates flows at nodes distributed along the length of a single channel with uniform cross-section, given the channel topography (slope and length), conveyance characteristics (top width and roughness coefficient) and the lateral inflow per unit length of channel. The lateral flow into each reach of the network is assumed to be uniform along the length of the reach, and directly proportional to both the area of the catchment draining into it and a runoff ratio. This information is combined in a variable

called channel 'weight', which is defined to be the area draining through the channel per unit length of channel multiplied by the runoff ratio. The instantaneous channel lateral inflow is obtained by multiplying the instantaneous values of rainfall intensity by the channel weight.

Determining the solution at time  $t$  requires knowledge of flows at nodes distributed along the channel length at time  $t - \Delta t$  and flow at the upstream boundary at time  $t$ . In channel networks where the kinematic wave approximation holds, the upstream boundary flow of any component channel at any time is the sum of discharges from all channels that lie immediately upstream at that time. This means that, solving flows for any channel requires the knowledge of which channels, if any, lie immediately upstream of that channel. To facilitate this, each channel of the network is assigned a unique identifier and associated with the list of channels corresponding to its immediate upstream channels. At the extremities of the network, this list is empty. Moreover, when calculating flows throughout the network, it is necessary to begin by solving for flows at the upstream extremities of the network and move systematically towards the outlet. To facilitate this, the channels constituting the network are ordered according to their relative position to the outlet. The ordering system is illustrated in Figure 4.4: once flows have been solved for all channels of  $n^{\text{th}}$  order, flows can be solved for all channels of  $(n-1)^{\text{th}}$  order.

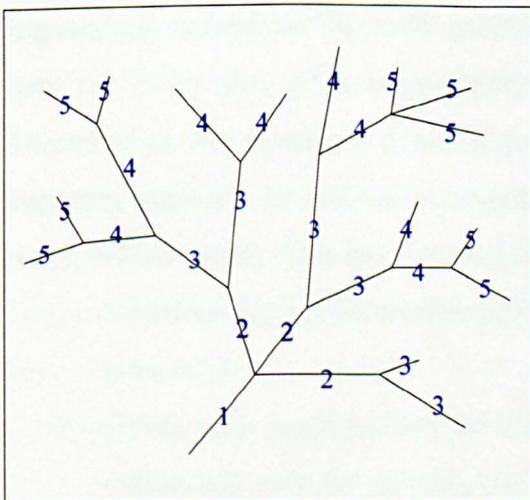


Figure 4.4 Network channels ordering system

A FORTRAN code was developed that takes details of the network and rainfall hyetograph and returns the hydrograph at the outlet of the network. It begins with the reaches at the extremities (level 5 in Figure 4.4) of the network. Given the flow at the previous time step ( $t - \Delta t$ ), it predicts current flow along each of these ( $t$ ). This then gives the upstream boundary conditions for reaches at the next level down (level 4 in Figure 4.4). In this way, water is successively routed through decreasing level reaches. When the flow has been simulated at time  $t$  for all reaches, the process is repeated to predict flow at time  $t + \Delta t$ .

### 4.3 Summary

A model has been developed that is suitable for flow routing through river networks where channels are steep enough for back water effects to be negligible. The model implements a numerical solution to the kinematic wave equations including lateral inflow and propagates the solution from the upstream extremities of the network towards the outlet. The finite-difference scheme used for solving the equations is unconditionally stable and allows for convergence with results derived by solving the equations analytically for a hypothetical case. The model, implemented in a FORTRAN code, takes inputs of network data and rainfall hyetograph and returns the hydrograph at the outlet of the network.

## *Chapter 5*

### **5 Network Structure Data**

The ultimate goal of this thesis is to determine a suite of variables that describe a network and can be used to classify the hydraulic response of river networks to runoff events. It has already been argued in Chapter 3 that the only practical way to characterise the network response in isolation from all the other catchment hydrological processes that produce a real hydrograph is using a comprehensive flow routing model. To this end, a kinematic wave routing model for large networks was developed. Whilst recognising that all the parameters required cannot be measured, in this thesis the aim is to make the river model as true to reality as possible. This is in contrast to many of the recent studies on network responses (Gupta et al, 1996; Veitzer and Gupta, 2001; Menabde et al, 2001). Therefore where possible, data from real river networks are used. Furthermore, if significant conclusions are to be drawn on network response and the influence of network structure, then many networks need to be simulated. The logistics of parameterising even a single large network on the basis of field surveys make it impractical. Indeed, no literature could be found where bed slopes and conveyance data for every link of a large dendritic network were derived from surveys. Therefore in this research, it was imperative that the best remotely sensed data currently available be utilised in an automated manner to parameterise models of many real networks. This has been achieved by:

- 1- Constructing a GIS database of the physical structure for Scottish river networks.
- 2- Developing an algorithm for automated extraction, from the database, of catchment network topological data characterising the network structure suitable for use in the flow routing model.
- 3- Generating physically reasonable distributions of conveyance parameters for each reach in the network, from which model realisations of the

conveyance parameters are selected. This later point is described in the following chapter.

## **5.1 Derivation of a GIS database for the physical structure of Scottish rivers**

The most comprehensive surveys of the UK have been conducted by the UK Ordnance Survey (OS). They produced detailed terrain and feature maps. During the 1990s, the Centre for Ecology and Hydrology (CEH) based at Wallingford recognised the potential for converting these data into a digital form that might be used in hydrological studies to determine flow pathways and to delineate catchments. These tasks used to be some of the most time consuming and error prone in a hydrological study. After many years of work, they produced digital maps of terrain and rivers for the UK (Morris and Flavin, 1990). Terrain data are raster and comprise the average elevation (accurate to 0.1m) in each cell of a regular 50m x 50m grid covering the UK. Such data are conventionally called a Digital Elevation Model (DTM). CEH have analysed this DTM to provide two additional data sets that potentially describe aspects of the hydrology. Firstly, a raster map of the direction in which water running across any cell in the grid would be likely to leave. This is based on the direction of greatest slope. Secondly, what is known as a flow accumulation map. This records the number of upstream cells that would drain through each cell on the grid if the terrain were covered in a film of water. This map is derived by analysing the flow directions. The river data are vector and comprise latitude and longitude data  $(x,y)$  for nodes that represent the start and end of reaches in the network, and  $(x,y)$  data for series of intermediate points, known as vertices, that define the shape of the reach. These data sets were imported into the Geographical Information System software ArcInfo.

To facilitate the automated extraction of river network data for the network flow routing model, aspects of these data sets have been combined. Namely, the elevation, flow direction and flow accumulation at each node on the river network

have been extracted from the raster maps and stored digitally along with the vectors that describe the rivers. ArcInfo allows a table of additional data to be associated with the nodes of each reach (node Attribute Table). In this table, each node and reach are assigned unique identifiers and the length of a reach corresponds to the sum of the Euclidean distances between successive vertices.

One might have expected that using ArcInfo GIS to interrogate the raster maps and find values at nodes in the river network would be a straightforward task. In principal, it is. However, it transpired that the standard Arc commands for achieving this did not work for interrogating the flow accumulation and flow direction data. This is because they assume that the raster data represents a continuous smoothly changing variable like elevation. Therefore, in interrogating the maps to find a value of the variable at a specific location, they take a distance weighted average of the values in neighbouring cells (Figure 5.1a). However, flow accumulation is a discrete variable and can change significantly between neighbouring cells. For example flow accumulation can drop dramatically in moving from a cell on the main channel, where flow accumulation will be high, to one on the neighbouring hillslope, where flow accumulation will be very much lower (Figure 5.1b). A distance weighted average in this case gives a very distorted representation of the flow accumulation at a node.

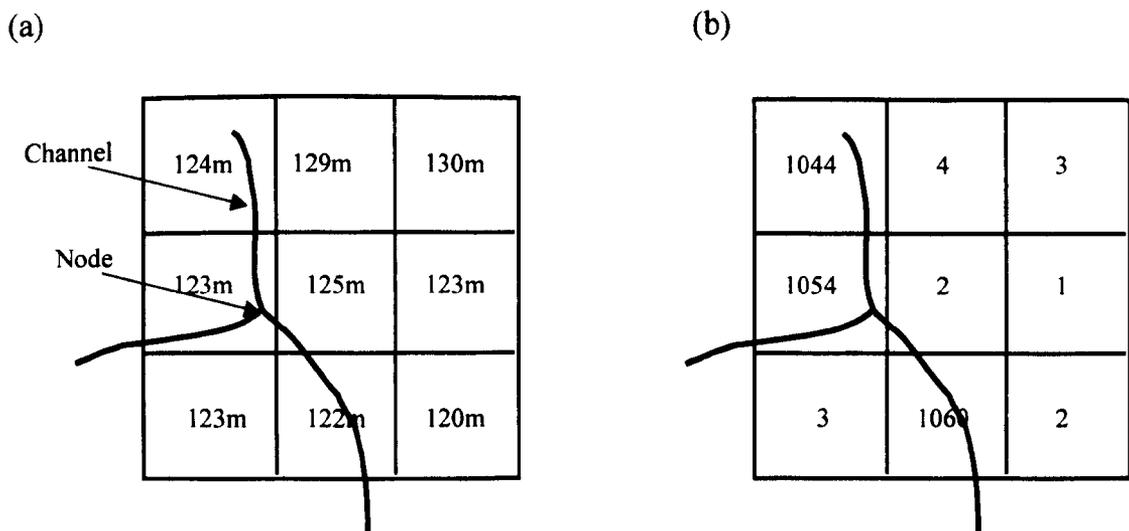


Figure 5.1 Raster representations of terrain elevation and flow accumulation variables: (a) Elevation; (b) Flow accumulation.

In addition to this problem with the standard ArcInfo commands, the software failed because the digital maps for the whole of Scotland were too large to hold simultaneously in the memory of a PC, even with 2GB RAM. Therefore, a computer program was written in the low level Arc Macro Language (AML) and FORTRAN to:

- Partition the raster and vector maps into smaller more manageable units.
- Interrogate the river vector map to find the location of nodes.
- Identify the cells on the raster maps that coincide with the river nodes.
- Extract a single value of flow accumulation corresponding to the cell that the node sits in.
- Take a weighted average of the elevation in the four closest cells to the node location.
- Add the elevation and flow accumulation data to the node attribute table.
- Reconstruct the database for the whole of Scotland.

Thus a comprehensive digital database for all the rivers in Scotland was constructed. It contains all the appropriate topological characteristics for every reach. However, despite all the nodes in the database being derived from Ordnance Survey data, it transpired that the river data and the DTM were inconsistent. Therefore, small but frequent corrections were required at the stage of deriving individual catchment networks for use in the flow routing model. Highlighting these inconsistencies, which is done later in this chapter, is extremely important. The CEH DTM, flow accumulation, flow direction and river network data is the industry standard dataset. It forms the basis of many decision support systems. Whilst it is marketed as being ‘hydrologically consistent’, it is not and this makes the automated extraction of a network a far less than straightforward procedure.

## **5.2 Automated extraction of catchment network data**

To implement the techniques for characterising the response of networks that is advocated throughout this thesis, it is necessary to extract the physical characteristics of river networks from a large number of catchments. This requires interrogating the raster maps and then the digital river database to first delineate catchments and then extract the networks they enclose.

Catchment, or watershed, delineation is a straightforward procedure in the ArcInfo GIS, provided that 'sinks' have been removed from the DTM. Sinks are cells on the terrain model that are surrounded, in all directions, by cells with higher elevations. Thus any flow entering the cell has no route by which to leave. Such sinks are normally 'filled' for hydrological analysis by increasing the elevation in such a way that there is a continuous flow pathway from any cell to the boundary of the DTM, in this case, the sea. CEH market all their DTMs as being 'hydrologically corrected' and therefore, sinks had already been removed. The delineation procedure comprises simply specifying the outlet of the catchment. The GIS then uses the flow direction raster map to trace back upstream of the outlet, to all cells that could potentially drain through it. The catchment is then stored as a binary raster map; one value for inside the catchment and another for outside.

### **5.2.1 Catchment network extraction**

Ideally, having derived the binary map for a catchment, the digital database could be interrogated in a straightforward manner to extract all nodes that fall within the catchment boundary. However, it was only after attempting this and inspecting the derived river networks in detail that the inconsistencies between the various data sets became apparent. Intermittently within a network, the downstream nodes of reaches were observed to have higher elevations than upstream nodes. Similarly, downstream nodes were often observed to have lower flow accumulations (less area draining through them) than upstream nodes, which is physically impossible. The inconsistencies arise because CEH have developed their own algorithm for generating the flow direction map. The DTM sold by CEH is as true as possible to

the original OS terrain data, with sinks removed. Using the GIS software, it is possible to generate flow directions from each cell based solely on the local slope. On comparing this to the flow accumulation map marketed by CEH, it transpired that there were many discrepancies. Furthermore, comparing it to the digitised rivers again there were significant differences. This is because basing flow direction solely on the local slope on a 50m x 50m grid neglects the role of micro topography on the path a river follows. Morris and Flavin (1990) recognised this and generated flow direction maps using additional information. They returned to the original digitised OS contour data and, in the proximity of rivers, identified the concave contour structure that would indicate a valley. During the interpolation procedure used to generate a DTM, such features are often smoothed out. Therefore, using the raw contour data, they were able to effectively mould the DTM so that the flow pathways as best as possible followed the river network. Thus the marketed flow direction and flow accumulation maps correspond not to the marketed DTM but rather to a modified DTM, which is not available. This fact is not advertised and is important to note for practitioners.

With regard to this research, it simply meant that the CEH flow direction and accumulation maps should be used. However, despite the good agreement between the digitised rivers and the CEH flow direction and accumulation maps, inconsistencies still arose. For example, with the flow direction indicating rivers running parallel but one cell removed (50m away) from the digitised river. The variety of anomalies encountered was such that, despite best efforts, no algorithm could be generated to automatically correct them. Therefore an alternative way of extracting local catchment river networks was sought.

### **5.2.2 Generating 'hydrologically consistent' network data based on digital terrain data**

Tarboton et al (1991) suggest an approach to generating network data that is consistent with an underlying DTM. It works on the assumption that the greater the area that can potentially drain through a point on a map the more likely there is to be a river at the point. Thus, their suggestion was that the flow accumulation map derived from a DTM could be used to approximate the path of watercourses,

in particular by setting a threshold in accumulation above which cells are retained and classified as rivers, and below which cells are discarded and assigned to hillslopes. The resulting raster map then comprises lines of cells representing rivers that can be converted into vector format. This method has been widely used (Naden et al, 1999, Moussa, 2003) to insure that the network is consistent with the underlying DTM and the flow direction and accumulation maps. However, clearly one runs the risk of departing from the true paths of rivers, should, for example, the DTM have too coarse a resolution. Therefore, in adopting such an approach here, it is implicitly assumed that Morris and Flavin's (1990) algorithm has been effective in pulling the flow pathways close to those of real rivers. This will subsequently be tested by comparing derived and digitised river networks. Using the Tarboton et al (1991) algorithm, networks were derived for many Scottish catchments. Figure 5.2 shows a comparison between the digitised river network and derived network in a single catchment using three different flow accumulation thresholds. It can be seen that there is very close agreement between the routes taken by the derived rivers and true river paths. However, Figure 5.2 also serves to illustrate a problem with Tarboton's algorithm in its original form. Even at the local scale of a catchment, no single threshold applies for all branches in the network. Some branches terminate at a source that coincides with the true source when the threshold is high (Figure 5.2a), but on the same map the extent of many of the other branches are under estimated. On the other hand with a very low threshold (Figure 5.2c) it is possible for the derived network to extend to the sources high in the headwaters but a high number of spurious reaches are introduced.

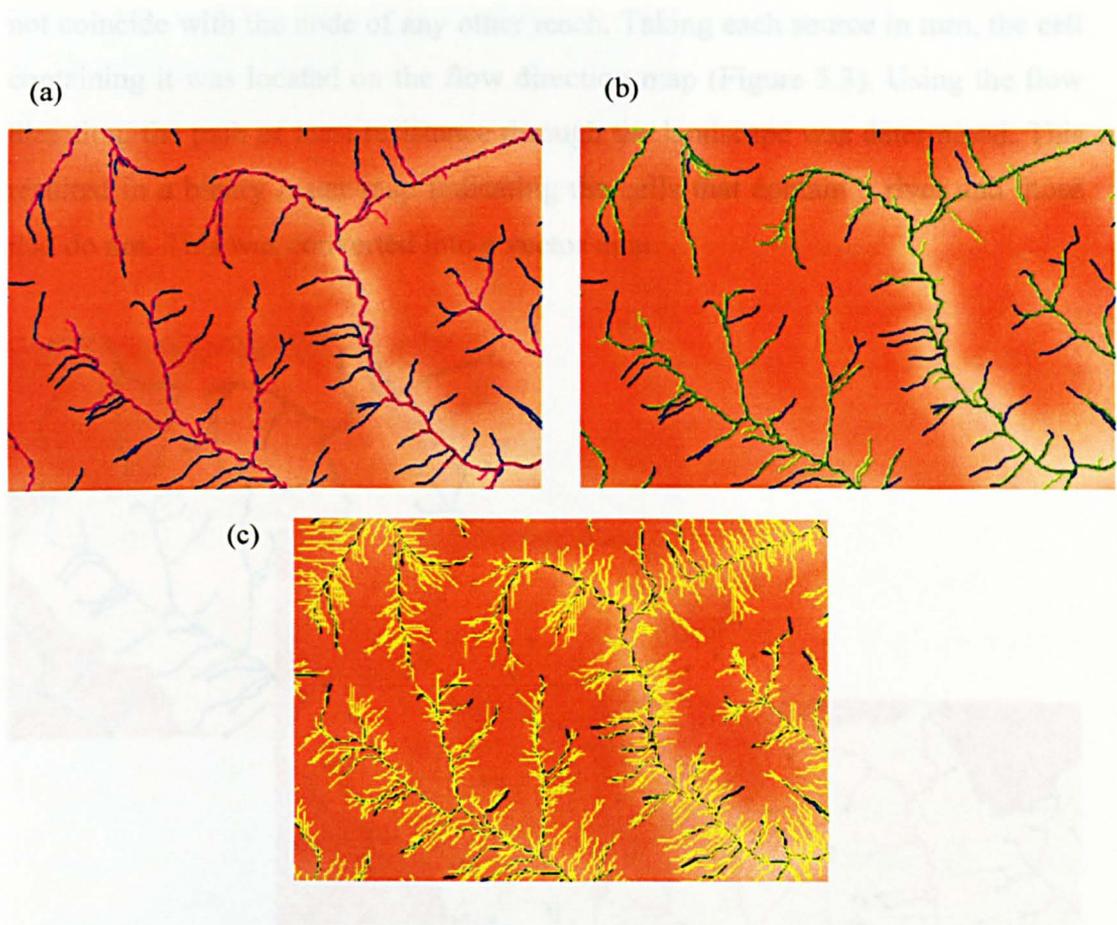


Figure 5.2 Digitised river network (blue lines) and DTM-generated flow paths derived for 3 decreasing flow accumulation threshold values: (a) Threshold  $0.250\text{km}^2$ ; (b) Threshold  $0.215\text{km}^2$ ; (c) Threshold  $0.025\text{km}^2$ .

The problem of selecting an appropriate threshold is compounded when deriving networks for many catchments automatically. The threshold is seen to vary significantly; this is probably as a result of land use and geological differences in the catchments. Thus, while the derived network paths were in agreement with the digitised paths (indicating that Morris and Flavin's flow accumulation map is a good representation), Tarboton's algorithm cannot reproduce the extent of networks. Therefore, an alternative that combined elements of the true digitised network and Tarboton's algorithm was developed.

Rather than use a threshold in flow accumulation to determine where the rivers rise (their sources), the location of sources was taken directly from the digitised river database for Scotland. Within this database, sources were easily identified as they correspond to the extremities of the network and therefore to nodes that do

not coincide with the node of any other reach. Taking each source in turn, the cell containing it was located on the flow direction map (Figure 5.3). Using the flow direction, the path of least resistance through the landscape was determined. This resulted in a binary raster map indicating the cells that contain a river and those that do not. This was converted into a vector map.

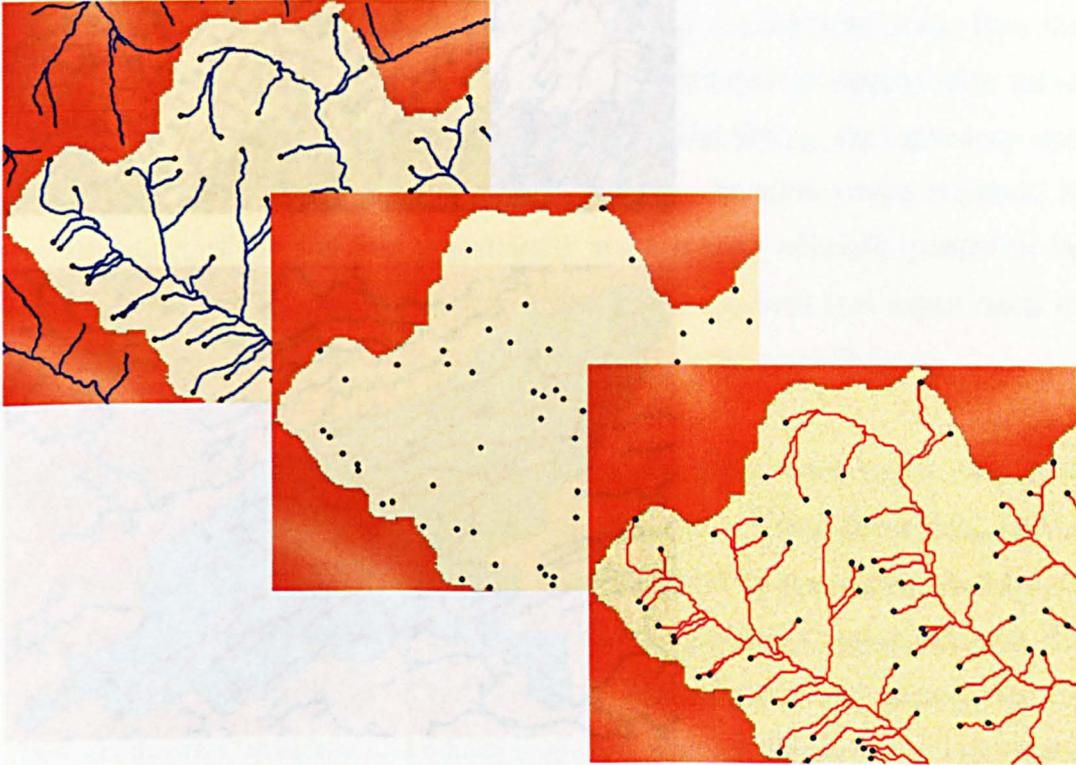


Figure 5.3 Watershed delineation, source-point localisation and drainage network derivation

Figure 5.4 shows the derived networks compared to the digitised networks for 3 catchments. The match is remarkable; there are a few minor anomalies.



Figure 5.4 Sampled examples of extracted DTM & sources - generated catchment drainage networks (red lines) compared with digitised rivers (blue lines)

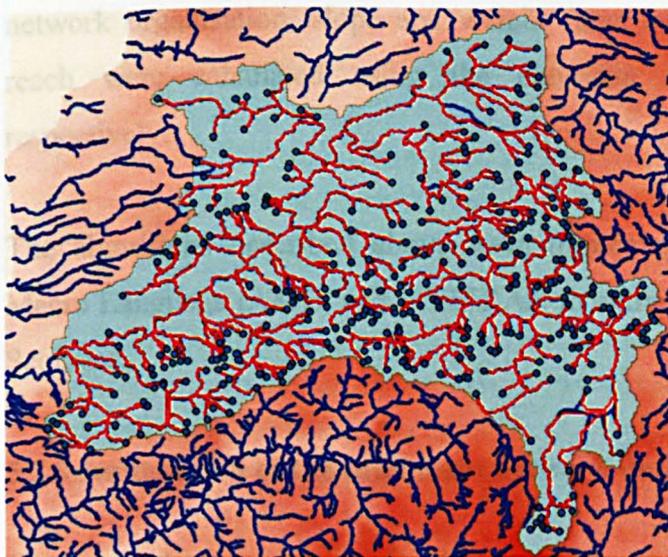
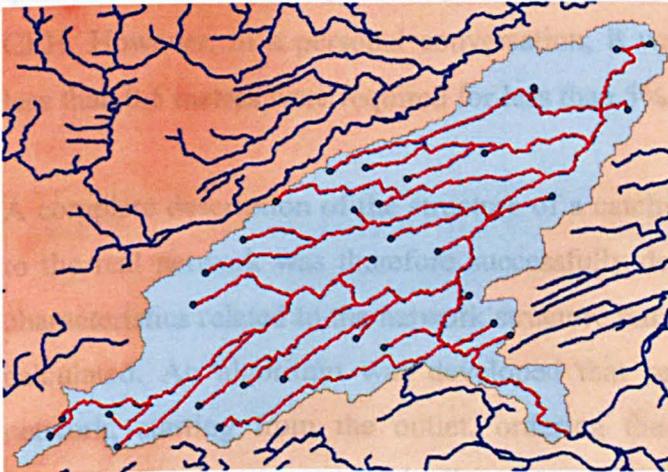
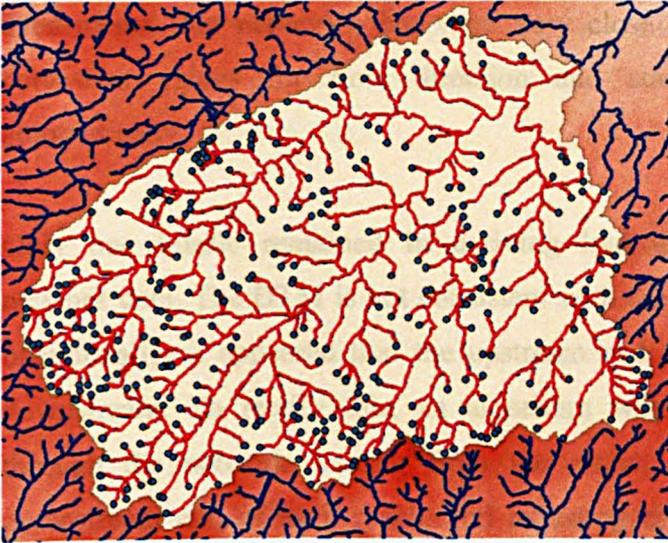


Figure 5.4 Sampled examples of extracted DTM & sources – generated catchment drainage networks (red lines) compared with digitised rivers (blue lines)

This meant that networks that were very close to the true networks and had consistent values for flow direction and accumulation could be derived automatically.

Only one hurdle remained in deriving hydrologically consistent catchment network data. The DTM is not consistent with the flow direction grid. Therefore, occasionally, it appeared that the upstream end of a reach was lower than the downstream. To rectify this, an algorithm was written to sequentially move through the network starting from the outlet, identifying the upstream and downstream end of each reach, and adjusting the elevation where required. A quantification of the required adjustments has not been officially published by CEH. However, in a personal conversation, it was reported that adjustments of less than 0.5 metres were required for less than 5% of the river channels.

A complete description of the structure of a catchment network that is very close to the real network was therefore successfully derived. From these data, all the characteristics related to the network structure required by the routing model were calculated. An algorithm was developed that sequentially moved through the network, starting from the outlet, ordering the channels and identifying the channels directly upstream and adjacent to each channel. Based on the interpreted network organisation, slope and weight (area contributing to runoff) for each reach were calculated using the elevation and flow accumulation data respectively.

The algorithms described above were implemented in a combination of Arc Macro Language (AML) and FORTRAN to extract many catchment networks for Scotland.

### **5.2.3 Automation of the catchment network data extraction**

The algorithm developed in the previous section automatically extracts a catchment network given the latitude and longitude of the outlet. A list of catchment outlet locations corresponding to the Scottish river gauging stations is available in the Flood Estimation Handbook (NERC, 1999). Based on this, a list

of  $x$  and  $y$  coordinates of gauging stations was created. The algorithm was enabled in a batch program that moved through the outlets list extracting all the corresponding network data.

Table 5.1 gives a sample of the network data file, which is extracted for each catchment. Appendix A gives a list of the catchment outlets for which network data were extracted. Estimates of catchment drainage area and median flood of the annual maximum series ( $Q_{med}$ ), as reported in the FEH (NERC, 1999), are also given, as well as network total length, as calculated by the GIS extraction algorithm. Figure 5.5 locates these catchments on the map of Scotland. In the case of nested catchments, for clarity of representation, only the hierarchically higher catchment is illustrated.

**channel ID,channel length,node ID,X,Y,Z,flow accumulation**

```

3,220.71068,1,239700.00000,938600.00000,520.600,41.000
3,220.71068,2,239750.00000,938400.00000,500.300,245.000
2,341.42136,3,239600.00000,938650.00000,550.100,18.000
2,341.42136,4,239750.00000,938400.00000,500.300,245.000
5,191.42136,5,239800.00000,938300.00000,500.000,19.000
5,191.42136,6,239650.00000,938200.00000,479.300,311.000
4,241.42136,7,239750.00000,938400.00000,500.300,245.000
4,241.42136,8,239650.00000,938200.00000,479.300,311.000
6,641.42136,9,237500.00000,938250.00000,473.100,70.000
6,641.42136,10,237600.00000,937650.00000,367.600,341.000
8,903.55339,11,238650.00000,937750.00000,647.400,11.000
8,903.55339,12,238050.00000,937500.00000,366.500,710.000
1,2210.66017,13,239100.00000,939000.00000,729.500,4.000
1,2210.66017,14,238050.00000,937500.00000,366.500,710.000
11,70.71068,15,238050.00000,937500.00000,366.500,710.000
11,70.71068,16,238000.00000,937450.00000,366.500,1393.000

```

Table 5.1 Sample of an extracted catchment network data file

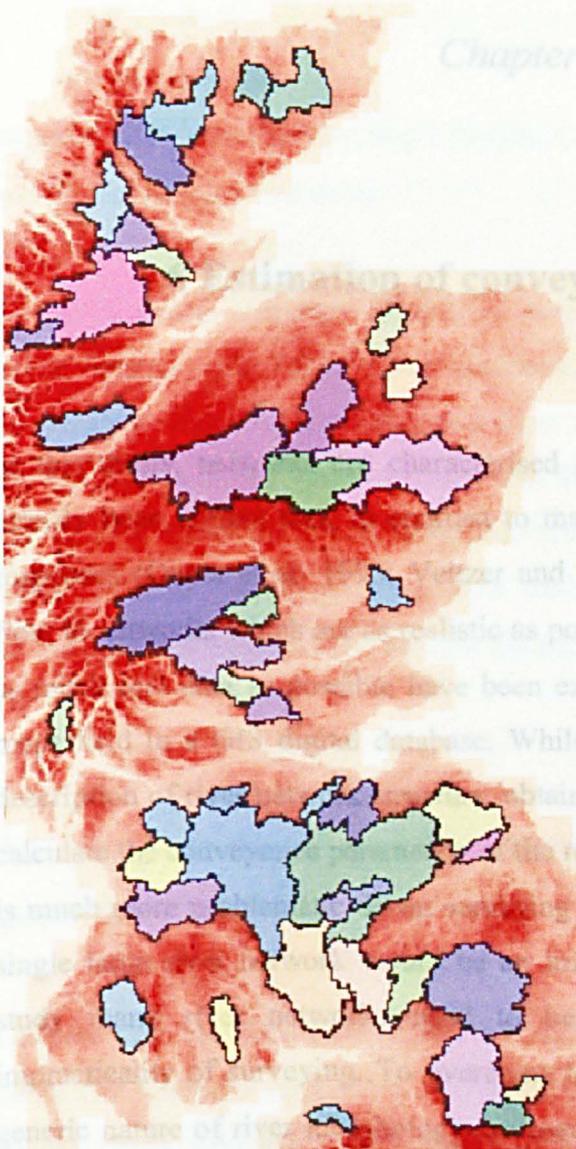


Figure 5.5 Catchment positions on the map of Scotland

### 5.3 Summary

A series of algorithms have been developed that perform automated extraction of catchment network data from a GIS database representing real Scottish river networks. Network data suitable for flow routing have been extracted for a large number of catchments. The extraction procedure uses an original method, which insures that the derived network structure data are very close to real networks and hydrologically consistent.

## Chapter 6

### 6 Estimation of conveyance parameters

In this thesis, networks are characterised using simulated hydrographs. It has already been argued that, in contrast to many of the recent studies on network responses (Gupta et al, 1996; Veitzer and Gupta, 2001; Menabde et al, 2001), flow in networks which are as realistic as possible should be simulated. Therefore as much real data as possible have been extracted from Ordnance Survey (OS) maps held in a GIS digital database. While OS maps provide a fairly accurate description of river network structure, obtaining cross-sectional data, necessary to calculate the conveyance parameters of the routing model, for large river networks is much more problematic. Even surveying one cross-section on each reach of a single large river network would be an impractical and expensive task. In this study many river networks need to be simulated, which compounds the impracticality of surveying. To overcome this problem, empirical results on the generic nature of river morphology are used. In particular, an empirical estimate of channel top width is required, since the routing model conveyance formula is given by,

$$A = \alpha Q^\beta, \quad (6.1)$$

where, based on Manning's equation applied assuming that channel wetted perimeter can be approximated by the value of the channel top width,

$$\alpha \approx \left( \frac{nW^{2/3}}{\sqrt{S_0}} \right)^{3/5} \quad (6.2)$$

and

$$\beta = 3/5 ,$$

where  $n$  is the channel Manning's roughness coefficient,  $W$  the channel top width and  $S_0$  the channel bed slope.

Although various empirical models for river cross-section geometry can be found in the literature (e.g. Leopold and Maddock, 1953; Hey and Thorne, 1986; Huang and Warner, 1995), these have never been tested against data for Scottish rivers. Therefore an analysis was conducted that involved collecting river cross-sections from a wide spread survey of Scottish rivers and seeking an empirical relationship between channel width and some flow statistic. This was the first time that such an analysis was made specifically for Scotland.

## 6.1 A brief review of generic patterns in river cross-section geometry

Leopold and Maddock (1953) first introduced the term 'Hydraulic geometry' to name the functions describing the variation of some streamflow characteristics (width, depth, velocity) in relation to discharge. They analysed current-meter field measurements, collected for the purpose of computing discharge, by district offices of the U.S. Geological Survey over a period of seventy years for rivers all over the United States. They observed that at a given cross-section, streamflow characteristics vary with discharge as simple power functions of the form,

$$W = aQ^b , \tag{6.3}$$

$$D = cQ^f , \tag{6.4}$$

$$V = kQ^m , \tag{6.5}$$

where  $W$  is the width of the water surface,  $D$  and  $V$  are the water mean depth and velocity across the cross-section,  $Q$  is the computed flow discharge and  $a, b, c, f, k$

and  $m$  are all constants. They called these power law relationships ‘at a station’ hydraulic geometry.

When examining the geometry of cross-sections at various gauging stations on the same river network, they noted similar power law relationships between  $W$ ;  $D$ ;  $V$  and flows of a particular return period. This they called ‘downstream’ hydraulic geometry.

Leopold and Maddock (1953) reported that average values of exponents  $b$ ,  $f$  and  $m$  are 0.26, 0.40 and 0.34 for at a station hydraulic geometry, whereas for downstream hydraulic geometry they are 0.5, 0.4 and 0.1 for a return period corresponding to bankfull flow.

They compared the downstream hydraulic geometry relationships they observed for river channels with ‘regime equations’ published by Lacey (1930, 1939) on stable irrigation canals. Given the similarity of both sets of equations, they concluded that the same factors, which tend to maintain equilibrium channel forms in stable reaches, are also acting in river reaches to produce consistent patterns in the relation among hydraulic variables, even though the rivers are not in equilibrium.

Nixon (1959) suggested that although rivers experience a range of discharges, they adjust to an effective single dominant discharge. This discharge is the constant flow rate that would develop the same shape and dimensions as the natural sequence of flows. Flume and field data suggest that this dominant flow rate is about bankfull discharge. Studies of sediment transport processes (Hey, 1975; Hey and Thorne, 1984) confirmed Nixon’s hypothesis. They showed that the frequency of flow, which transports most sediment in the long term, equates to the frequency of bankfull flow.

Hey and Thorne (1986) analysed data from 62 stable gravel-bed river reaches in England and Wales and related bankfull discharge to bankfull hydraulic geometry variables. Motivated by practical engineering design, they recommended the following equation,

$$W = aQ_h^{0.5}, \quad (6.6)$$

where  $a$  takes the value 4.33, 3.33, 2.73 or 2.34 depending on the type of vegetation and where the use of the 1.5-year return period flood is recommended if bankfull discharge is not available.

Huang and Warner (1995) analysed 529 observations from both stable canals and natural rivers in the USA and the UK in order to obtain general values of exponents for a quantitative multivariate downstream hydraulic geometry model incorporating the influence of bank strength. On the basis of this quantitative model, it is shown that downstream hydraulic geometry is determined not only by flow discharge, but also by channel slope, channel average roughness and sediment composition of the channel boundary. They proposed the following relations as the general form of river channel geometry,

$$W = C_w Q^{0.5} n^{0.355} S^{-0.156}, \quad (6.7)$$

$$D = C_D Q^{0.3} n^{0.383} S^{-0.206}, \quad (6.8)$$

$$A = C_A Q^{0.8} n^{0.738} S^{-0.362}, \quad (6.9)$$

where  $W$ ,  $D$ ,  $A$ ,  $Q$ ,  $n$  and  $S$  represent the channel width, depth, cross-sectional area, flow discharge, channel average roughness and slope respectively. When the coefficients  $C_w$ ,  $C_D$  and  $C_A$  take constant values of 4.059, 0.427 and 1.733 respectively, the above relations explain 87, 92 and 99 per cent of the variances, respectively, for 529 field observations mostly from natural rivers.

## 6.2 Validation of a downstream hydraulic geometry model for Scottish rivers

In an attempt to validate a downstream hydraulic geometry model applicable for estimating Scottish rivers top width, records of real cross-sections were analysed.

The Scottish Environmental Protection Agency (SEPA) holds historical records of all the flow measurements made in constructing stage-discharge relationships for their gauging stations. For most natural channels, flow is estimated by integrating velocity measurements over the cross-sectional area of flow. Area is estimated by measuring the depth at discrete points on the cross-section. These data are held on a variety of hard copy formats at the local SEPA offices. Therefore, their archives were trawled and the hard copy data were translated into a digital database. This allowed the discrete measurements of depth to be integrated (using a trapezoidal rule) to get cross-sectional areas associated with flows. It also allowed estimation of top-widths (difference between start and end points of depth measurements).

Nixon (1959) suggested that rivers adjust to a single dominant discharge, and Hey (1975) that the frequency of this dominant discharge equals the frequency of bankfull flow. Therefore a good estimation of stable river top widths would be given by the application of a downstream hydraulic geometry corresponding to bankfull flow. Hey and Thorne (1986) recommended the use of the 1.5-year return period flood if bankfull discharge was not available. However these channel shaping bankfull discharges will occur with different frequencies in different networks. Indeed Nixon (1959) reported them as occurring with return periods in the range 0.09 to 2.7 years for UK rivers. When upland rivers are isolated from this study, for example the upper Tyne or the upper Wye, the frequency with which bankfull discharges occur is lower and consequently their return period is longer. For each of the sites analysed in this thesis, the value of  $Q_{med}$ , which is the FEH estimation of the median flood of the annual maximum series, is available (NERC, 1999). This corresponds to the 2-year return period flood. In the light of Nixon's report (1959), this value is used as an acceptable estimate of bankfull flow.

From SEPA's records, there were a range of flow rates, cross-sectional areas and widths associated with each measurement of flow at all the gauging stations. Furthermore, we had estimates of bankfull discharges courtesy of the statistical analyses conducted by CEH. However, gauging stations are rarely visited when the river is at bankfull. Therefore, we needed to infer top-width at bankfull from the available data if we were to derive a downstream hydraulic geometry relationship for Scottish rivers. This was achieved by extrapolating to bankfull characteristics using at a station hydraulic geometries.

Therefore, for each of 30 gauging stations located at individual sites in Scotland, an at a station hydraulic geometry relationship for top-width was derived. This is illustrated for one station in Figure 6.1. It involved plotting top-width versus flow rate on a logarithmic scale, for the available range of flow measurements, and applying linear regression to derive the hydraulic geometry equation. Using the derived at a station hydraulic geometry equation and  $Q_{med}$ , as an estimate of bankfull flow, the value of top width corresponding to bankfull discharge was extrapolated.

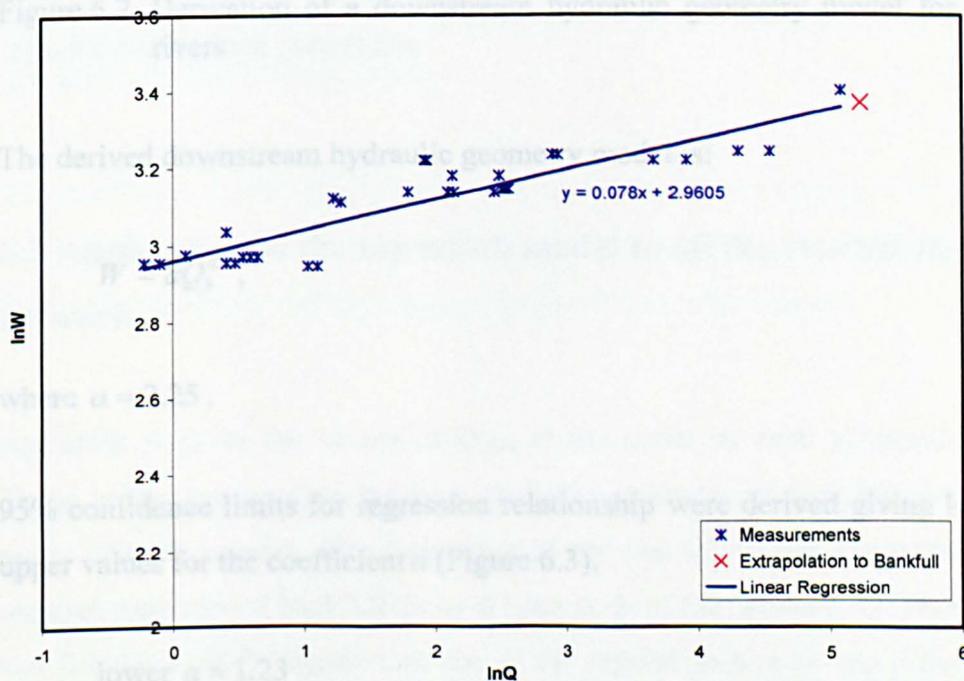


Figure 6.1 Derivation of at a station hydraulic geometry

This was repeated for each site, allowing a relationship to be developed between estimates of top-widths and bankfull discharges. Figure 6.2 shows estimated top-widths versus bankfull discharges on a logarithmic scale for the 30 sites. Linear regression was used to derive a downstream hydraulic geometry model corresponding to bankfull suitable for Scottish rivers.

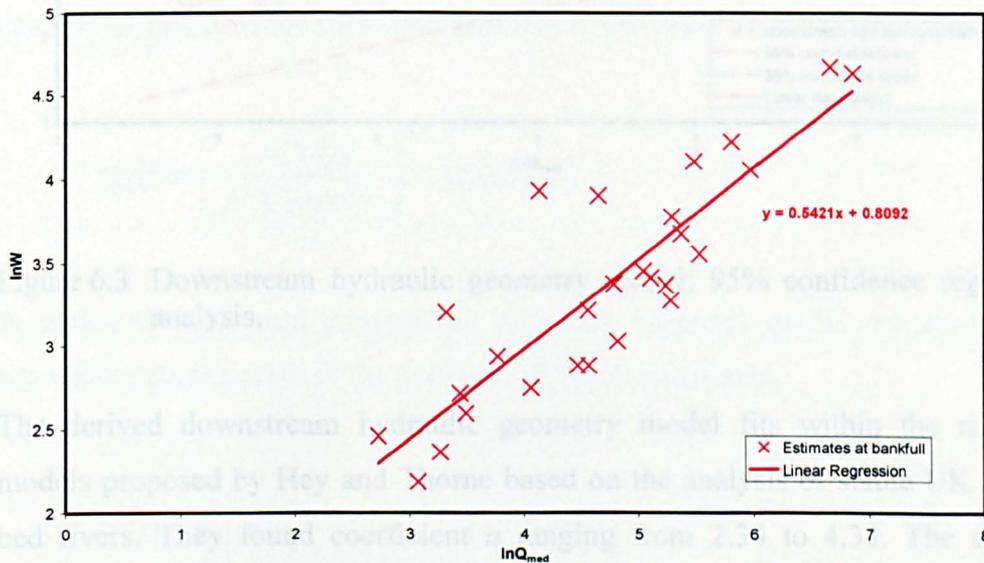


Figure 6.2 Derivation of a downstream hydraulic geometry model for Scottish rivers

The derived downstream hydraulic geometry model is:

$$W = aQ_b^{0.5}, \quad (6.10)$$

where  $a = 2.25$ .

95% confidence limits for regression relationship were derived giving lower and upper values for the coefficient  $a$  (Figure 6.3),

lower  $a = 1.23$

upper  $a = 4.10$ .

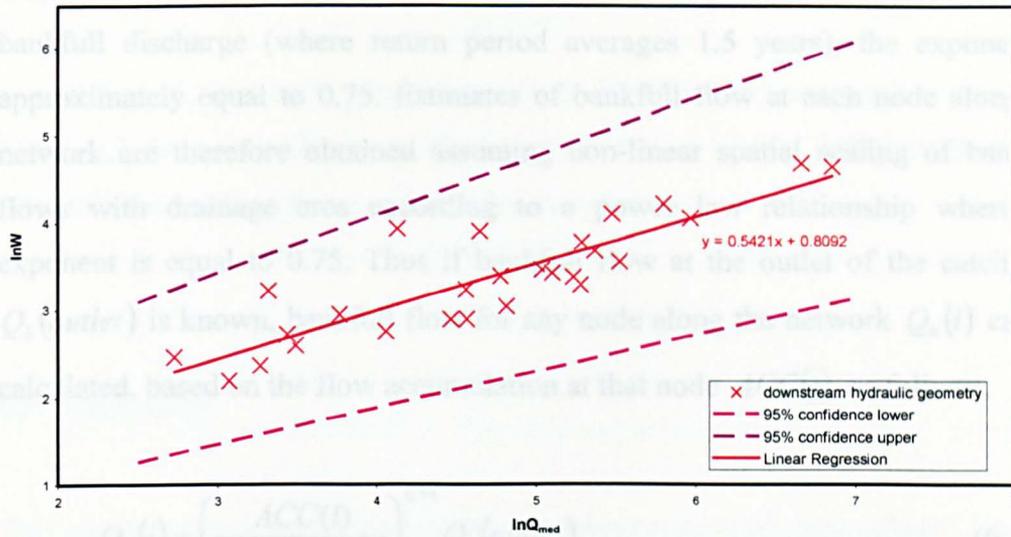


Figure 6.3 Downstream hydraulic geometry model; 95% confidence regression analysis.

The derived downstream hydraulic geometry model fits within the range of models proposed by Hey and Thorne based on the analysis of stable UK gravel-bed rivers. They found coefficient  $a$  ranging from 2.34 to 4.33. The analysis suggests that for Scottish rivers, slightly lower values of coefficient  $a$  apply. Therefore it would appear that hydraulic geometry relationships could be used to estimate conveyance parameters.

### 6.3 Application of the top width model to all the reaches in a network

Appendix A gives the values of  $Q_{med}$  at the outlet of each extracted network. These values are used as the best available estimates of bankfull flows at the outlet of each network. The estimation of channel top-widths along the network requires estimates of bankfull flows at each node of the network. Rodriguez-Iturbe and Rinaldo (1997) suggest that one of the regular factors among different river basins is the relationship between discharge of a given frequency of occurrence and drainage area, and that this relationship is a power law of the form  $Q \propto A^n$ ,

where the exponent  $n$  varies according to the frequency of the considered flow. At bankfull discharge (where return period averages 1.5 years), the exponent is approximately equal to 0.75. Estimates of bankfull flow at each node along the network are therefore obtained assuming non-linear spatial scaling of bankfull flows with drainage area according to a power law relationship where the exponent is equal to 0.75. Thus if bankfull flow at the outlet of the catchment  $Q_b(outlet)$  is known, bankfull flow for any node along the network  $Q_b(i)$  can be calculated, based on the flow accumulation at that node  $ACC(i)$ , as follows,

$$Q_b(i) = \left( \frac{ACC(i)}{ACC(outlet)} \right)^{0.75} Q_b(outlet). \quad (6.11)$$

Applying the validated downstream hydraulic geometry model (equation 6.10), top widths at each node of the network,  $W(i)$ , are calculated,

$$W(i) = a \left( \left( \frac{ACC(i)}{ACC(outlet)} \right)^{0.75} Q_b(outlet) \right)^{0.5}. \quad (6.12)$$

The top-width along each reach is assumed to be an average of the top widths at upstream and downstream nodes.

The coefficient  $a$  is uncertain but known to vary between 1.23 and 4.33. Therefore an ensemble of possible realisations of channel morphology is developed for each channel by selecting values of  $a$  lying between these two values.

## 6.4 Roughness coefficient

Similarly, parameterisation of the networks in term of Manning's roughness coefficient is performed by selecting a variety of Manning's  $n$  covering a wide range of possible values varying from 0.02 to 0.08 (Chow, 1959; Hornberger et al., 1998).

In practice, it was only feasible to select 10 physically reasonable values of Manning's  $n$  and coefficient  $a$ . This yielded a total of 100 different physically plausible parameter sets for each catchment network.

## **6.5 Summary**

This chapter completes the derivation of catchment network data suitable for flow routing. Although all the characteristics describing catchment network structure could be extracted from real river networks, the estimation of channel conveyance parameters required using empirical results of river morphology, in particular for channel cross-section geometry. A generic downstream hydraulic geometry model, validated based on the analysis of flow and cross-section measurements for Scottish rivers, was applied to estimate top-width along the network. In order to take into account uncertainties in the estimation of network conveyances, 100 combinations of hydraulic geometry and Manning's roughness coefficients have been selected, leading to the generation of 100 plausible representations of each catchment network structure.

## *Chapter 7*

### **7 Simulation**

The central tenet of this thesis is that, since hydrological data on any individual component in the hydrological cycle of an entire basin are difficult to obtain, the only pragmatic alternative for characterising them is simulation using an appropriately parameterised physically based model. In addition, to allow patterns to be sought and statistically meaningful relationships between the simulated hydrological response and potential classifier variables to be derived, the model has to be applied to as many river basins as possible. This chapter describes flow routing simulations performed on more than 50 river basins throughout Scotland. The study has been restricted to Scotland for two reasons: firstly, the data on network morphology used in Chapter 6 could be obtained from SEPA and secondly, it limited the number of basins to that which could be reasonably handled during the course of a three-year PhD. First, the rainfall scenarios used to force the model in the later simulations are described, along with the pair of variables that are used to characterise responses. Then, a suite of preliminary simulations was conducted to determine whether the premise of the approach to characterising networks was valid and how best to characterise responses. Finally, flow through all the networks is simulated and the hydrographs characterised.

#### **7.1 Input hyetographs and response hydrographs**

Having attempted to retain as much information from real networks as possible, it is imperative that the flow scenarios simulated are physically realistic. However, if the local climatic conditions that prevailed at each of the networks were explicitly incorporated into the simulations, then it would be very difficult to

compare networks. Therefore, the same input hydrographs were applied to the entire set of networks. Input hydrographs were derived by assuming a rainfall hyetograph that was typical for Scotland and that rainfall is uniformly distributed over the catchment and a fixed proportion of it (runoff ratio) runs off immediately. This gross simplification of real runoff dynamics is warranted because the interest is in network responses and not runoff mechanisms. However, it is important to note that much of the shape of a runoff hydrograph is retained. This is in contrast to many other studies (e.g. Gupta et al, 1996; Veitzer and Gupta, 2001; Menabde et al, 2001) where an instantaneous pulse is used as the runoff hydrograph to force the network routing model. This may be warranted if the system behaves in a truly linear fashion. However, the indications from field studies where the flood peaks follow a power law relationship with area (Smith 1992; Gupta and Dawdy, 1995) are that this is not the case. The network routing model developed in Chapter 4 is inherently non-linear since flow velocities increase with flow depth according to Manning's equation. Thus, non-linearities in the response can be captured. Indeed it will be seen that the shape of the hyetograph affects the outlet hydrograph response in a non-linear manner; in particular that the storm duration affects the way that peak flows change as one moves down through the network. Therefore, a range of realistic hyetographs was used.

### **7.1.1 Hyetographs**

The input hyetograph corresponds to a design extreme rainfall event, constructed using the FEH storm design procedures (NERC, 1999). In constructing such a rainfall hyetograph, the aim is not to reproduce any one particular storm event but rather to have a storm profile with standardised shape and typical rainfall intensities with which to compare the hydrological responses at the outlet of the tested networks.

FEH storm design procedures define rainfall intensity, duration and profile. The FEH contains UK wide maps showing median annual maximum rainfall values (RMED) for 8 durations varying between 1 hour and 8 days, which have been interpolated from rain gauge data. RMED events correspond roughly to a 1 in 2 year extreme rainfall. According to the storm depth versus flood peak return

period graph for the UK (Shaw, 1994), a rainfall event that has a given frequency of occurrence will produce a flood peak that has a higher frequency of occurrence. Therefore the choice of RMED rainfall events insures that the flows generated in channels designed based on bankfull equal to  $Q_{med}$  will stay in bank. For a 1-hour duration, the median total depth of rain falling during a RMED event in Scotland is 12mm. This value drops once an areal reduction factor is applied. The areal reduction factor is assumed to vary only with area and rainfall duration. For a 1-hour duration, the areal reduction factors are 0.80 and 0.50 for 100 km<sup>2</sup> and 5000 km<sup>2</sup> areas respectively. As the same input hydrograph needs to be applied to the entire set of networks which drain areas of varied sizes, an average reduction factor of 0.65 corresponding to the arithmetic average between 0.80 and 0.50 is applied to reduce the total depth of the RMED event from 12mm to 7.8mm. Following FEH recommendations, the '75% winter profile' design storm formula is used to distribute 7.8mm over 1 hour and design a symmetric, single-peaked and bell-shaped storm profile, which is shown in Figure 7.1.

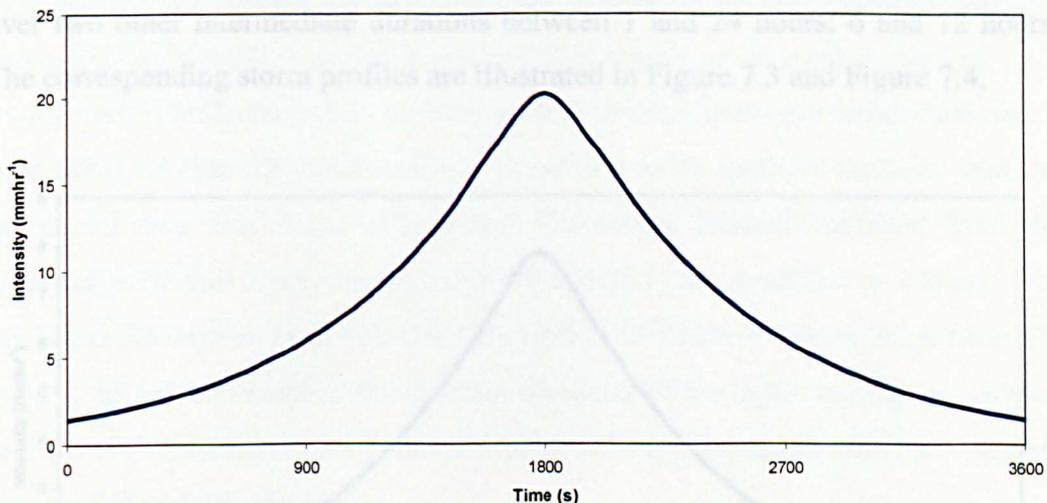


Figure 7.1 Profile of the 1-hour design storm used for simulation

For a 1-day duration, the RMED average value is 60 mm. For a 1-day duration, the areal reduction factors are 0.94 and 0.85 for 100 km<sup>2</sup> and 5000 km<sup>2</sup> areas respectively. An average reduction factor of 0.90 corresponding to the arithmetic average between 0.94 and 0.85 is applied to reduce the total depth of the RMED event from 60mm to 54mm. The storm profile obtained by distributing these

54mm using the FEH '75% winter profile' design storm formula is shown in Figure 7.2.

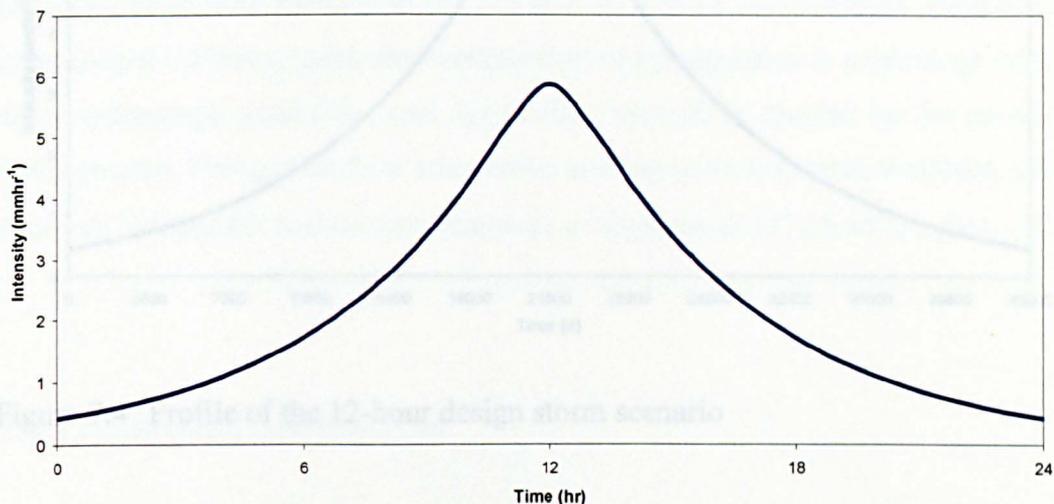


Figure 7.2 Profile of the 1-day design storm used for simulation

Linear interpolation was used to calculate areal rainfall amounts to be distributed over two other intermediate durations between 1 and 24 hours: 6 and 12 hours. The corresponding storm profiles are illustrated in Figure 7.3 and Figure 7.4.

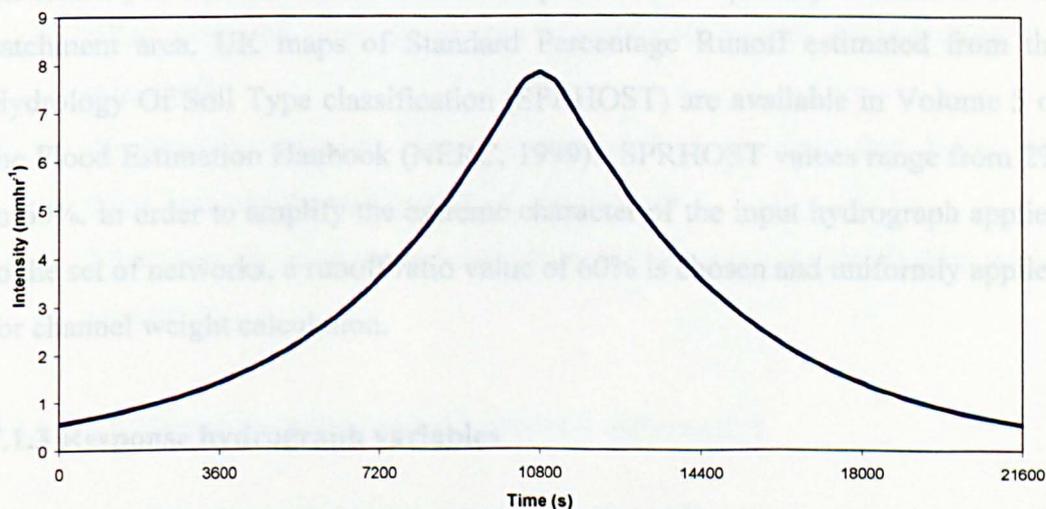


Figure 7.3 Profile of the 6-hour design storm scenario

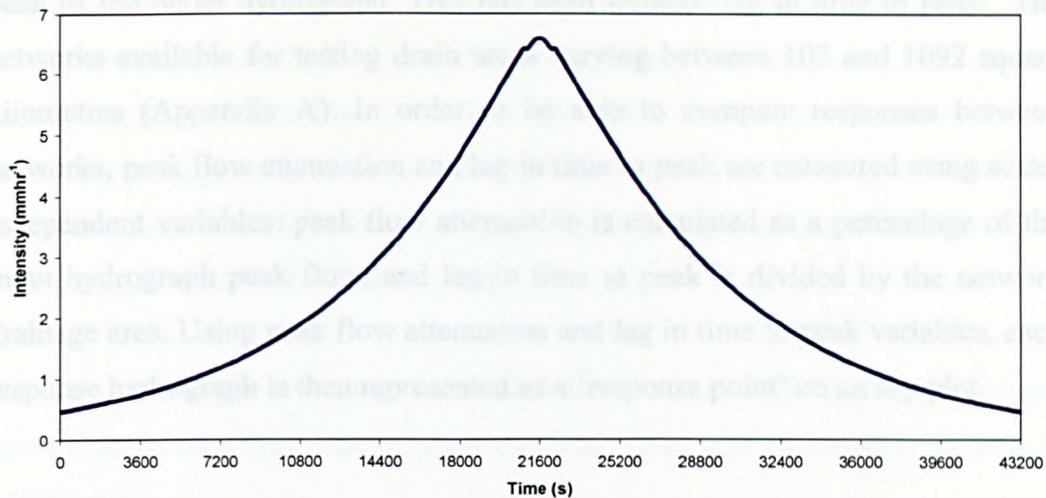


Figure 7.4 Profile of the 12-hour design storm scenario

### 7.1.2 Lateral inflow

As explained in section 4.2 of Chapter 4, the rainfall hyetographs are multiplied by channel weight for each reach in order to produce runoff that enters directly into the channel (lateral inflows). Channel weight is the area draining, laterally across the river bank, into the reach per unit length of channel multiplied by a runoff ratio. The focus here is on flow routing through networks rather than runoff generation, so that the runoff ratio is simplified to be spatially uniform over the catchment area. UK maps of Standard Percentage Runoff estimated from the Hydrology Of Soil Type classification (SPRHOST) are available in Volume 5 of the Flood Estimation Handbook (NERC, 1999). SPRHOST values range from 2% to 60%. In order to amplify the extreme character of the input hydrograph applied to the set of networks, a runoff ratio value of 60% is chosen and uniformly applied for channel weight calculation.

### 7.1.3 Response hydrograph variables

For the preliminary simulations, a fairly standard approach was used to characterise the simulated outlet response hydrographs. Attenuation was measured using the relative drop in peak flow rate between the input runoff hydrograph and the outlet hydrograph. The translation, or travel time of the flow wave, was measured using the time between the peak of the input runoff hydrograph and the

peak of the outlet hydrograph. This has been termed 'lag in time to peak'. The networks available for testing drain areas varying between 102 and 1092 square kilometres (Appendix A). In order to be able to compare responses between networks, peak flow attenuation and lag in time to peak are measured using scale-independent variables: peak flow attenuation is calculated as a percentage of the input hydrograph peak flow; and lag in time to peak is divided by the network drainage area. Using peak flow attenuation and lag in time to peak variables, each response hydrograph is then represented as a 'response point' on an x-y plot.

## **7.2 Preliminary simulations**

Prior to describing the simulation whose results were eventually used to characterise the river networks, a suite of preliminary simulations were conducted to determine:

1. Whether the underlying premise that different shaped networks produced different shape hydrographs was valid?
2. What was the most appropriate way to incorporate uncertainty in conveyance parameters?
3. What were the most appropriate variables to use in characterising the response?
4. Whether the networks alone, uncoupled from the land that drains into them, have their own signature hydrographs?
5. What were the most appropriate storm hyetographs (or rainfall time series) to use?

### **7.2.1 Do different shaped networks respond differently?**

Inherent in this whole thesis is the premise that different shaped networks will respond differently. Whilst this may seem intuitive, it has not been explicitly tested. Some authors have intuited similarly, for example Rodriguez-Iturbe and Rinaldo (1997), in a highly influential book where the fractal characteristics of river networks are used to describe network responses. Their theory is called the

Geomorphologic Unit Hydrograph (GUH). However, the GUH is explicitly built on an assumption that network shape dominates the response, not on the underlying principles of open channel hydraulics. Indeed, it violates some of these, the assumption of constant water velocity everywhere in a network being a case in point. Therefore, it is perhaps unsurprising that the GUH produces different hydrographs for different shaped networks.

Here, the premise is tested using one network extracted from GIS databases and two synthetic networks. The network extracted from GIS databases is illustrated in Figure 7.5.

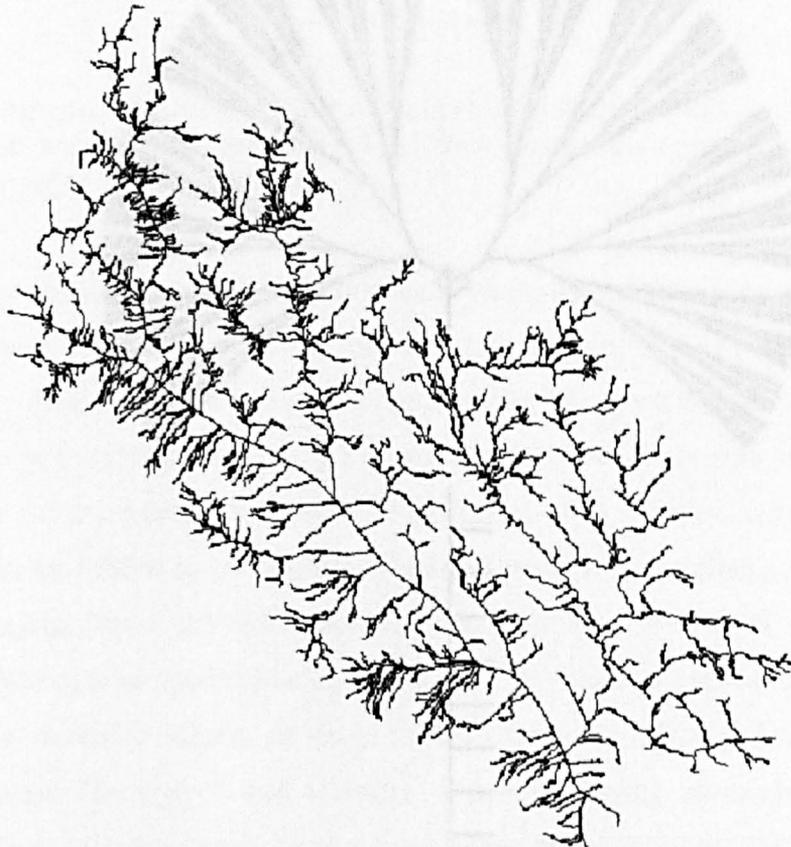
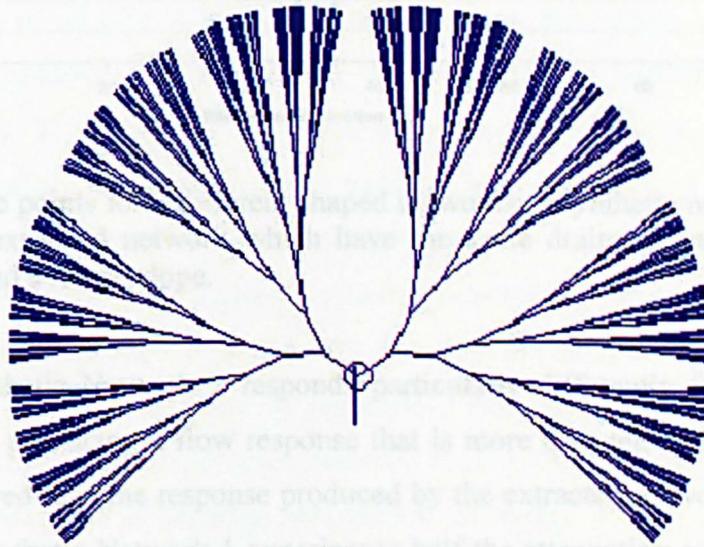


Figure 7.5 The network extracted from GIS databases which is used as 'model' for generation of synthetic networks. It has code w3001 (see Appendix A).

In order to compare network structures, the area draining through every unit length of this extracted network was normalise to an average calculated based on

50 real Scottish river basins. Similarly, the channel slope in every link was normalised to the average corresponding to these 50 river basins. The two synthetic networks were generated in a MATLAB program such that they had the same total length, area draining per unit length and slope as the normalised extracted network. The first network is very broad; if described in the upstream direction each channel splits into two channels until the total length is reached (Figure 7.6a). The second is long and thin; again describing it in the upstream direction, when a channel splits into two upstream channels, only one of these goes on to split again (Figure 7.6b).

(a)



(b)



Figure 7.6 Diagram of 2 synthetic networks with same total length: (a) Synthetic Network 1; (b) Synthetic Network 2 (partial representation).

Figure 7.7 shows the simulated response points to the 24-hour storm for these three networks for one value of the conveyance parameters  $a$  and  $n$ .

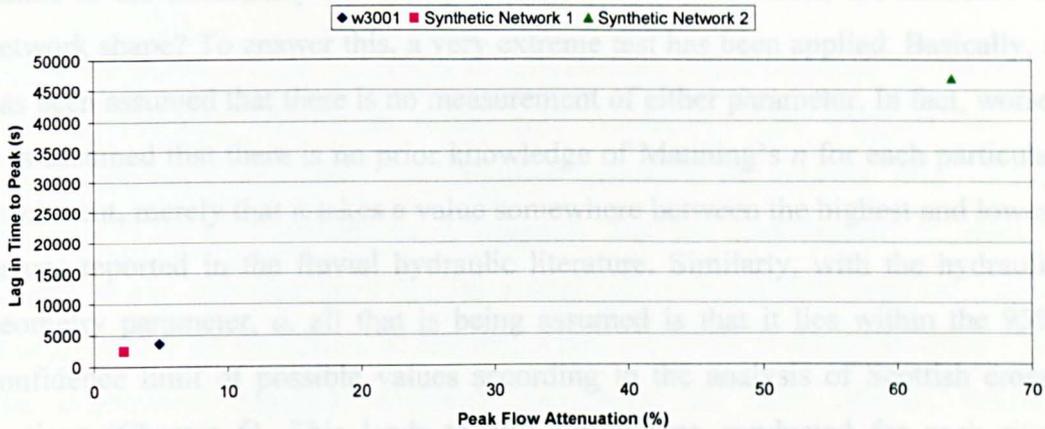


Figure 7.7 Response points for 3 different shaped networks; 2 synthetic networks and an extracted network which have the same drainage area, total length and average slope.

It is clear that Synthetic Network 2 responds particularly differently from the other two networks, producing a flow response that is more than ten times more attenuated and delayed than the response produced by the extracted network. The flow response in Synthetic Network 1 experiences half the attenuation and delay that the flow in the extracted network experiences. The extreme response of Network 2 can be related to its 'extreme shape': Network 2, extremely long and thin, looks nothing like a network extracted from real data, whereas the shape of Network 1, although far too regular to correspond to a natural network, is more similar to the dendritic nature of most natural networks. The differences in response between Network 1 and Network 2 can intuitively be explained by looking at expected travel times of the flood wave from the extremities of the network to the outlet. Conveyance parameters and slope being the same for the two networks, travel time through each link of the network will be dictated by link length and therefore flow response by the distribution of travel distances to the outlet. The maximum distance to the outlet is much higher in Network 2 than in Network 1, which will be much more effective at routing water toward the outlet through its broad shape.

Although it is clear that network shape does affect hydrograph shape, the demonstration assumed absolute knowledge of both the roughness and the hydraulic geometry parameters. The question that immediately arises is, if one builds in the uncertainty in these parameters, would this mask the influence of network shape? To answer this, a very extreme test has been applied. Basically, it has been assumed that there is no measurement of either parameter. In fact, worse, it is assumed that there is no prior knowledge of Manning's  $n$  for each particular catchment, merely that it takes a value somewhere between the highest and lowest values reported in the fluvial hydraulic literature. Similarly, with the hydraulic geometry parameter,  $a$ , all that is being assumed is that it lies within the 95% confidence limit of possible values according to the analysis of Scottish cross-sections (Chapter 6). This leads to 100 simulations conducted for each river network using 100 values of the conveyance parameters array  $(a_i, n_j)$ ,  $i, j = 1, 10$ ; where  $n$  is Manning's roughness coefficient, which takes 10 evenly spaced values between 0.02 and 0.08 and  $a$  is the hydraulic geometry coefficient, which assumes 10 evenly spaced values between 1.23 and 4.33 based on the analysis in Chapter 6. This is the most conservative suite of parameters that could be selected to describe the conveyance properties of each network. Figure 7.8 shows lag in time to peak per square kilometre of drainage area versus percentage of attenuation in peak flow for multiple realisations of each of 48 catchments in Scotland.

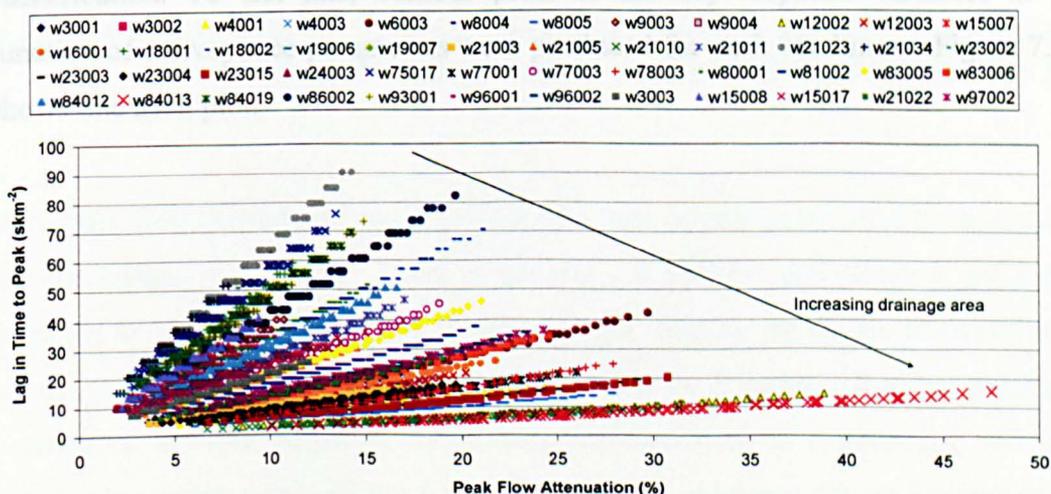


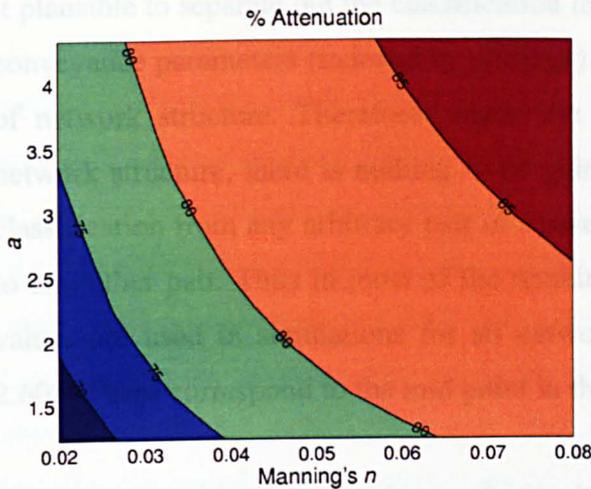
Figure 7.8 Multiple realisations of flow simulated for 48 catchments

On the basis of these results, the effects of conveyance and scale on catchment flow response will now be investigated.

### **7.2.2 Importance of the conveyance parameters**

Figure 7.8 shows a wide variability in the resulting outlet response hydrographs' time to peak and attenuation within each catchment. This suggests that the conveyance parameters have an important influence on network flow routing response. This is important as it suggests that the combination of channel roughness and hydraulic geometry has a huge effect on network response, almost enough to obscure the effects of network shape. However, it should be made clear that experienced practical engineers can place far more realistic bounds, particularly on the roughness coefficient, than have been assumed here. Local knowledge and knowledge of the geology will reduce the bounds on potential values significantly. Basically, no prior knowledge has been assumed. For it to be plausible to apply a geologically based initial classification of river networks to group them together based on their likely conveyance parameters, it is imperative to show that the conveyance parameters do not react in some non-linear way with the network structure to produce idiosyncratic, unexpected responses for some networks. If such idiosyncratic responses were prevalent then small changes in network structure could lead to highly divergent responses and, therefore, it would be impossible to separate off the geological classification from the network shape classification. To test this, contour plots of the key response variables as a function of conveyance parameters were produced for each catchment. Figure 7.9 shows one such plot.

(a)



(b)

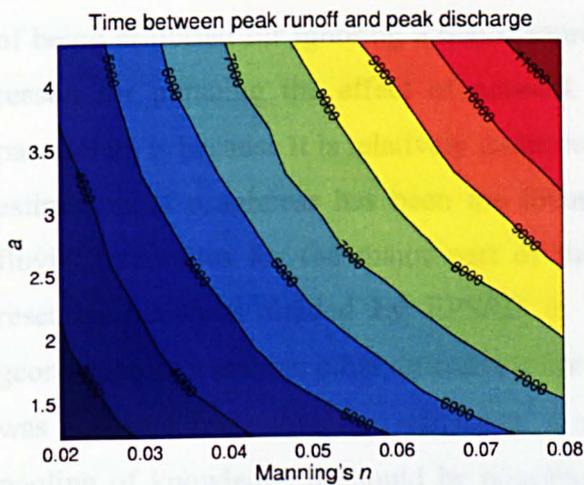


Figure 7.9 Variability of network hydrograph responses with conveyance parameters: (a) Percentage attenuation of peak flow; (b) Lag in time to peak.

Both peak flow attenuation and lag of time to peak appear to be linearly related to the conveyance parameters. There is generally a positive and constant gradient from the bottom left of the plots to the top right. Similar trends are observed for all the other catchments. This is important because it implies that the relative position of network response coordinates corresponding to a particular pair of conveyance parameters will not vary. Thus network response cluster patterns will be conserved from one set of value of conveyance parameters to the other. The

conservation of this pattern is explicitly checked later in this chapter. This makes it plausible to separate out the classification into two steps: first, classifying on the conveyance parameters (indexed by geology) and second, classifying on the basis of network structure. Therefore, where the aim is to classify on the basis of network structure, there is nothing to be gained from multiple simulations. The classification from any arbitrary pair of conveyance parameters can be transposed to any other pair. Thus in most of the remainder of this chapter, a single pair of values are used in simulations for all networks; Manning's  $n = 0.047$  and  $a = 2.608$ . These correspond to the mid point in the range of both.

Whilst the results of the harsh test above clearly demonstrate the importance of having some prior knowledge of the conveyance properties of the network, in this thesis, the effect of network structure is concentrated on. As such, it runs the risk of being criticised for ignoring a major source of the variability in response. The reason for pursuing the effect of network shape over and above conveyance parameters is because it is relatively under researched in the UK. Conveyance and estimation of roughness has been the focus of the UK research community in fluvial hydraulics for the major part of the last century and still is. The two research networks funded by EPSRC in the last 5 years - one on fluvial geomorphology and the other on conveyance, where one of the intended outcomes was software for estimating roughness – are evidence of this. Thus with this pooling of knowledge it should be possible in the near future to make a crude estimate of, for example, Manning's  $n$  from key descriptive variables of the catchment. However, there has been almost no research into the effect of network structure, this is in contrast to the USA where it has assumed a very high profile in the last decade (e.g. Rodriguez-Iturbe and Rinaldo, 1997; Veitzer and Gupta, 2001; Morrison and Smith, 2001). From the few notable exceptions in the UK, such as Naden's research using a network 'width function' to derive a hydrograph on the River Thames (Naden, 1992), it would appear that structure is important. However, what is not clear from this and other research like it (Mesa and Mifflin, 1988) is how best to describe network structure and which elements of the structure exert the strongest influence in shaping a flood wave.

### 7.2.3 The importance of scale

On the basis of Figure 7.8 it is possible to differentiate between catchment responses. However, much of the distinction between catchments can be attributed to differences in catchment drainage area: the response points appear closer to the x-axis for large catchments, whereas for smaller catchments response points are closer to the y-axis (catchment drainage areas are given in Appendix A). Thus the catchment's responses show a dependency on spatial scale. This is despite the fact that the drainage area has been implicitly incorporated in the variables used to characterise the response hydrographs. This suggests that the variables chosen to describe the response hydrographs are not in fact scale independent. If a set of scale independent variables could be found this would greatly ease the classification of networks. This indicates that in order to quantify the effects of channel shape, the influence of both the uncertainty in the conveyance parameters and the area need to be removed.

It is clear that catchment scale has an effect in the network response. This is the case even though the rainstorms in all the simulations to date are uniformly distributed over the entire catchment area. Many researchers have observed a power law relationship between catchment area and peak flow which they attribute to the heterogeneous distribution of rainfall over the catchment area, with small areas experiencing more intense rainfall more frequently than larger areas (e.g. Gupta and Dawdy, 1995; Gupta et al, 1996; Morrison and Smith, 2001). Whilst this will undoubtedly affect the hydrograph, it appears from the simulations presented above that the scale in itself will affect the hydrograph shape. This then begs the question: is the scale effect purely a result of the area that drains into each reach or is it a property of the network structure? In addition, whilst the scale effect is undoubtedly present, it may reflect the inappropriateness of the particular pair of variables used to characterise the hydrological response of the whole network. Perhaps there are better, scale independent variables that might be used. Furthermore, until now, the simulations have been driven by a long 24-hour storm. Storm duration clearly has a strong influence on the hydrograph shape, but does it have an effect in such a way that it would change the classification of networks?

### ***Uniform Lateral inflow***

In the first set of simulations, lateral inflow to each channel is assumed to be proportional to the area draining through that channel. This means that a form of runoff generation process is incorporated in the network flow response simulation (albeit in a very simple way). A more sophisticated runoff generation process could be developed, but this would divert the simulation analysis from its original objective, which consists in characterising the flow routing through river network component in isolation from any other component constituting the overall hydrological response of a catchment. To achieve this goal, network flow has to be completely divorced from runoff generation processes. To this end, the simulations have been modified so that the runoff into each channel per unit length of channel is constant. This in effect means that every stretch of river in the networks drains the same area and the networks are divorced from the catchments they sit in. In Chapter 4, the concept of channel weight was introduced as being the product of drainage area per unit length and runoff ratio. Here, the uniform channel weight value was chosen equal to the mean channel weight values calculated for the 48 catchments used in the first set of simulations.

### ***Response variables spatial scaling***

By removing the dependency on runoff mechanisms and normalising the area drained per unit length of channel it is now possible to consider the channel structure in isolation. However, the very large variation in lag in time to peak per unit catchment area and percentage attenuation hinted that this particular pair of variables might not be as scale independent as was initially thought. In almost all other studies of how networks respond to a flood wave, a similar pair of variables is used (Moussa, 2003). By dividing variables that describe the downstream hydrograph by total catchment area, one is implicitly assuming that these variables scale linearly with area, or some measure of spatial scale. However, some network flow routing simulation studies, such as the ones conducted by Veitzer and Gupta (2001) and Menabde et al (2001), show that peak flows do not increase linearly as a function of drainage area as you move downstream in a network, but rather that the scaling dependence with drainage area follows a

power law relationship. Reported scaling exponents vary between 0.49 and 0.79 depending on the type of routing model and synthetic networks used. Both studies use a highly simplified description of the flow in channels: analysis of peak flows based on the width function in the first study and flow routing modelled by a linear mass conservation equation in the second one. However, they both focus on network routing rather than on runoff generation, and use the same assumption of spatially uniform runoff into the network as is used here. Prior to these studies, it had long been assumed that the scale dependence of peak flows was borne out of spatial heterogeneities in the rainfall distribution and runoff mechanisms. Therefore, they were significant in changing perspective on the role of network structure. However, the highly simplified models used in these studies means that one is unsure whether the scaling observed is a manifestation of the model structure the authors have imposed or a true feature of networks. For example, in Menabde et al (2001) the cascade of linear reservoirs forced by an instantaneous pulse of rainfall gives a neat analytic form to the hydrograph that falls out of the linear assumption, but does it truly represent the way that a flood wave is modified as it moves through the network? In essence, both models assume uniform constant flow velocity along the network. Whilst it is possible to speculate about this, a more pragmatic approach is to use a model that better represents the hydraulics and put up with the inconvenience of not having a neat analytic solution (that might be wrong anyway). Here, the kinematic wave routing model is a much more realistic representation of channel flows. In addition, description of network structure (derived from OS maps) and of the way the channel cross-sections change within the network (derived from field data), is far more detailed than in either previous study. Therefore, a much more rigorous test of the scale dependence of these variables can be implemented.

The way in which peak flows scale with drainage area was ascertained for 8 networks, arbitrarily chosen but evenly distributed about Scotland. The time and magnitude of the peak flow were recorded at each node in the network. Figure 7.10 shows the peak flows simulated along the 8 networks for a storm of duration 24 hours. Since spatially uniform runoff is applied, the contributing area at each node of the network is measured as the total length of channel upstream of that node.

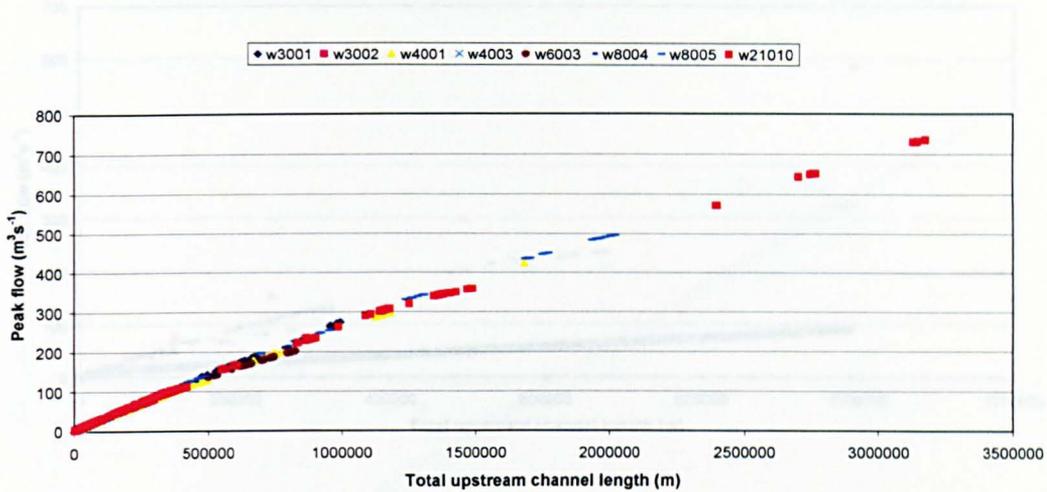


Figure 7.10 Simulated peak flows along 8 networks using the 24-hour duration storm

These simulations show a linear scaling of peak flows with contributing area in all eight networks. This is a surprising result because it does not coincide with the power law relationship obtained by previous investigators. This may be a result of the assumption, made in both Veitzer and Gupta (2001) and Menabde et al (2001) studies, that networks are self-similar. Their studies simulate flow in synthetic self-similar river networks, whereas the simulations presented here are performed on networks extracted from GIS databases.

### *Spatial scaling and network shape*

To see whether radically different network structures that have the same macro-scale properties (drainage area and total channel length) as an extracted network can produce a different relationship between peak flow and scale, the simulations were conducted on the two synthetic networks introduced earlier in this chapter: one broad (Figure 7.6a) and the other very long and narrow (Figure 7.6b). Figure 7.11 shows the results of simulations performed on these two synthetic networks and the extracted network from which their total channel length was derived.

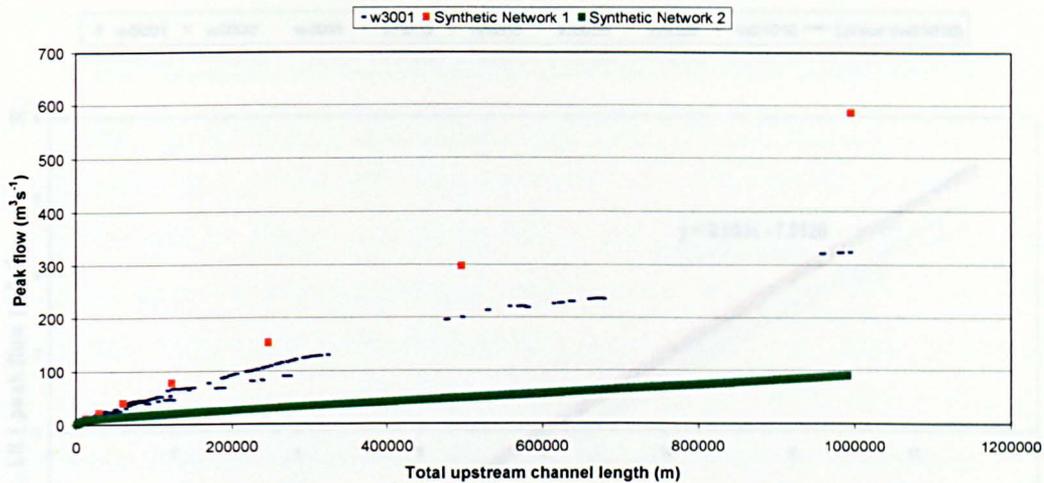


Figure 7.11 Simulated peak flows along 2 synthetic networks and an extracted network which have the same drainage area, total length and average slope

Again, even with these highly divergent network structures, peak flows show linear scaling dependence on drainage area. However, the gradient does differ between each of the networks. Thus, network structure does not seem to be responsible for the linearity of peak flow scale dependency on drainage area. However, as previously shown (Figure 7.7), it does influence the network response to a runoff event.

This calls into question the generality of the scaling relationships reported by Veitzer and Gupta (2001) and Menabde et al (2001). Perhaps it is a function of the scenarios used in their simulation. Both these studies apply an instantaneous runoff input hydrograph, whereas the simulation scheme used here applies the 24-hour input hyetograph described in section 7.1.1 (Figure 7.2).

### *Spatial scaling and storm duration*

The network routing model was forced by the suite of storm profiles of durations varying between 1 and 24 hours described earlier. The simulated peak flows are plotted against the length of channel upstream in Figure 7.12.

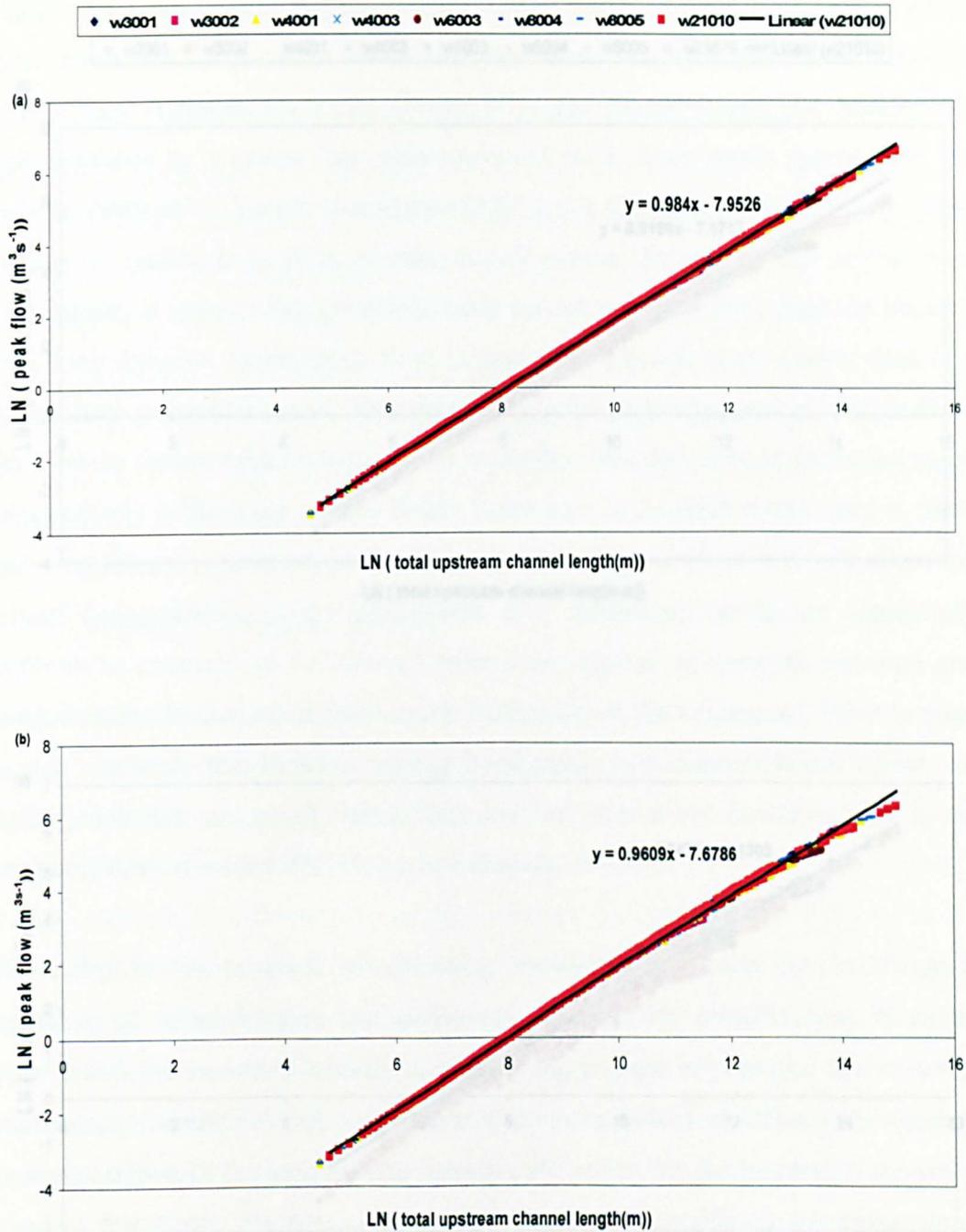


Figure 7.12 Logarithmic plots of simulated peak flows along 8 networks for 5 storm durations: (a) 24-hour storm; (b) 12-hour storm.

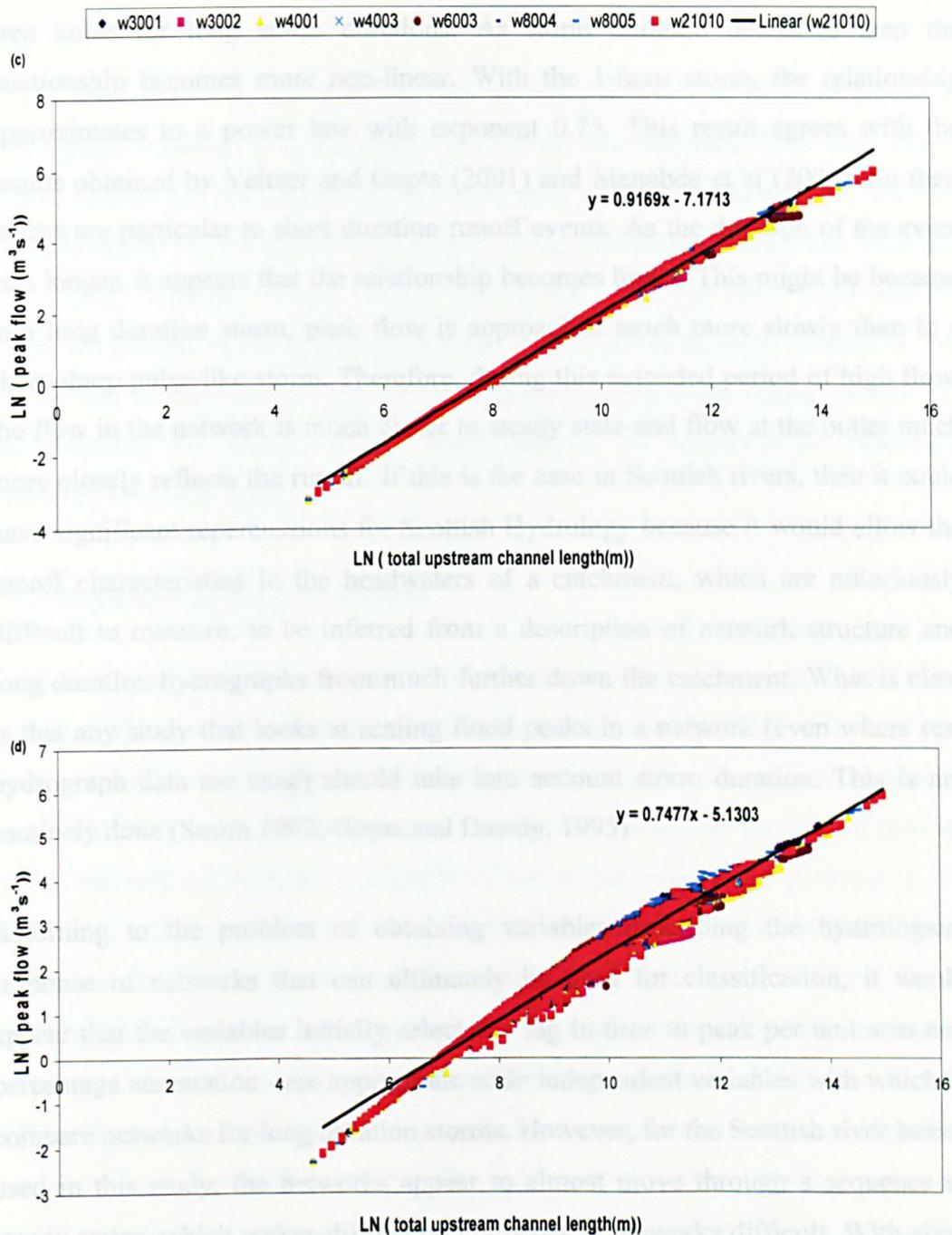


Figure 7.12 (continued) (c) 6-hour storm; (d) 1-hour storm.

Figure 7.12 shows that the linear relationship between peak flow and drainage area holds for long storm durations. As storm duration decreases then the relationship becomes more non-linear. With the 1-hour storm, the relationship approximates to a power law with exponent 0.75. This result agrees with the results obtained by Veitzer and Gupta (2001) and Menabde et al (2001). So their results are particular to short duration runoff events. As the duration of the event gets longer, it appears that the relationship becomes linear. This might be because in a long duration storm, peak flow is approached much more slowly than in a short-sharp pulse-like storm. Therefore, during this extended period of high flow, the flow in the network is much closer to steady state and flow at the outlet much more closely reflects the runoff. If this is the case in Scottish rivers, then it could have significant repercussions for Scottish Hydrology because it would allow the runoff characteristics in the headwaters of a catchment, which are notoriously difficult to measure, to be inferred from a description of network structure and long duration hydrographs from much further down the catchment. What is clear is that any study that looks at scaling flood peaks in a network (even where real hydrograph data are used) should take into account storm duration. This is not routinely done (Smith 1992; Gupta and Dawdy, 1995)

Returning to the problem of obtaining variables describing the hydrological response of networks that can ultimately be used for classification, it would appear that the variables initially selected – lag in time to peak per unit area and percentage attenuation - are appropriate scale independent variables with which to compare networks for long duration storms. However, for the Scottish river basins used in this study, the networks appear to almost move through a sequence of steady states, which makes differentiating between networks difficult. With short duration storms, it is easier to differentiate between networks. However, the variables are scale dependent. Therefore, in the section that follows, a 1-hour storm duration is used. The scale problem is circumvented by only comparing networks that drain the same area.

### 7.3 'At a scale' simulations

It has been shown that in order to characterise the effect of network shape on hydrograph response it is necessary to divorce that response from both runoff characteristics and the uncertainty in conveyance parameters. Applying a spatially uniform lateral inflow achieves the first of these. It was also shown that selecting one pair of 'average' values for the conveyance parameter array ( $a, n$ ) will not affect the pattern of the resulting network flow response clusters. Finally, it was shown that networks could best be differentiated using short 1-hour storms, but that reducing the duration of the storm leads to the constraint of comparing networks of similar scale.

#### 7.3.1 Definition of the scale criteria

Performing 'at a scale' analysis of river network flow response requires the selection of sets of river networks of similar spatial scale or 'size'. The size of a river network is commonly identified with the size of its associated river basin, and therefore measured by its drainage area. In this thesis, the analysis of the role of the network in shaping hydrograph response has been stripped down to the network in isolation from the catchment it sits in. Too many previous studies have become embroiled in representing the combined effects of network and runoff (Smith, 1992; Gupta and Dawdy, 1995; Gupta et al, 1996) and thus weaken their conclusions about network structure because it is impossible to determine whether it is network shape, hyetographs or runoff characteristics that dominate. This then prompted some researchers to describe response of highly idealised networks forced with highly idealised storms. For example, the use of fractals to describe network structure and in particular the highly idealised Peano networks became the focus (Rodriguez-Iturbe and Rinaldo, 1997; Veitzer and Gupta, 2001; Menabde et al, 2001). Using these as a surrogate for true networks in the belief that almost all natural networks adhere to a self similar structure, it was shown that the magnitude of a flood peak scales with network size (Veitzer and Gupta, 2001; Menabde et al, 2001). However, this essentially assumes that all networks behave in the same way. Morrison and Smith (2001) were the first to attempt to disentangle the effects of runoff, hyetograph, rainfall spatial distribution and

network shape, in a study that used real networks. They used a less sophisticated routing model than that which is used in this thesis which, like those of Veitzer and Gupta (2001) and Menabde et al (2001), assumed a constant water velocity throughout the network. Nonetheless, their conclusion was that at scales greater than 100 square kilometres, the structure of the network became extremely important. However, implicit in their discussion of the results is an admission that this is in essence an empirical observation from their model results and, thus, to really test whether it is possible to distinguish between networks, the network has to be divorced from runoff mechanisms. Only then is it possible to begin to speculate on the effects of spatially varying rainfall or heterogeneous runoff mechanisms. It is for this reason that, whilst retaining the essential structure of the real networks, the area draining into them has been standardised. This is achieved by assuming a uniform lateral inflow per unit channel length. Under this assumption, two catchments with the same drainage area will not necessarily receive the same total runoff because similarity in drainage area does not imply similarity in total channel length. However, if uniform lateral inflow per unit channel length is applied on two catchments with same total length, both of them will receive the same total runoff, and each of them will route that amount of water towards the outlet. Therefore, when comparing network structure in isolation, networks with the same total length should be selected. This allows simple un-scaled variables like the lag in time to peak, in seconds, and the percentage of attenuation of the peak flow to be used to characterise the network outlet response hydrographs.

### **7.3.2 Isolation of networks with same total length**

Two scales were investigated: networks with an approximate total length of 300 kilometres and networks with an approximate total length of 600 kilometres. Clearly, if restricted to river basins in Scotland with precisely these total lengths of channel upstream of the gauging station (Appendix A), the sample would be very small indeed. Therefore, in catchments with longer total channel lengths, a single nested network was extracted. This was achieved using an algorithm that moves from the outlet of the network towards the extremities, searching for a node with the desired total upstream channel length. In a large network, there can

be many sub-networks with similar total length. To ensure that the statistical analysis of responses was not biased by having many networks with very similar structures, only one nested network was selected from each of the larger basins.

### 7.3.3 Synthetic channel top widths

The estimation of channel top widths along the network is based on the values of bankfull flow at nodes. These values are calculated based on the value of bankfull flow at the outlet of the network, scaling dependence on drainage area according to a power law. This in a way incorporates characteristics of the drainage area into the network structure model, which this attempt to characterise the flow routing component in isolation is trying to avoid. Therefore, synthetic values of contributing areas (flow accumulation) at each node of the network are generated that correspond to the synthetic uniform channel weight used for lateral inflow calculation. These synthetic values of flow accumulation are then used for bankfull flow scaling, and therefore channel top width scaling. Moreover, in order to eliminate any source of variability in the network structure that is related to characteristics of the catchment area, the same synthetic value of outlet bankfull flow is used for all the networks. Both values of synthetic channel weight and outlet bankfull flow equal the 'real' values for one network arbitrarily chosen amongst the networks extracted from GIS. So in other words, the values of channel top widths along all the networks are based on the same value of outlet bankfull flow, and scale according to the same value of drainage area per unit channel length.

Before presenting the final simulation results, which will be retained for the statistical multivariate analysis, a list will be drawn of the different steps that have gradually made possible a complete isolation of the network shape from the characteristics of the catchment it sits in. These consist in:

1. Selecting one pair of average values for the conveyance parameter array  $(a, n)$ , which is uniformly applied along the network.
2. Generating a synthetic uniform contributing area per unit length of network, which is used for the application of a uniform lateral inflow.

3. Generating synthetic values of channel top widths, which are scaled on the basis of the synthetic uniform contributing areas and a value of outlet bankfull flow common for all the networks.
4. Isolating networks of same total channel length.

### 7.3.4 Simulation results

A final pool of 46 networks with 300-kilometre total length was created from the networks in Table 5.2. One network was arbitrarily chosen to serve as a ‘model’ for generation of synthetic flow accumulation and outlet bankfull flow values. The ‘real’ values of total drainage area and total network length associated with that network were used to calculate an average value of contributing area per unit length of channel. These were then used in the simulations of flow in all 46 networks. Scaling of top width as one moves up the network was performed as previously described using the upstream hydraulic geometry relationship,  $W = aQ^b$ , where  $a$  equals 2.61 and  $b$  equals 0.5. Manning’s  $n$  was fixed at 0.047. Figure 7.13 shows time to peak lag versus peak flow percentage of attenuation corresponding to flow routing simulations conducted on the pool of 46 networks with 300-kilometre total length.

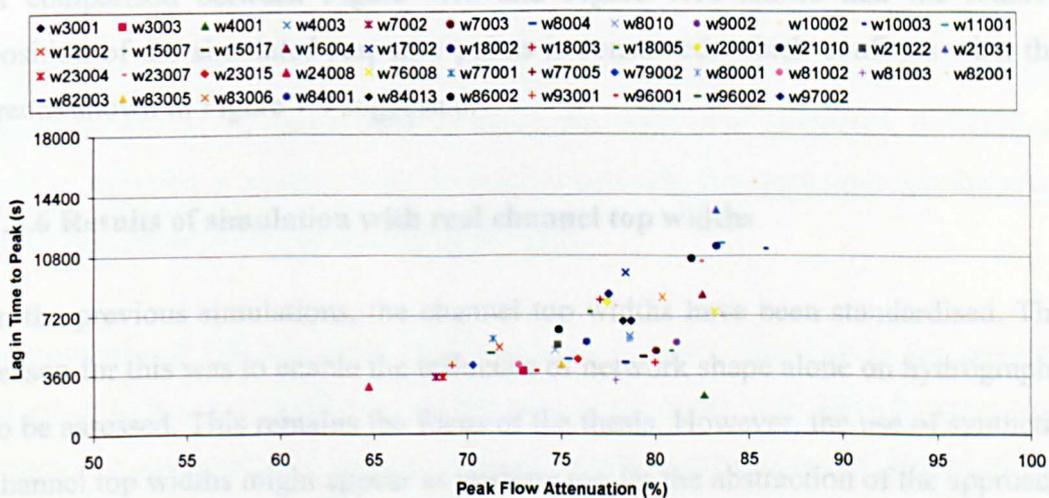


Figure 7.13 Simulated responses for 46 networks of 300km total length

### 7.3.5 Results of simulation for another conveyancing value

In order to test the assumption that the relative position of the coordinates of lag in time to peak and percentage attenuation are independent of the specific values chosen for the conveyance parameters, the simulations were repeated for a second arbitrarily chosen pair of conveyance parameters. Results of simulations conducted for (a,n) equal to (3.64,0.067) are shown in Figure 7.14.

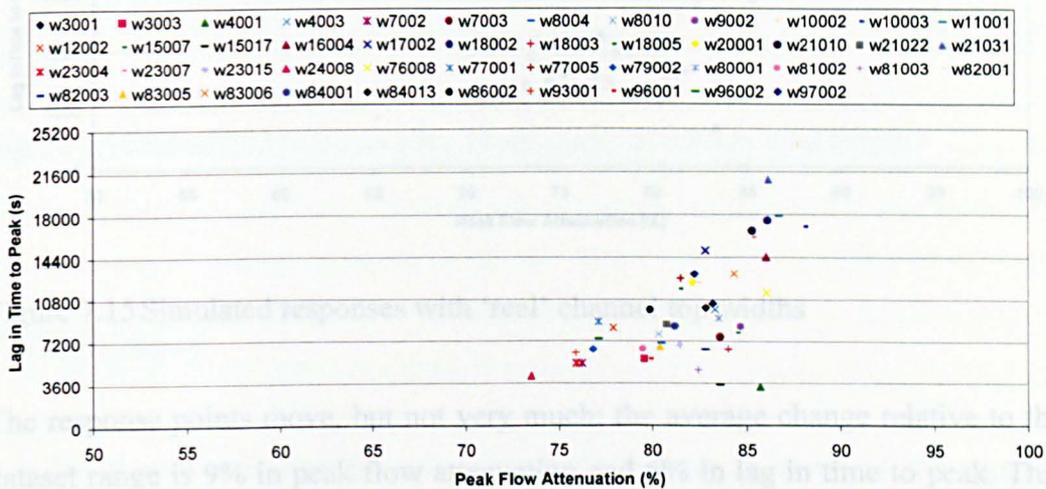


Figure 7.14 Simulated responses for conveyance parameters (a,n) equal to (3.64,0.067)

A comparison between Figure 7.13 and Figure 7.14 shows that the relative position of the simulated response points is conserved, which confirms what the trends shown in Figure 7.9 suggested.

### 7.3.6 Results of simulation with real channel top widths

In the previous simulations, the channel top widths have been standardised. The reason for this was to enable the influence of network shape alone on hydrographs to be assessed. This remains the focus of the thesis. However, the use of synthetic channel top widths might appear as pushing too far the abstraction of the approach by standardising an element of the network which is an integral part of it. Therefore it is interesting and might be reassuring to see what happens to the network responses when 'real' top widths are reintroduced. Figure 7.15 shows the

responses for the same storm and same networks as in Figure 7.13, but with ‘real’ top widths.

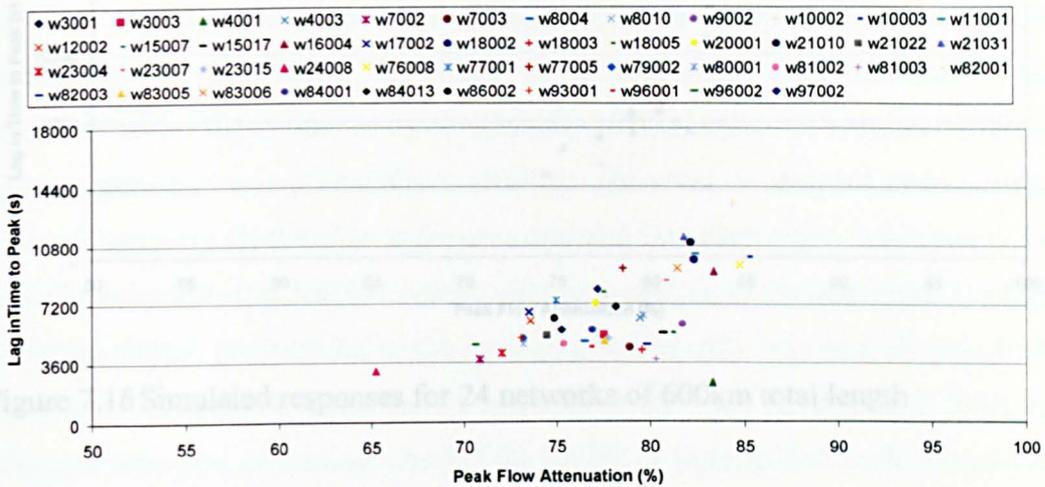


Figure 7.15 Simulated responses with ‘real’ channel top widths

The response points move, but not very much; the average change relative to the dataset range is 9% in peak flow attenuation and 6% in lag in time to peak. This suggests that network shape dominates the response and the cluster pattern identified in the following chapter.

### 7.3.7 Results of simulation at another scale

In order to test the assumption that network flow response best classifiers will not be dependent on the spatial scale at which the analysis is conducted, the analysis will be repeated at another scale. Simulated network flow responses are therefore generated for networks of another total length.

24 networks of 600-kilometre total length are isolated. Simulations are repeated identical to the ones conducted on the pool of networks with 300-kilometre total length: the same value of uniform channel weight as applied on the 300-kilometre networks is used; the value of synthetic outlet bankfull flow is derived by up-scaling the one applied to the 300-kilometre networks. The simulation results are shown in Figure 7.16. In the next chapter, these simulation results are analysed.

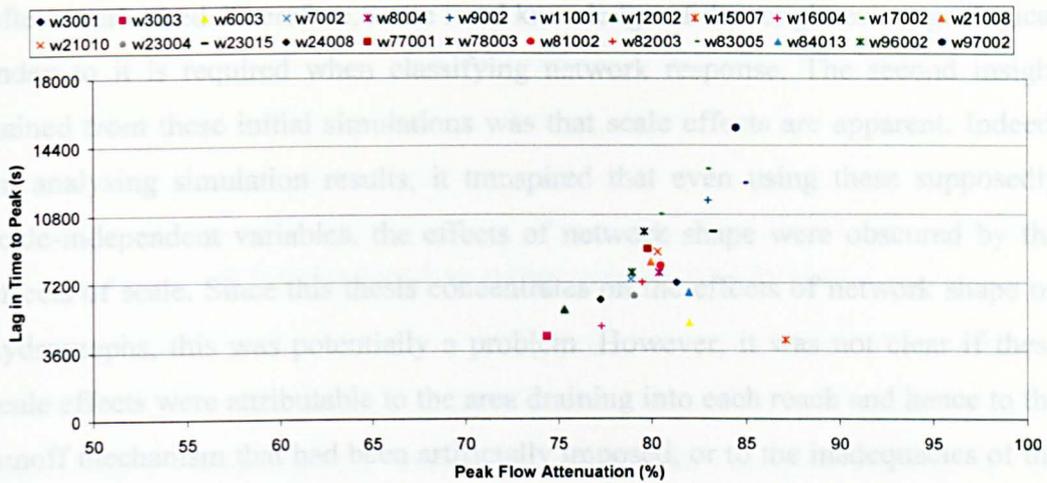


Figure 7.16 Simulated responses for 24 networks of 600km total length

## 7.4 Summary

Initially flow was routed through catchments that lie upstream of the gauging stations in 48 different river basins in Scotland. A long duration (24-hour) design storm was applied uniformly over the catchment area and a very crude representation of runoff in the form of a runoff ratio was applied. In these simulations, the bare minimum knowledge of conveyance parameters was assumed; simply that Manning's  $n$  was equally likely to fall between the maximum and minimum values reported in the literature, and that the scaling exponent of top width fell within the 95 percentile range identified in the previous chapter for Scotland. The resulting hydrograph at the outlet of the river network was characterised using difference between the time to peak of the storm hydrograph and the hydrograph peak, which was termed the 'lag in time to peak', and the percentage attenuation of the flood peak. In an attempt to make these scale-independent measures, and in accordance with previous studies on large-scale network hydrological responses, they were divided by catchment area. The simulations served to highlight two points. Firstly, it highlighted that the conveyance parameters do play a dominant role in shaping the flood hydrograph. This may seem a trivial finding to a hydraulic engineer used to calibrating hydraulic models in small reaches, but, in studies that look at entire networks, it is

often overlooked. Therefore, some local knowledge of the roughness or geological index to it is required when classifying network response. The second insight gained from these initial simulations was that scale effects are apparent. Indeed, on analysing simulation results, it transpired that even using these supposedly scale-independent variables, the effects of network shape were obscured by the effects of scale. Since this thesis concentrates on the effects of network shape on hydrographs, this was potentially a problem. However, it was not clear if these scale effects were attributable to the area draining into each reach and hence to the runoff mechanism that had been artificially imposed, or to the inadequacies of the variables chosen to characterise the hydrograph. The only way to truly determine the effects of network shape on the response was firstly to divorce it from the drainage area then determine whether the variables were indeed scale-dependent. On analysing 8 networks in detail, it was found that the variables were scale-dependent. However, the degree of non-linearity in the scaling of flood peaks was a function of storm duration: near-linear scaling for long duration storms and power law scaling with an exponent of 0.75 for short duration storms. This is another important result that has been overlooked in previous studies. It has two consequences. Firstly, when analysing real hydrological data on how a flood wave is modified as it moves through a large network, it should be perceived as a function not only of in stream variables but also storm duration. Secondly, all previous theoretical studies on the response of large networks impose a short duration storm, in fact usually instantaneous. The generality of their results for real river basins and longer duration storm has then to be questioned. In the remaining simulations the problem of scale is circumvented by only comparing catchments of similar total lengths, in which case simply the difference between the time to peak of the storm hyetograph and the hydrograph peak and the percentage attenuation of the flood peak could be used. The resulting simulations gave synthetic hydrographs for Scottish networks at two scales (300 and 600-kilometre total channel length) with which to characterise specifically the effects of network shape on the hydrograph response.

## Chapter 8

### 8 Statistical multivariate analysis

This chapter describes the statistical multivariate analysis leading to the identification of the best classifiers for surface flow through river networks. As outlined in Chapter 3, the first step in this analysis consists in clustering the simulated network flow routing responses derived in Chapter 7. In doing this, the aim is to form groups (clusters) of networks, which respond similarly to a uniform pulse of runoff. The first section of this chapter describes the derivation of these network response clusters. In the following section, a discriminant analysis, aiming at identifying the network descriptor variables that are best able to reproduce the cluster structure, is conducted. This involves selection and derivation of variables describing the network structure. The contribution to discrimination of these potential classifiers is then tested.

#### 8.1 Cluster analysis of simulated network responses

##### 8.1.1 Clustering method

Cluster analysis techniques are used for grouping objects according to their similarity when no *a priori* division of the objects into categories is available. A large number of clustering techniques have been developed due to the fact that appropriate definition of a cluster varies according to the type and field of application. Cluster analysis techniques are commonly divided into two types; 'hierarchical' and 'optimisation' methods. Hierarchical cluster analysis methods form clusters by hierarchically grouping sub-clusters ('agglomerative' hierarchical methods) or splitting parent clusters ('divisive' hierarchical methods) according to

their similarity or dissimilarity respectively. Divisive and agglomerative methods are clearly complementary and the majority of the published applications of hierarchical cluster analysis techniques use agglomerative methods (Hand, 1981). The main differences between the various hierarchical agglomerative techniques are in the different ways of deciding which pair of clusters should next be merged. The decision is a function of the different ways of measuring inter-cluster distances (similarities). The way the distance between two clusters is calculated is called the 'linkage' method. For example, in the 'nearest neighbour' (also called 'single') linkage method, the distance between two clusters is the distance between the closest points, one from each of the two clusters, whereas it is the distance between the centroids of the clusters in the 'centroid' linkage method. Moreover, the measure of distance can also differ from application to application, although the most commonly used measure is the Euclidian distance (Everitt and Dunn, 1991). The results of a hierarchical cluster analysis are commonly represented by means of tree graphs called 'dendrograms' which provides a scale by which to measure the similarity between the merged clusters at each hierarchical step. The difference between hierarchical and optimisation clustering approaches is that in hierarchical methods, once assigned to a cluster, a point cannot be transferred to another cluster. With optimisation methods, points can be transferred between clusters in an attempt to optimise a given clustering criteria. Optimisation clustering techniques differ in the criteria and the searching algorithm that they use.

Hand's practical suggestion is that when exploring data for possible clusters, different techniques should be tried (Hand, 1981). Indeed, most clustering analysis carried out in the field of hydrology, in the context of regionalisation studies, have incorporated and compared various clustering techniques (e.g.: Acreman and Sinclair, 1986; Nathan and MacMahon, 1990; Burn and Boorman, 1993). Moreover, studies such as the ones of Acreman and Sinclair (1986) on Scottish catchments and Nathan and MacMahon (1990) on catchments located in south-eastern Australia suggest that the most appropriate partition may be identified by using a combination of clustering techniques, applying hierarchical techniques to obtain a first partition which can then be improved by applying optimisation approaches. Hierarchical methods provide a first visual screening of

the data through the dendrograms and optimisation approaches refine the initial partition by reallocating individual points. The appeal of this approach is reinforced by the fact that optimisation techniques have been shown to perform better when a good initial partition is provided (Milligan, 1980; Acreman and Sinclair, 1986).

The clustering method adopted here consists of combining hierarchical and optimisation methods available in the MINITAB statistical analysis commercial package. An initial partition is defined by comparing the dendrograms obtained by different hierarchical clustering linkage methods. The similarity measure most commonly used is the Euclidian distance (Acreman and Sinclair, 1986). Defining a partition by applying hierarchical methods requires a decision to be made about the desired number of clusters, which corresponds to 'cutting' the dendrogram at a certain step. Measurements of similarity level at each step are used to make a decision about the most appropriate step at which to cut the dendrogram. The similarity level at any step is the percentage of the minimum distance at that step relative to the maximum inter-object distance in the data. The dendrogram is cut at the step where an abrupt change in similarity level values is observed (Figure 8.1).

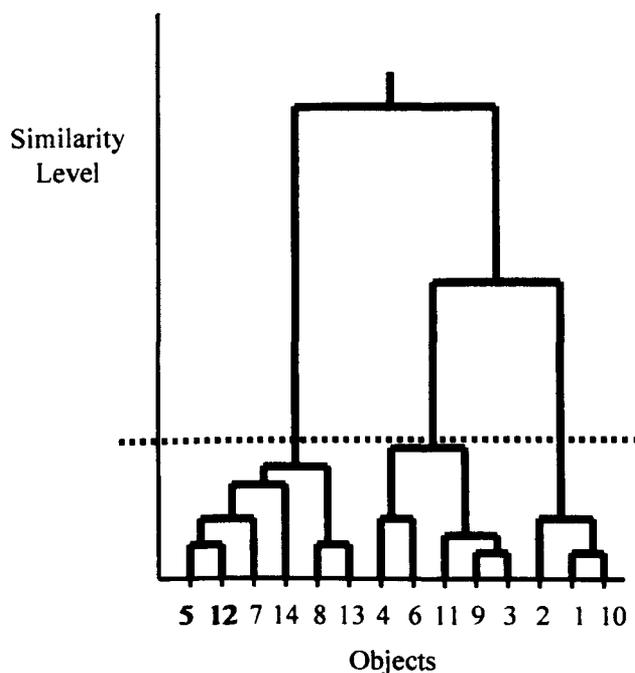


Figure 8.1 Dendrogram suggesting an optimised partition of 3 groups

Choice of the most appropriate linkage method is made according to the dendrogram that shows the clearest abrupt change in similarity level. An initial partition is thus obtained using a hierarchical approach. This partition is then improved applying an optimisation clustering approach. The application of optimisation clustering techniques requires initialisation of the procedure using an arbitrarily predefined partition or number of clusters. Here, the initial partition generated through the hierarchical approach is used to initialise the optimisation procedure. The optimisation technique used is the K-means approach, which was used by Burn and Boorman (1993) to classify UK catchments based on their similarity in flow response. This method applies MacQueen's algorithm (MacQueen, 1967), which evaluates each object of the predefined partition, moving it to the nearest cluster. The nearest cluster is the one that has the smallest distance between the object and the centroid of the cluster. When a cluster changes, by losing or gaining an object, the cluster centroid is recalculated. The process is repeated until no more objects can be moved into a different cluster.

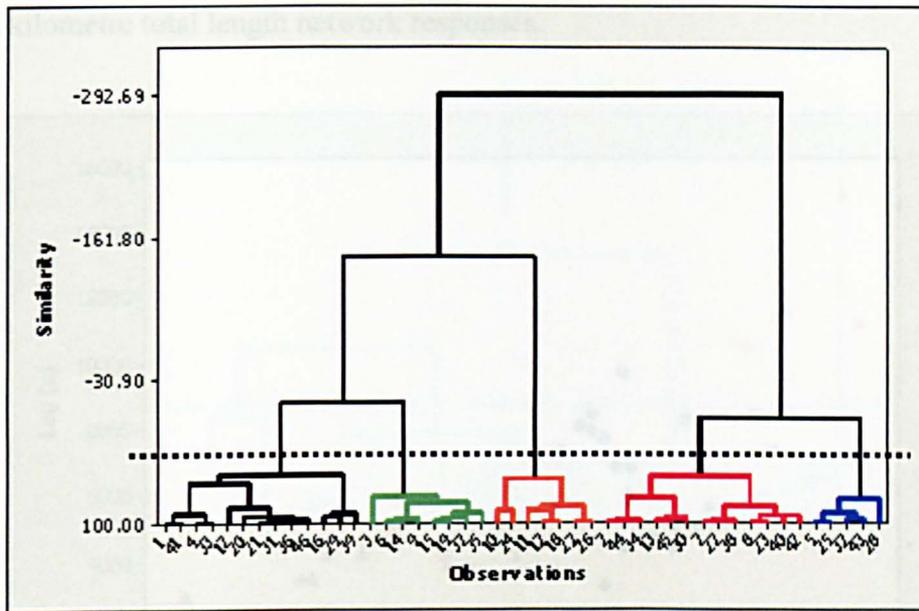
The final partition is therefore obtained by a combination of clustering approaches. The choice of the number of clusters is defined by the hierarchical approach, which generates an initial partition. This initial partition is improved by applying the K-means optimisation clustering method. Results of the application of this procedure for clustering the simulated network flow routing responses at both 300 and 600-kilometre network total length scales are presented in the following sections. The variables used to describe each network response are the lag in time to peak and peak flow percentage of attenuation of the network outlet response hydrographs. These variables being in different units, they are standardised by subtracting the mean and dividing by the standard deviation.

### **8.1.2 Network response clusters at 300-kilometre total length scale**

The simulated hydrographs at the outlet of the 46 networks of 300-kilometre total length were initially clustered using hierarchical agglomerative procedures. Six different linkage methods were tried - Single, Complete, Average, Median, Centroid, McQuitty's and Ward's methods – all of these are outlined in detail in Hand (1981). Visual comparison of the similarity level patterns on the

dendrograms produced by each method, as recommended (Hand, 1981), revealed that Ward's linkage method yielded the greatest similarity within clusters and the best differentiation between them (Figure 8.2a). In Ward's linkage method, the distance between two clusters is the sum of the squared deviations from points to the joint cluster centroids minus the sum of the squared deviations from points to their individual cluster centroids. The objective of this method is to minimise the within-cluster sum of squares. A scatter plot of the network responses showing the obtained initial partition is given in Figure 8.2b.

(a)



(b)

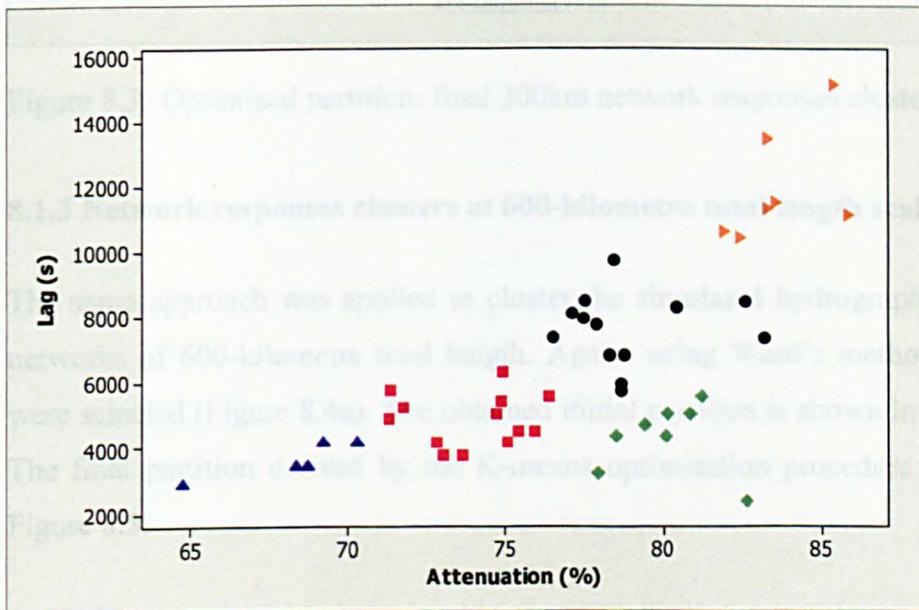


Figure 8.2 Hierarchical agglomerative clustering of the hydrographs for the 46 networks of 300km total length: (a) Dendrogram using Ward's linkage method; (b) Initial partition.

The improved partition obtained when the K-means optimisation clustering procedure is applied to the data, using the initial partition to initialise the procedure is shown in Figure 8.3. This partition, which is very similar to the initial one - only two points situated close to a boundary between two groups have

changed membership – is retained as the final cluster structure for the 300-kilometre total length network responses.

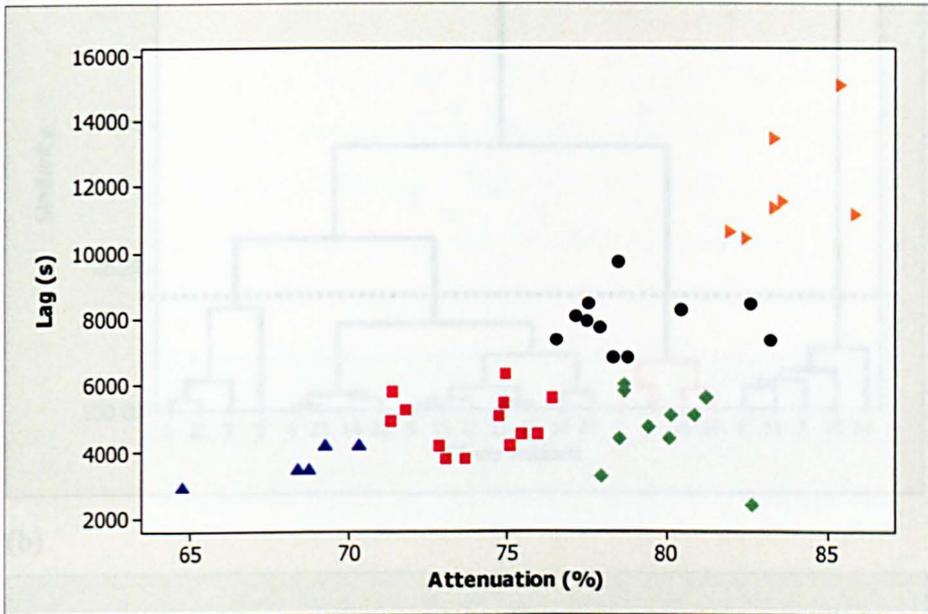


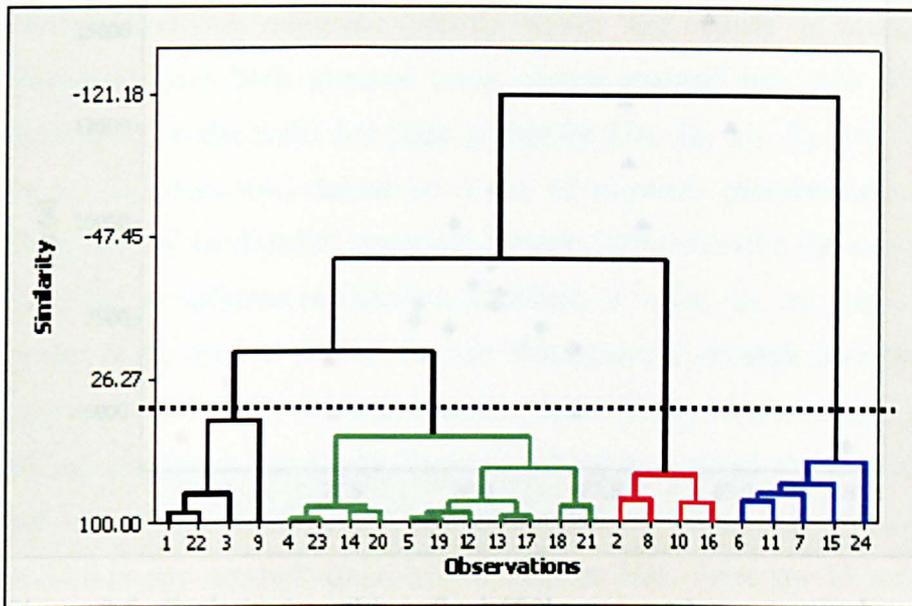
Figure 8.3 Optimised partition; final 300km network responses clusters.

### 8.1.3 Network responses clusters at 600-kilometre total length scale

The same approach was applied to cluster the simulated hydrographs for the 24 networks of 600-kilometre total length. Again, using Ward's method, 4 clusters were selected (Figure 8.4a). The obtained initial partition is shown in Figure 8.4b. The final partition derived by the K-means optimisation procedure is shown in Figure 8.5.

Figure 8.4 Hierarchical agglomerative clustering of the hydrographs for the 24 networks of 600km total length: (a) Dendrogram using Ward's linkage method; (b) Initial partition.

(a)



(b)

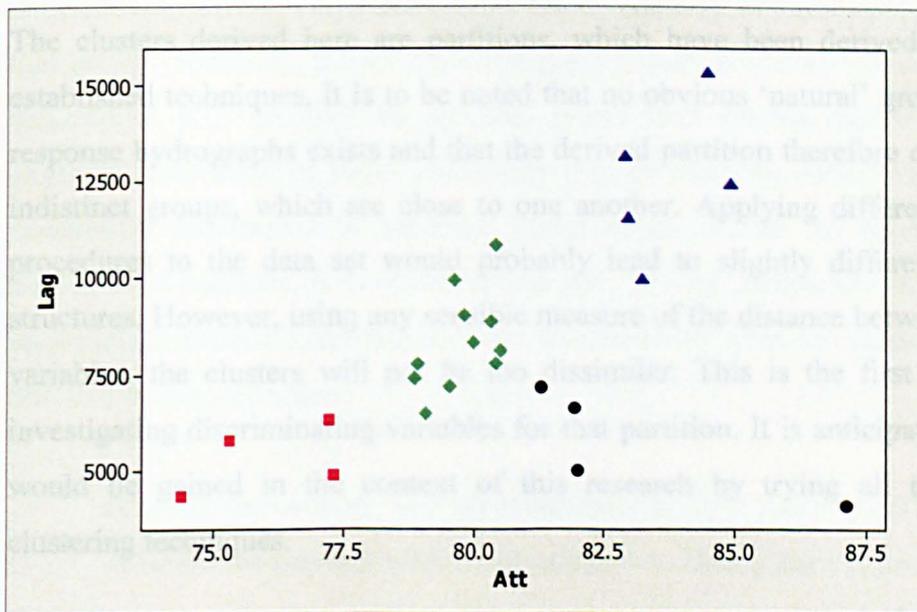


Figure 8.4 Hierarchical agglomerative clustering of the hydrographs for the 24 networks of 600km total length: (a) Dendrogram using Ward's linkage method; (b) Initial partition.

### 8.2.1 Method

The generic statistical term "discrimination" refers to the process of deriving classification rules from samples of already classified objects (Hand, 1981). The objective of a classification rule, also called "decision rule", is to minimise the probability of making an error in classifying an object based on that rule.

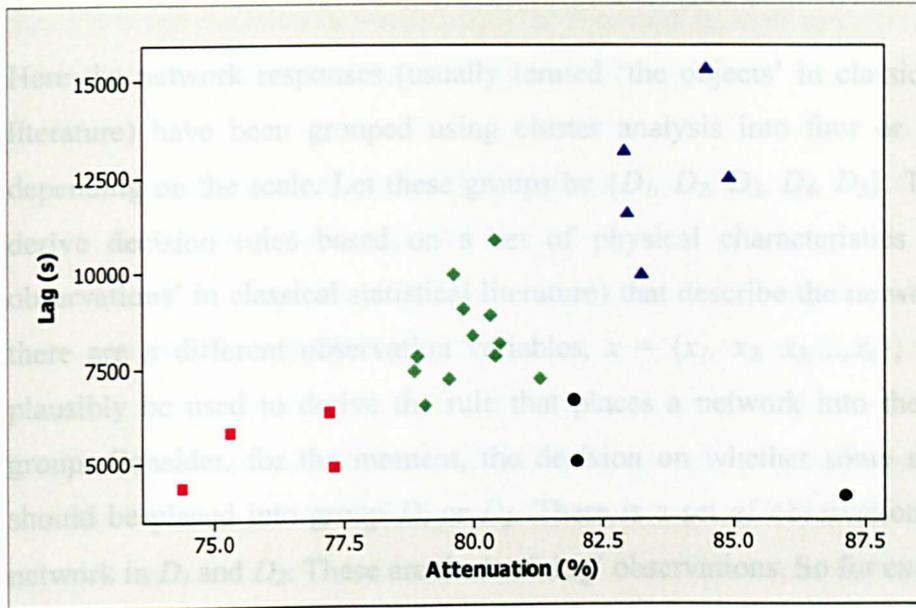


Figure 8.5 Optimised partition; final 600km network responses clusters.

The clusters derived here are partitions, which have been derived using well-established techniques. It is to be noted that no obvious ‘natural’ grouping of the response hydrographs exists and that the derived partition therefore creates rather indistinct groups, which are close to one another. Applying different clustering procedures to the data set would probably lead to slightly different clustering structures. However, using any sensible measure of the distance between response variables, the clusters will not be too dissimilar. This is the first step toward investigating discriminating variables for that partition. It is anticipated that little would be gained in the context of this research by trying all the available clustering techniques.

## 8.2 Discriminant analysis

### 8.2.1 Method

The generic statistical term ‘discrimination’ refers to the process of deriving classification rules from samples of already classified objects (Hand, 1981). The objective of a classification rule, also called ‘decision rule’, is to minimise the probability of making an error in classifying an object based on that rule.

Here the network responses (usually termed ‘the objects’ in classical statistical literature) have been grouped using cluster analysis into four or five groups, depending on the scale. Let these groups be  $\{D_1, D_2, D_3, D_4, D_5\}$ . The aim is to derive decision rules based on a set of physical characteristics (called ‘the observations’ in classical statistical literature) that describe the network. Suppose there are  $n$  different observation variables,  $x = \{x_1, x_2, x_3, \dots, x_n\}$ , which could plausibly be used to derive the rule that places a network into the appropriate group. Consider, for the moment, the decision on whether some new network should be placed into group  $D_1$  or  $D_2$ . There is a set of observations  $x$  for each network in  $D_1$  and  $D_2$ . These are the ‘training’ observations. So for example, in the 300-kilometre network classification (Figure 8.3), there are 13 networks in the group coloured with red squares, each of which has a set of observations. Thus it is possible to derive, empirically, a probability density function (pdf) for  $x$  within that group. Similarly, for all the other groups, pdfs could be derived. Let these conditional pdfs be  $f_i(x)$ . In practice, rather than relying completely on the empirical derivation of these probability distributions, an underlying parametric probability distribution is assumed. Here, as in all but the most specialised discriminant analysis, it is assumed that the  $f_i(x)$  are adequately modelled by a multivariate normal distribution. Now an intuitively obvious classification rule falls out of this definition:

Allocate the network with observations  $x$  to  $D_1$  if  $f_1(x) > f_2(x)$

Allocate the network with observations  $x$  to  $D_2$  if  $f_2(x) > f_1(x)$

It can be shown (e.g. Hand, 1981) that under the assumption that the  $f_i(x)$  are multivariate normals, this rule can be conveniently written in terms of the means of the observed  $x$ s in the groups,  $\mu_1$  and  $\mu_2$ , and the covariance matrix  $S$  for observed  $x$  in the pair of groups. The rule then becomes, allocate the network with observations  $x$  to group  $D_1$  if,

$$h_{12}(x) = (\mu_1 - \mu_2)' S^{-1} \left( x - \frac{1}{2}(\mu_1 + \mu_2) \right) > 0. \quad (8.1)$$

Thus, when a decision between which of 5 groups is most appropriate has to be made, the rules are, allocate to,

$$\begin{aligned}
 &D_1 \text{ if } h_{1j}(x) > 0 \text{ and } h_{ij}(x) < 0 \text{ for } i, j \neq 1 \\
 &D_2 \text{ if } h_{2j}(x) > 0 \text{ and } h_{ij}(x) < 0 \text{ for } i, j \neq 2 \\
 &\quad \vdots \\
 &D_5 \text{ if } h_{5j}(x) > 0 \text{ and } h_{ij}(x) < 0 \text{ for } i, j \neq 5
 \end{aligned} \tag{8.2}$$

The particular discriminant analysis outlined above is called a ‘linear discriminant analysis’ because on multiplying out the matrix equation 8.1, the  $h_{ij}(x)$  become linear functions in  $x$ . These  $h_{ij}(x)$  are also called ‘linear discriminant functions’. Although this arises from assuming multivariate normality within each group, Fisher (1936) derived the same rules based on defining a dummy variable, which was simply a linear combination of the observation variables, and looking for the combination that would separate the groups as much as possible. Fisher found that the coefficients of the linear equations that maximised group separation were the same as the coefficients defined by the linear discriminant functions described above. Consequently, one might expect the discrimination rules given in (8.2) to perform reasonably well even when the assumption of normality is not wholly justified (Everitt and Dunn, 1991). It is of some interest to note that the coefficients of the linear discriminant functions can be derived by regression (Everitt and Dunn, 1991).

Because of their analytical tractability, linear discriminant analyses are widely used (Hand, 1981). In the hydrology field, Bates et al (1998) used a linear discriminant analysis approach to study the discriminating role of basin characteristics for regional floods in southeastern Australia.

The relative importance of the discriminant function coefficients indicates the relative contribution of each variable to the discrimination. In order to evaluate the performance of the discriminant function, the group membership allocated by discrimination is compared with the true membership. A ‘classification index’, corresponding to the percentage of correctly classified objects, is calculated. A cross-validation procedure is used in order to avoid optimistic estimation of this

index due to the fact that the same data is used for deriving the discriminant functions and allocating the objects to a group. This procedure consists in omitting each object one at a time, recalculating the discriminant functions using the remaining data, and then classifying the omitted object.

In this thesis, the training set is the set of river networks clustered according to their similarity in hydrographs produced by a pulse of runoff. The observations made on these classified objects are variables describing the network structure. These variables are potential classifiers whose relative contribution to discrimination is tested. The selection and calculation of these variables are described in the following section.

### **8.2.2 Network descriptors**

The aim here is to assess potential classifiers and select the most appropriate for describing network response. The obvious place to start is with a review of the classifying variables used in the regionalisation studies that were described in Chapter 2. However, these studies try to capture the entire hydrological response of a river basin, whereas here the focus is on the flow routing through river networks. Therefore, only a subset of these catchment descriptors, which are relevant to network routing, is selected as a potential set of classifiers for network response. Many of these traditionally used descriptors have their roots in the analysis of extreme flow events that were conducted in an era before the wide spread use of computers and digital maps. In the last two decades, Geographical Information Systems and the increase in computational power on personal computers mean that a range of alternative classifiers, that were previously deemed too difficult to calculate, can be investigated. These derive from studies like the classical work of Shreve (1966) or the more recent work of Rodriguez-Iturbe and Rinaldo (1997), which aim at characterising the physical structure of networks, without necessarily relating it to flood response. Therefore, the potential for these to be used as alternative new classifiers is reviewed. Finally, a set of completely new variables, unique to this thesis, is introduced. These variables are based on some very basic observations on the physics of fluid flow in networks.

### *Classifying variables used in the past*

Variables relevant to network routing that have been extensively used as classifiers include: the 'main stream length', which is defined as the longest flow path from a source to the outlet; the 'main stream slope', which is defined as the slope of the central 75% of main stream; the 'stream frequency', which corresponds to the number of stream junctions per unit of drainage area; and the 'stream density', which corresponds to the length of channel per unit of drainage area. These classifiers have been used in regionalisation studies by Acreman and Sinclair (1986), Nathan and McMahon (1990), Burn and Boorman (1992) and are still used in more recent studies such as the ones conducted by Bates et al (1998), Hall and Minns (1999) and Calver et al (2001). Other descriptors commonly used in regionalisation studies are the catchment 'relief', equal to the maximum elevation difference in the basin (Bates et al, 1998; Wolock et al, 2003) and a variable describing the 'shape' of the catchment, which is defined in various ways according to whether it measures the catchment elongation (Post and Jakeman, 1996), rotundity (Bates et al, 1998) or elliptical properties (Moussa, 2003).

Here, the variables used as possible descriptors of the way a network modifies a flood wave have been selected because it is perceived that they might be able to classify network structure in isolation. Variables that describe the geometry of the network in relation to other characteristics of the basin, such as its drainage area, are adapted, when possible, into similar variables that capture the network structure in isolation. Moreover, and importantly, the selected variables need to be easily derived from digital network data.

Reviewing the literature (mainly Acreman and Sinclair, 1986; Nathan and MacMahon, 1990; Burn and Boorman, 1992) led to seven previously used classifiers being considered. These, which will be referred to as 'standard' network descriptor variables, are:

- NUM\_Link; the number of links. Links are the segments of channel between two successive nodes in the network. This variable is equivalent

to the commonly used ‘stream frequency’ variable in the context of networks draining the same catchment area.

- AV\_LLength; the average link length.
- AV\_LSlope; the average link slope.
- MAX\_Dist; the commonly called ‘main stream length’ corresponding to the maximum distance of any point on the network to the outlet.
- DROP\_MaxD; the difference in elevation between the point of maximum distance to the outlet and the outlet. This variable is an adaptation of the traditional field measurement of the slope of the central 75% of main stream, which it is impractical to calculate from synthetic data.
- MAX\_Drop; the maximum elevation difference between any point on the network and the outlet. This variable is an adaptation, focused on the network structure in isolation from its drainage area, of the traditional ‘relief’ catchment descriptor, which is defined as the maximum elevation difference between any point in the basin and the outlet.
- NUM\_Source; the number of extremities in the network. This variable is a proxy for the commonly used ‘shape’ catchment descriptor, traditionally defined in relation to the catchment perimeter.

### ***Extending the set of possible classifiers***

With the objective of extending the possible classifiers to include variables that have not been used in the past for river basin classification but that do describe the network geometry, a review of the geomorphologically based hydrological models was conducted. There exists a suite of models that attempt to explicitly incorporate a description of the network structure. This is typified by the analytic approach developed by Rodriguez-Iturbe and Valdes (1979), who interpreted the response of a network to an instantaneous unit volume of runoff uniformly distributed in space as the travel time distribution of the particles of water to the basin outlet. The application of this model, which they called the ‘geomorphologic instantaneous unit hydrograph’ (GIUH), requires ordering the streams constituting the network according to the Strahler ordering system (Strahler, 1957) defined as follows: exterior links have order 1; where two or

more links of order  $m_1, m_2, m_3 \dots$  join, with  $m_1 \geq m_2 \geq m_3 \dots$ , the order of the next downstream link is the greater of  $m_1$  and  $m_2 + 1$ . A 'Strahler stream' is defined as a sequence of links of the same order and the 'order of the network' as the maximum order identified in the network. The determination of the GIUH is accomplished by analysing the detailed motion of water particles in space according to the following rule: the only possible transitions out of a link of order  $m_i$  are of the type  $m_i$  to  $m_j$  for  $j > i$ . This rule defines a collection of routes that a particle may follow to the basin outlet. The number of Strahler streams of a given order divided by the number of Strahler streams of the following order, called the 'bifurcation ratio', was first defined by Horton (1945) who observed it to be approximately constant through semilog plots of number of Strahler streams against order. Horton's bifurcation ratio and other similar laws have led to the development of the theory of the fractal structure of river networks and the study of the impact of such fractal characters on the GIUH in synthetic Hortonian networks (Rinaldo et al, 1991; Rodriguez-Iturbe and Rinaldo, 1997). Such theories suggest that both network order and bifurcation are instrumental in describing network hydrologic response. Although the real networks analysed in this thesis have similar total length, they might not be of the same order. Moreover, they are certainly not strictly Hortonian. A Hortonian network shows a constant bifurcation ratio with order. The networks analysed here show a bifurcation ratio that varies with order in the range of a couple of units. Therefore, variables incorporating the network order and the average bifurcation ratio are added to the selection of potential classifiers to be tested. These variables are:

- MAX\_Order; the maximum order identified in the network according to the Strahler ordering system.
- AV\_Bif; the network average Horton bifurcation ratio.

Rodriguez-Iturbe and Valdes GIUH model (1979) has also been formulated in term of the 'width function'. The width function, first introduced by Shreve (1969), is defined as the number of links in the network at a given flow distance from the outlet. The first linkage of network hydrologic response with the width function came from Kirkby (1976), who argued that, for spatially uniform inputs, the maximum flood was proportional to the maximum of the width function. The concept was later refined and combined to the travel time interpretation of the

GIUH (Mesa and Mifflin, 1988; Rodriguez-Iturbe and Rinaldo, 1997). Two variables based on the width function were added to the list of potential classifiers. These are:

- MAX\_Width; the maximum of the width function.
- Dist-MaxW; the distance from the outlet at the maximum of the width function.

It is of interest to note that, in a recent study on continuous flood simulation, Calver et al (2001) incorporated variables based on the 'width function' in regionalising their model parameters to extend application to ungauged basins.

### *New potential classifiers*

Hydrological models based on a network width function have proved popular and successful in modelling hydrographs, especially of large river basins. For example, Naden (1992) successfully applied such a model to the whole of the River Thames. This might indicate that the width function in itself is a good way of classifying the response of a network. However, the models are laden with a raft of additional assumptions that provide a suite of free parameters, all of which can be calibrated and mask the true influence of the width function. In the discriminant analysis below, the classifying power of the width function is tested. Clearly however, the entire distribution of distances to the outlet cannot be used. Therefore, the distribution is described using its moments. Here, the first four moments are used:

- MEAN\_Dist; the mean of the distance to the outlet distribution.
- STD\_Dist; the standard deviation of the distance to the outlet distribution.
- KURT\_Dist; the kurtosis of the distance to the outlet distribution.
- SKEW\_Dist; the skewness of the distance to the outlet distribution.

The width function takes no account of typical slopes in the network. This may not matter for very large (regional to continental scale) river networks, where models based on it are increasingly being applied (Abdulla and Lettenmaier,

1997). However, at a regional scale, the elevation drop, particularly in the headwaters of a river network, must play a role in shaping the hydrograph. Therefore an attempt has been made to characterise the elevation drop from any point in the network. This is achieved using a similar approach to the width function in that a distribution is collated from every reach in the network and the outlet. Again, the whole distribution cannot be used as a classifier, so its first four moments are used:

- MEAN\_Drop; the mean of the elevation drop to the outlet distribution.
- STD\_Drop; the standard deviation of the elevation drop to the outlet distribution.
- KURT\_Drop; the kurtosis of the elevation drop to the outlet distribution.
- SKEW\_Drop; the skewness of the elevation drop to the outlet distribution.

Finally, energy concepts play such a dominant role in hydraulics it is perhaps surprising that they have remained on the periphery of hydrological modelling. Channel flow in large networks clearly adheres to the same fundamental hydraulic principles that can be expressed in terms of conservation of energy, and yet no researchers use an energy based classification scheme. Describing the potential kinetic and heat energy at any instant in time for an entire network would require a fully distributed description of the network. Here, indices are derived that are intended to measure both the potential energy and likelihood of energy being dissipated into heat. It is assumed that the potential energy in any reach is proportional to its elevation above the outlet of the network. Clearly, this ignores the mass of the water stored in the reach. However incorporating this level of detail into an index would make it impossible to derive from most digital databases of river networks. Further, it is assumed that the capacity to dissipate the potential energy into heat through friction is a function of the distance travelled through the network. Thus, the energy index is defined by the ratio of elevation drop to distance downstream to the outlet. A distribution of this index for the entire network can be derived by considering each reach that comprises it in turn. High reaches that are close to the outlet will have a high value of the

index, whereas low reaches that are distant from the outlet will have a low value for the index. Thus, in essence, the index combines the network width function with the elevation distribution giving a three dimensional picture of the network structure. Again, as with the previous distributions, it is impractical to consider the whole distribution as a classifier. Therefore, the first four moments are used as well as the maximum of the distribution:

- MEAN\_Eng; the mean of the energy index distribution.
- STD\_Eng; the standard deviation of the energy index distribution.
- KURT\_Eng; the kurtosis of the energy index distribution.
- SKEW\_Eng; the skewness of the energy index distribution.
- MAX\_Eng; the maximum of the energy index distribution

A set of 24 variables suitable for the discriminant analysis is therefore defined. A set of algorithms was written in Arc Macro Language (AML) to interrogate the digital databases and derive all the variables above defined. Values for these 24 network descriptor variables for the 300 and 600-kilometre total length networks are given in Appendix B.

### **8.2.3 Discriminant analysis at 300-kilometre scale**

The objective is to find the variable or combination of variables, amongst the 24 variables available for discriminant analysis, which are best able to reproduce the network response clusters. The number of possible combinations of these variables is too high for considering applying a trial and error analysis. Therefore a strategy for analysis is defined.

#### ***Definition of a strategy for analysis***

The analysis is conducted in two steps. In the first one, a reduced set comprising the 7 'standard' descriptor variables described in section 8.2.2 is tested. The aim of this first step is to investigate to what extent the variables commonly used as classifiers are able to predict the network response clusters and which of these variables contribute most to the discrimination. The approach consists in applying the discriminant analysis described in section 8.2.1 to the set of variables and

observing the variation of the classification index when this set of variables is gradually reduced. Variables are discarded from the set starting from the variable that shows the lowest discriminant coefficient. An abrupt decrease in the classification index indicates the optimum set of variables. Hence a set of 'best standard classifiers' is defined. In a second step, the other 17 variables are pooled together with the best standard classifier variables and the analysis is repeated in order to determine the overall set of 'best classifiers'.

Prior to discriminant analysis, a correlation analysis including the 24 candidate variables is conducted in order to measure the degree of linear relationship between each pair of variables in order to discard unnecessary variables. For each pair of variables, Pearson's product moment correlation coefficients and associated p-values for the hypothesis test of the correlation coefficient being zero are calculated. If one variable tends to increase as the other decreases, the correlation coefficient is negative. Conversely, if the two variables tend to increase together the correlation coefficient is positive. Two variables are considered to be highly linearly correlated when the p-value is lower than 0.0005, in which case one of the two variables is discarded.

The correlation matrix corresponding to the 24 potential classifier variables calculated for the 300-kilometre networks is given in Table 8.1.

	NUM_Link	AV_LLength	AV_LSlope	MAX_Dist	Drop_MaxD	Dist_MaxW	MAX_Width	MEAN_Dist	STD_Dist	KURT_Dist	SKEW_Dist	MEAN_Drop	STD_Drop
<b>AV_LLength</b>	-0.862 0.000												
<b>AV_LSlope</b>	0.501 0.000	-0.439 0.002											
<b>MAX_Dist</b>	-0.470 0.001	0.543 0.000	-0.262 0.079										
<b>Drop_MaxD</b>	0.296 0.046	-0.247 0.098	0.648 0.000	0.172 0.252									
<b>Dist_MaxW</b>	-0.131 0.384	0.088 0.563	-0.040 0.790	0.579 0.000	0.077 0.611								
<b>MAX_Width</b>	0.530 0.000	-0.424 0.003	0.264 0.076	-0.720 0.000	-0.042 0.784	-0.473 0.001							
<b>MEAN_Dist</b>	-0.420 0.004	0.414 0.004	-0.243 0.103	0.917 0.000	0.135 0.371	0.749 0.000	-0.690 0.000						
<b>STD_Dist</b>	-0.408 0.005	0.515 0.000	-0.182 0.227	0.967 0.000	0.218 0.146	0.565 0.000	-0.748 0.000	0.885 0.000					
<b>KURT_Dist</b>	-0.059 0.698	-0.028 0.856	-0.187 0.214	-0.228 0.128	-0.212 0.157	-0.201 0.180	0.477 0.001	-0.103 0.495	-0.404 0.005				
<b>SKEW_Dist</b>	-0.172 0.252	0.321 0.030	-0.050 0.742	0.296 0.046	0.156 0.301	-0.321 0.029	-0.295 0.046	-0.074 0.623	0.328 0.026	-0.495 0.000			
<b>MEAN_Drop</b>	0.432 0.003	-0.362 0.013	0.693 0.000	-0.033 0.827	0.821 0.000	-0.036 0.812	0.159 0.290	-0.054 0.721	0.022 0.885	-0.135 0.371	0.097 0.521		
<b>STD_Drop</b>	0.305 0.039	-0.261 0.080	0.859 0.000	-0.017 0.912	0.775 0.000	-0.054 0.722	0.143 0.345	-0.043 0.776	0.043 0.778	-0.144 0.340	0.075 0.619	0.812 0.000	
<b>KURT_Drop</b>	0.116 0.442	-0.119 0.431	-0.267 0.072	-0.205 0.172	-0.250 0.094	-0.168 0.265	0.189 0.209	-0.144 0.340	-0.220 0.141	0.234 0.117	-0.183 0.223	-0.244 0.102	-0.292 0.049

Table 8.1 Correlation matrix for the 300km network descriptor variables (upper number = correlation coefficient; lower number = p-value)

	NUM_Link	AV_LLength	AV_LSlope	MAX_Dist	Drop_MaxD	Dist_MaxW	MAX_Width	MEAN_Dist	STD_Dist	KURT_Dist	SKEW_Dist	MEAN_Drop	STD_Drop
<b>SKEW_Drop</b>	-0.131	0.060	-0.149	-0.100	-0.213	-0.121	0.086	-0.041	-0.141	0.297	-0.192	-0.447	-0.134
	0.386	0.690	0.324	0.508	0.156	0.422	0.568	0.786	0.349	0.045	0.201	0.002	0.376
<b>MAX_Drop</b>	0.372	-0.321	0.826	-0.068	0.795	-0.097	0.171	-0.081	-0.011	-0.080	0.029	0.846	0.933
	0.011	0.030	0.000	0.654	0.000	0.522	0.256	0.593	0.941	0.598	0.848	0.000	0.000
<b>MEAN_Eng</b>	0.564	-0.464	0.749	-0.425	0.623	-0.360	0.405	-0.486	-0.337	-0.172	0.154	0.864	0.700
	0.000	0.001	0.000	0.003	0.000	0.014	0.005	0.001	0.022	0.253	0.308	0.000	0.000
<b>STD_Eng</b>	0.427	-0.337	0.837	-0.262	0.447	-0.100	0.244	-0.303	-0.154	-0.252	0.066	0.552	0.686
	0.003	0.022	0.000	0.079	0.002	0.509	0.103	0.041	0.306	0.091	0.665	0.000	0.000
<b>KURT_Eng</b>	0.211	-0.199	-0.021	-0.096	0.060	0.081	0.210	-0.018	-0.134	0.247	-0.192	0.193	-0.083
	0.159	0.184	0.887	0.527	0.691	0.593	0.161	0.905	0.376	0.098	0.200	0.199	0.583
<b>SKEW_Eng</b>	0.246	-0.233	0.030	-0.047	0.099	0.133	0.158	0.023	-0.071	0.217	-0.175	0.181	-0.058
	0.099	0.119	0.843	0.759	0.513	0.379	0.295	0.879	0.641	0.148	0.245	0.228	0.700
<b>MAX_Eng</b>	0.461	-0.379	0.570	-0.244	0.437	-0.035	0.295	-0.243	-0.184	-0.072	-0.007	0.586	0.458
	0.001	0.009	0.000	0.103	0.002	0.816	0.047	0.103	0.221	0.634	0.961	0.000	0.001
<b>MAX_Order</b>	0.553	-0.560	0.189	-0.305	0.017	-0.015	0.297	-0.232	-0.280	-0.006	-0.163	0.151	0.089
	0.000	0.000	0.209	0.040	0.911	0.921	0.045	0.121	0.060	0.971	0.281	0.317	0.556
<b>AV_Bif</b>	0.037	-0.040	0.195	-0.064	0.173	-0.097	0.020	-0.062	-0.050	0.031	-0.053	0.159	0.172
	0.809	0.790	0.193	0.671	0.251	0.522	0.897	0.682	0.741	0.841	0.729	0.291	0.253
<b>NUM_Source</b>	1.000	-0.860	0.508	-0.474	0.297	-0.132	0.534	-0.425	-0.411	-0.060	-0.171	0.433	0.309
	0.000	0.000	0.000	0.001	0.045	0.382	0.000	0.003	0.005	0.690	0.256	0.003	0.037

Table 8.1 Correlation matrix for the 300km network descriptor variables (continued)

	KURT_Drop	SKEW_Drop	MAX_Drop	MEAN_Eng	STD_Eng	KURT_Eng	SKEW_Eng	MAX_Eng	MAX_Order	AV_Bif
<b>SKEW_Drop</b>	0.727 0.000									
<b>MAX_Drop</b>	-0.050 0.741	-0.031 0.837								
<b>MEAN_Eng</b>	-0.142 0.346	-0.374 0.011	0.746 0.000							
<b>STD_Eng</b>	-0.207 0.168	-0.189 0.209	0.640 0.000	0.717 0.000						
<b>KURT_Eng</b>	0.040 0.792	-0.293 0.048	-0.036 0.814	0.164 0.277	0.156 0.300					
<b>SKEW_Eng</b>	0.068 0.655	-0.213 0.156	0.002 0.988	0.154 0.307	0.243 0.104	0.954 0.000				
<b>MAX_Eng</b>	-0.106 0.482	-0.348 0.018	0.455 0.001	0.674 0.000	0.785 0.000	0.648 0.000	0.680 0.000			
<b>MAX_Order</b>	0.084 0.578	0.015 0.920	0.149 0.324	0.222 0.138	0.123 0.414	0.067 0.659	0.071 0.641	0.089 0.555		
<b>AV_Bif</b>	-0.103 0.498	-0.145 0.336	0.128 0.396	0.180 0.230	0.239 0.109	0.093 0.540	0.090 0.553	0.271 0.069	-0.736 0.000	
<b>NUM_Source</b>	0.117 0.440	-0.130 0.390	0.375 0.010	0.567 0.000	0.434 0.003	0.211 0.159	0.245 0.100	0.468 0.001	0.547 0.000	0.043 0.777

Table 8.1 Correlation matrix for the 300km network descriptor variables (continued)

None of the variables is completely independent of all the others; all show a high degree of linear correlation with at least one other variable. For example, as one might expect, the third and fourth moments of the distributions of distance to the outlet, elevation drop to the outlet and energy index are correlated. The negative correlation between the number of links in the network, NUM\_Link, and the average link length, AV\_LLength, is expected as the networks have the same total length; the higher the number of links, the shorter the links and therefore the lower the average link length. The positive correlation between the first moment of the elevation drop to the outlet distribution, MEAN\_Drop, and the average link slope, AV\_LSlope, again, is expected; a high average elevation drop from any point along the network to the outlet describes a steep network where the average link slope is expected to be high. Many other examples of such correlations could be highlighted and interpreted here. However, whilst on the basis of the correlation analysis it is apparent that many of the variables are correlated, it is not clear at this stage whether two highly correlated variables have the same 'discriminating power'. Thus, rather than arbitrarily discard one variable in favour of another that is highly correlated, both are thrown into the discriminant analysis. If two variables are highly correlated, the one that has the most discriminating power is retained and the other is discarded. The discriminating power of each variable is determined based on the relative importance of the linear discriminant function coefficients described in section 8.2.1; for each variable, a discriminating power corresponding to the average for all groups of the absolute values of the linear discriminant function coefficients is calculated. The higher the discriminating power is, the more the variable contributes to between-group separation. Therefore the derived strategy for selecting the best classifying variables is:

1. Run a discriminant analysis with the entire set of variables to be tested.
2. Isolate the variable that shows the highest discriminating power.
3. Discard the variable or set of variables that are highly correlated (p-value lower than 0.0005) with the 'best discriminating variable' isolated in step 2.
4. Recalculate the discriminating powers for the remaining variables.
5. Repeat steps 2 to 4 until the remaining variables are the least correlated (p-value higher than 0.0005).

6. Calculate the classification index corresponding to discrimination with the 'best discriminating least correlated variables' isolated in step 5.
7. Discard the variable that shows the lowest discriminating power.
8. Recalculate the classification index and discriminating powers for the remaining variables.
9. Repeat step 7 and 8, observing the variations of the classification index in order to identify the optimum set of variables.

### ***Results of the discriminant analysis at 300-kilometre scale***

Table 8.2 summarises the analysis results.

Set	Variables	Classification Index (%)
All Standard Variables		34.8
Best Discriminating least correlated	AV_LSlope, AV_LLength	26.1
discard lowest contribution	AV_LSlope	26.1
Standard Best Standard Classifiers + Others		52.2
Best Discriminating least correlated	MEAN_Eng; MAX_Order; KURT_Dist; MAX_Width; SKEW_Drop; SKEW_Eng; DIST_MaxW	50.0
discard 1 lowest contribution	MEAN_Eng; MAX_Order; KURT_Dist; MAX_Width; SKEW_Drop; SKEW_Eng	45.7
discard 2 lowest contribution	MEAN_Eng; MAX_Order; KURT_Dist; MAX_Width; SKEW_Drop	52.2
discard 3 lowest contribution	MEAN_Eng; MAX_Order; KURT_Dist; MAX_Width	52.2
discard 4 lowest contribution	MEAN_Eng; MAX_Order; KURT_Dist	34.8
discard 5 lowest contribution	MEAN_Eng; MAX_Order	23.9
discard 6 lowest contribution	MEAN_Eng	28.3

Table 8.2 Results of discriminant analysis at 300km scale. The variables are listed in decreasing order of discriminating contribution.

When the 7 standard variables are used for reproducing the network response clusters, only 34.8% of the networks are correctly classified. When this set of variables is reduced to the pair of least correlated variables AV\_LSlope and AV\_LLength, the classification index drops to 26.1%. The same classification performance is obtained when only AV\_LSlope is used. These results suggest that the variables commonly used for characterising river basins hydrology are actually very poor classifiers of networks and, furthermore, that too many highly correlated variables are often used. AV\_LSlope was identified here as the best discriminating standard variable and the other standard variables were discarded for the rest of the analysis.

Turning now to the extended set of 17 'possible classifiers' and 'original' variables devised within this research, these are pooled together with AV\_LSLope, and the same systematic weeding out of highly correlated and poorly discriminating variables was conducted. With all 18 variables, 52.2% of the networks were correctly classified. Thus, just under half of the networks were misclassified. This can rightly be seen as a worrying result in the context of the networks tested in this thesis. The reasons why this level of misclassification occurs are discussed in detail later. However, it should be borne in mind that when the 'standard' classifying variables are used, the misclassification is very much higher. With 18 variables to base the classification scheme on then one might expect a much higher level of classification especially if all variables are linearly independent (i.e. uncorrelated). However, there is a high degree of multicollinearity in the variables, which not only means that a reduced set of classifiers should be found, but that this set can also hamper the performance of the linear discriminant analysis. Moreover, increasing the number of variables used for discrimination does not necessarily increase the classification performance (Hand, 1981). It is seen below that as discriminant variables are discarded from the analysis, the number of networks that are correctly classified can in fact increase. Therefore, first of all, the set of variables is reduced to a set of least correlated variables by discarding the correlated variables that show the lowest discriminating power. Discriminant analysis using these 'best discriminating least correlated variables' leads to a classification performance that is only slightly lower (50%) than when the entire set is used (52.2%), even though the set has been reduced from 18 to 7 variables. The list of these 7 variables is given in Table 8.2. This set is then gradually reduced from 7 to 1, discarding the variable that shows the lowest discriminating contribution at each step. Classification index and discriminating powers are calculated for each step. An abrupt decrease in classification performance is observed when the 4<sup>th</sup> 'lowest contributing' variable is discarded. This indicates that the set of best classifying variables is the one used in the previous step, including MEAN\_Eng, MAX\_Order, KURT\_Dist and MAX\_Width. Further evidence that this is the best set of classifiers is given by the fact that 52.2% of the networks are correctly

classified. This classification performance, obtained using 4 discriminating variables, is the same as the one obtained using 18 potential classifiers.

Therefore the set of variables identified as best classifiers includes, in decreasing order of contribution, MEAN\_Eng, MAX\_Order, KURT\_Dist and MAX\_Width. Thus the discriminant analysis has objectively picked up a set of only four, linearly independent classifiers that pertain directly and intuitively to key features of the shape of networks. None of these variables have been routinely used in previous classification studies. The values for the coefficients of the linear discriminant functions obtained for each group are reported in Table 8.3.

Group	1	2	3	4	5	Discriminating Power
MEAN_Eng	241.1	583.8	560.1	1006.3	100.2	498.3
MAX_Order	33.2	35.5	34.5	34.8	33.3	34.26
KURT_Dist	-20.7	-23.3	-23.1	-23.4	-24.5	23
MAX_Width	3.1	3.7	3.1	3.9	2.6	3.28

Table 8.3 Linear discriminant function coefficients and derived potential classifier's discriminating powers at 300km scale

This set of variables is able to reproduce the clusters with a performance of 52.2%. This is illustrated in Figure 8.6.

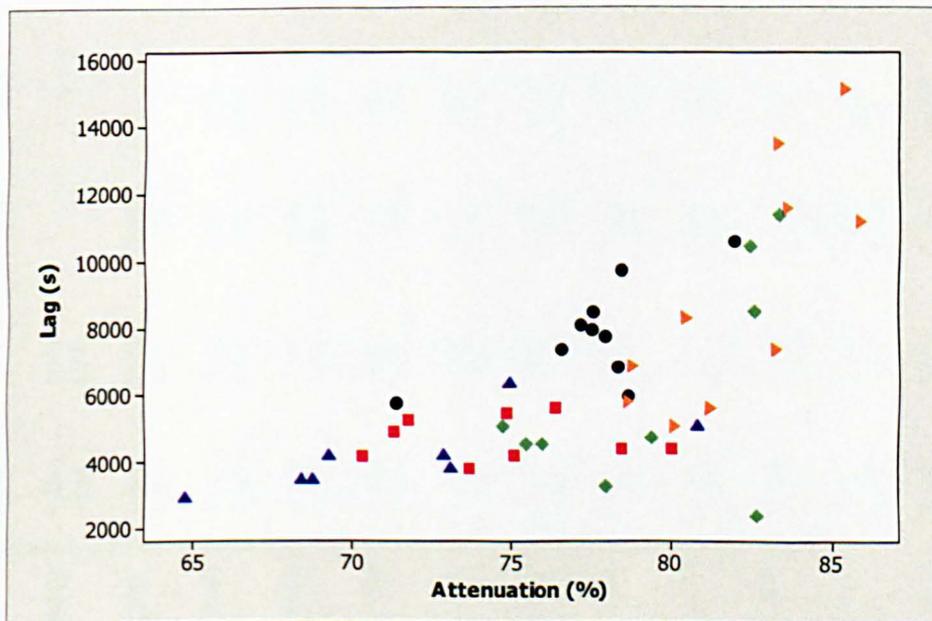


Figure 8.6 300km network grouping using the identified best classifiers

#### **8.2.4 Discriminant analysis at 600-kilometre scale**

In order to verify that the best classifiers identified at the 300-kilometre scale are not scale dependant, the analysis is repeated at another scale, based on the cluster analysis conducted on the 600-kilometre network responses.

The same strategy as for analysis of the 300-kilometre networks was applied for selecting the best classifiers. The correlation matrix calculated for these data is given in Table 8.4.

	NUM_Link	AV_LLength	AV_LSlope	MAX_Dist	Drop_MaxD	Dist_MaxW	MAX_Width	MEAN_Dist	STD_Dist	KURT_Dist	SKEW_Dist	MEAN_Drop	STD_Drop
<b>AV_LLength</b>	-0.932 0.000												
<b>AV_LSlope</b>	0.224 0.293	-0.303 0.150											
<b>MAX_Dist</b>	-0.370 0.075	0.336 0.109	-0.076 0.723										
<b>Drop_MaxD</b>	0.141 0.510	-0.318 0.131	0.633 0.001	0.005 0.982									
<b>Dist_MaxW</b>	-0.168 0.432	0.168 0.434	-0.010 0.964	0.765 0.000	-0.054 0.803								
<b>MAX_Width</b>	0.368 0.077	-0.271 0.201	0.052 0.809	-0.648 0.001	-0.056 0.795	-0.493 0.014							
<b>MEAN_Dist</b>	-0.409 0.047	0.377 0.069	-0.077 0.719	0.958 0.000	-0.115 0.594	0.808 0.000	-0.648 0.001						
<b>STD_Dist</b>	-0.310 0.140	0.265 0.210	-0.020 0.925	0.951 0.000	0.052 0.810	0.820 0.000	-0.773 0.000	0.913 0.000					
<b>KURT_Dist</b>	0.006 0.978	0.068 0.753	-0.126 0.558	-0.297 0.159	-0.201 0.347	-0.363 0.081	0.745 0.000	-0.243 0.252	-0.554 0.005				
<b>SKEW_Dist</b>	0.148 0.489	-0.236 0.267	0.276 0.191	0.123 0.568	0.430 0.036	-0.227 0.286	-0.254 0.231	-0.090 0.677	0.151 0.481	-0.375 0.071			
<b>MEAN_Drop</b>	-0.004 0.985	-0.143 0.504	0.756 0.000	0.168 0.433	0.749 0.000	0.188 0.380	-0.293 0.164	0.142 0.507	0.247 0.245	-0.333 0.112	0.342 0.102		
<b>STD_Drop</b>	0.030 0.888	-0.155 0.471	0.822 0.000	0.071 0.741	0.787 0.000	0.092 0.670	-0.097 0.651	0.015 0.945	0.123 0.568	-0.249 0.241	0.324 0.122	0.907 0.000	
<b>KURT_Drop</b>	0.356 0.088	-0.417 0.043	-0.235 0.270	-0.306 0.145	0.074 0.732	-0.197 0.357	0.497 0.013	-0.363 0.081	-0.312 0.138	0.196 0.359	-0.077 0.722	-0.322 0.125	-0.250 0.238

Table 8.4 Correlation matrix for the 600km network descriptor variables (upper number = correlation coefficient; lower number = p-value)

	NUM_Link	AV_LLlength	AV_LSlope	MAX_Dist	Drop_MaxD	Dist_MaxW	MAX_Width	MEAN_Dist	STD_Dist	KURT_Dist	SKEW_Dist	MEAN_Drop	STD_Drop
<b>SKEW_Drop</b>	0.244	-0.273	-0.212	-0.198	0.023	-0.176	0.478	-0.250	-0.254	0.257	-0.098	-0.462	-0.268
	0.251	0.197	0.320	0.353	0.915	0.411	0.018	0.239	0.231	0.225	0.648	0.023	0.206
<b>MAX_Drop</b>	0.188	-0.313	0.794	-0.058	0.849	0.018	0.098	-0.132	0.003	-0.182	0.312	0.808	0.911
	0.378	0.136	0.000	0.786	0.000	0.935	0.649	0.539	0.990	0.394	0.137	0.000	0.000
<b>MEAN_Eng</b>	0.158	-0.283	0.726	-0.316	0.757	-0.184	-0.002	-0.367	-0.184	-0.269	0.364	0.839	0.793
	0.462	0.180	0.000	0.132	0.000	0.391	0.992	0.077	0.390	0.204	0.080	0.000	0.000
<b>STD_Eng</b>	0.247	-0.298	0.715	-0.257	0.685	-0.073	0.104	-0.345	-0.137	-0.241	0.324	0.618	0.673
	0.244	0.158	0.000	0.226	0.000	0.734	0.630	0.099	0.524	0.257	0.122	0.001	0.000
<b>KURT_Eng</b>	0.264	-0.330	-0.036	-0.194	0.196	-0.301	0.251	-0.275	-0.220	0.150	0.151	0.059	-0.001
	0.213	0.116	0.869	0.363	0.359	0.154	0.236	0.194	0.301	0.483	0.480	0.785	0.996
<b>SKEW_Eng</b>	0.350	-0.426	-0.001	-0.035	0.238	-0.208	0.180	-0.138	-0.087	0.133	0.266	0.044	-0.009
	0.093	0.038	0.996	0.872	0.262	0.330	0.399	0.521	0.687	0.536	0.210	0.840	0.967
<b>MAX_Eng</b>	0.334	-0.457	0.287	-0.286	0.485	-0.325	0.123	-0.383	-0.231	-0.082	0.309	0.398	0.361
	0.110	0.025	0.174	0.176	0.016	0.122	0.567	0.065	0.278	0.704	0.142	0.054	0.083
<b>MAX_Order</b>	0.106	-0.178	-0.071	-0.379	0.064	-0.227	0.287	-0.384	-0.361	0.078	0.018	-0.002	0.068
	0.621	0.404	0.741	0.068	0.766	0.285	0.174	0.064	0.083	0.716	0.933	0.993	0.753
<b>AV_Bif</b>	0.323	-0.246	0.191	0.194	0.026	0.132	-0.130	0.178	0.209	-0.098	0.089	0.011	-0.034
	0.123	0.247	0.371	0.363	0.904	0.537	0.544	0.405	0.327	0.649	0.681	0.960	0.873
<b>NUM_Source</b>	1.000	-0.930	0.225	-0.375	0.142	-0.167	0.382	-0.414	-0.318	0.017	0.141	-0.007	0.033
	0.000	0.000	0.291	0.071	0.508	0.435	0.066	0.044	0.130	0.938	0.512	0.974	0.877

Table 8.4 Correlation matrix for the 600km network descriptor variables (continued)

	KURT_Drop	SKEW_Drop	MAX_Drop	MEAN_Eng	STD_Eng	KURT_Eng	SKEW_Eng	MAX_Eng	MAX_Order	AV_Bif
<b>SKEW_Drop</b>	0.882 0.000									
<b>MAX_Drop</b>	0.080 0.710	0.013 0.952								
<b>MEAN_Eng</b>	-0.062 0.773	-0.265 0.211	0.809 0.000							
<b>STD_Eng</b>	0.105 0.625	0.013 0.952	0.789 0.000	0.832 0.000						
<b>KURT_Eng</b>	0.635 0.001	0.411 0.046	0.156 0.465	0.222 0.297	0.227 0.287					
<b>SKEW_Eng</b>	0.627 0.001	0.485 0.016	0.153 0.476	0.119 0.578	0.197 0.357	0.934 0.000				
<b>MAX_Eng</b>	0.446 0.029	0.218 0.306	0.478 0.018	0.594 0.002	0.573 0.003	0.853 0.000	0.796 0.000			
<b>MAX_Order</b>	0.469 0.021	0.350 0.094	0.251 0.237	0.180 0.400	0.133 0.534	0.396 0.056	0.298 0.157	0.424 0.039		
<b>AV_Bif</b>	-0.294 0.163	-0.229 0.281	-0.139 0.516	-0.084 0.698	0.008 0.972	-0.251 0.237	-0.122 0.571	-0.253 0.233	-0.892 0.000	
<b>NUM_Source</b>	0.361 0.083	0.251 0.236	0.192 0.370	0.157 0.465	0.250 0.239	0.262 0.216	0.348 0.096	0.331 0.115	0.106 0.623	0.324 0.122

Table 8.4 Correlation matrix for the 600km network descriptor variables (continued)

35 linear relationships with p-value lower than 0.0005 are identified. Most of these relationships were also identified at the 300-kilometre scale. Two are different: the correlations between STD\_Eng and DROP\_MaxD and between KURT\_Dist and MAX\_Width were identified with a p-value equal to 0.002 and 0.001 respectively based on the 300-kilometre dataset, whereas they are identified with a p-value lower than 0.0005 based on this dataset. 20 correlations that were identified with a p-values lower than 0.0005 at the 300-kilometre scale are identified with higher p-values based on this dataset and 14 of these show p-values higher than 0.05. This suggests that when the total length of the networks analysed increases from 300 to 600 kilometres, the linear independency between the potential classifying variables increases.

A summary of the discriminant analysis results is given in Table 8.5.

Set	Variables	Classification Index (%)
All Standard Variables		33.3
Best Discriminating least correlated	AV_Lslope; NUM_Source; Drop_MaxD; MAX_Dist	50
discard 1 lowest contribution	AV_Lslope; NUM_Source; Drop_MaxD	45.8
discard 2 lowest contribution	AV_Lslope; NUM_Source	50
discard 3 lowest contribution	AV_Lslope	45.8
Standard Best Standard Classifiers + Others		
Best Discriminating least correlated	MEAN_Eng; MAX_Eng; MAX_Order; SKEW_Dist; KURT_Dist; SKEW_Drop; DIST_MaxW	37.5
discard 1 lowest contribution	MEAN_Eng; MAX_Eng; MAX_Order; SKEW_Dist; KURT_Dist; SKEW_Drop	50.0
discard 2 lowest contribution	MEAN_Eng; MAX_Eng; MAX_Order; SKEW_Dist; KURT_Dist	54.2
discard 3 lowest contribution	MEAN_Eng; MAX_Eng; MAX_Order; SKEW_Dist;	25.0
discard 4 lowest contribution	MEAN_Eng; MAX_Eng; MAX_Order	25.0
discard 5 lowest contribution	MEAN_Eng; MAX_Eng	20.8
discard 6 lowest contribution	MEAN_Eng	37.5

Table 8.5 Results of discriminant analysis at 600km scale. The variables are listed in decreasing order of discriminating contribution.

The best standard classifier identified at this scale is AV\_LSlope, as it was at the 300-kilometre scale. Classification based on this single variable leads to a performance of 45.8%, which is much higher than that obtained based on that same variables at a smaller scale (26.1%).

The second step of the analysis including the extended set of ‘possible classifiers’ and ‘original’ variables leads to the identification of the set of best classifiers,

which includes, in decreasing order of contribution, MEAN\_Eng, MAX\_Eng, MAX\_Order, SKEW\_Dist and KURT\_Dist. Classification based on these 5 variables leads to a performance of 54.2%. The values for the coefficients of the linear discriminant functions obtained for each group are reported in Table 8.6

Group	1	2	3	4	Discriminating Power
MEAN_Eng	2357.9	3114	1889.8	1013.7	2093.85
MAX_Eng	-127	-131.2	-122.6	-121.7	125.625
MAX_Order	33.3	36.1	34.5	35.1	34.75
SKEW_Dist	-15.3	-17.6	-15	-11.5	14.85
KURT_Dist	-5.3	3.2	-4.2	-11	5.925

Table 8.6 Linear discriminant function coefficients and derived potential classifier's discriminating powers at 600km scale

3 of the 4 best discriminating variables identified at the 300-kilometre scale are included in the set of best classifiers identified at this scale. These are MEAN\_Eng, MAX\_Order, and KURT\_Dist. More variables are necessary to achieve a similar classification index at the 600-kilometre scale than were at the 300-kilometre scale. However the size of the dataset available at the 600-kilometre scale is much smaller (24 networks) than the one used for discriminant analysis at the 300-kilometre scale (46 networks). This might explain in part the differences in the results, which are however not divergent. At both scales, MEAN\_Eng is the network descriptor that shows by far the highest discriminating contribution.

### 8.3 Best Network Classifiers

24 network descriptor variables have been tested for their potential capacity to characterise network flow response in 46 networks whose total length is 300 kilometres. 4 have been identified as best able to reproduce network flow response clusters. It is worth noting that none of the variables are from the set that was called 'standard' because of their prevalence in previous classification studies. However, it should be noted that the 'standard' variable that shows the best discriminating power, AV\_LSlope, bears a marked similarity to the overall

best classifier, MEAN\_Eng, as both variables take into account a measure of the average slope of the network. The variable with the most discriminating power, MEAN\_Eng, is one that has been devised as part of this thesis, with the intention of describing the energy stored and the potential to dissipate it in the network. The other three best classifiers all pertain to the plan-view shape of the network; two describe the network width function. MAX\_Width is an index of how splayed out a dendritic network is. KURT\_Dist, the kurtosis in the width function, is a measure of the peakiness of the distance to the outlet distribution. Distributions that are flatter than a normal distribution, and consequently have higher values in their tails, will have a negative value of kurtosis; whereas those that are 'spikier' than a normal distribution, with most of the distribution bunched up around the mean (or some central value) and lower probabilities in the tails, will have a positive value of kurtosis. Thus in terms of the network width function, it is an index of whether the number of channels occurring varies fairly uniformly with distance (negative kurtosis) or there is a high concentration of channels at one particular distance (positive kurtosis). The final discriminating variable, MAX\_Order, is interesting. It dates back to insights of Horton, Strahler and Shreve at the beginning of the last century. They believed that the way rivers bifurcate as you move upstream towards the headwaters would play a key role in shaping hydrographs. MAX\_Order is an index of the nature of the bifurcations that have taken place. If MAX\_Order is high, it indicates that the river splits into fairly evenly sized rivers, whereas a low value of MAX\_Order means that when the channels split, there is a dominant main channel and smaller tributaries.

When the total length of the networks analysed is increased from 300 to 600 kilometres, the best set of classifiers changes slightly. Three of the variables remain the same: MEAN\_Eng; MAX\_Order and KURT\_Dist. MAX\_Width is replaced by SKEW\_Dist and MAX\_Eng. SKEW\_Dist measures how skewed the network width function is. A high positive skew means that there are more reaches of the network that are close to the outlet than distant, whereas a negative skew indicates that there are more reaches distant from the outlet than close. It would appear, therefore, that as scale increases, how splayed the network is (as measured by MAX\_Width) is less important in influencing the hydrograph than where the splaying occurs (as measured by SKEW\_Dist). The inclusion of

MAX\_Eng is perhaps surprising since this is an extremely poor statistic to use to describe the distribution of the energy index in the whole network. It only pertains to the one reach where the energy is at a maximum.

The sets of best classifiers at both scales only allow for approximately 50% of the networks to be allocated to their true membership. Nonetheless, the classification based on the new variables devised here is far more successful than the one based on the 7 variables commonly used in the context of regionalisation studies (which leads to a classification index of 35%). Moreover, only 4 variables are required and all of these are relatively easily derivable from a digital database. Therefore it is anticipated that the use of these variables will improve the classification of river networks in regionalisation studies. The question still remains as to why almost 50% of the networks are misclassified. The answer may lie in the original groups identified during the cluster analysis of network responses. The boundaries of the clusters (Figure 8.3 and Figure 8.5) are close to one another. This means that there is no absolutely obvious grouping in the networks, and hence a network that sits close to the boundary within its own group is also fairly close to the adjacent group. This might be due to the nature of Scottish river networks. It may be that they are actually relatively homogeneous in their structure and that much more distinct groups would have fallen out of an analysis of all UK river networks. For example, one might envisage that rivers that drain the 'chalky' South Downs of England would have a very different structure and signature hydrograph than typical Scottish rivers. The consequence of this relative homogeneity can be seen when the nature of the misclassification is analysed. Close comparison of Figure 8.3 and Figure 8.6 reveals that where a network has been misclassified, for most part, it falls into a group that sits adjacent in the plots of hydrograph responses. Thus while the misclassification is occurring, the decision rules that have been derived by discriminant analysis are not placing the networks in wildly different groups.

## 8.4 Summary

A multivariate statistical analysis was conducted on 46 Scottish river networks and their simulated responses to a pulse of runoff, the aim being to identify a set of descriptor variables that best characterise surface flow through river networks. The analysis was conducted in two steps. Firstly, cluster analysis was applied to form groups of networks that responded similarly to the same pulse of runoff. Secondly, linear discriminant analysis was applied to a set of network descriptor variables with the aim of testing these for their capacity in reproducing the network response clusters.

Of all the variables tested, 4 stand clear in showing the best capacity to reproduce these clusters: the mean of the energy index (ratio of elevation drop to distance to the outlet) spatial distribution along the network; the network order as defined by Strahler (1957); the Kurtosis of the distance to the outlet spatial distribution along the network; and the maximum of the width function as defined by Shreve (1969). These variables have not traditionally been used in regionalisation studies and, therefore, should prompt a reassessment in the research community of what are the important properties of networks in shaping a flood wave. The discriminating rules based on these variables failed to classify 50% of the networks. However, considering the relative and, prior to this thesis, unexpected homogeneity in the responses of Scottish rivers, it is remarkable that the discrimination rules managed to classify 50% of the networks correctly. It is anticipated that if the procedures that led to the selection of the best classifiers were applied to a more diverse group of networks, selected from across the whole of the UK, more distinct clusters would arise, for example with all Scottish networks falling into one or two more populous groups. Then, discriminating functions based on the same 4 variables may be able to correctly classify a higher proportion of the networks.

## Chapter 9

### 9 Conclusions and Future Research

The first section of this chapter aims to conclude with a concise summary of the research advances made in this thesis. A second section underlines areas of future research.

#### 9.1 Conclusions

Research advances in the area of hydrological modelling have been made in this thesis at three different levels: a conceptual one; a practical one; and a scientific one.

##### *At a conceptual level:*

- A novel method for classifying river basins specially designed for estimation of hydrological model parameters in ungauged basins applying regionalisation techniques has been developed. This method is based on two novel concepts: the first is that the constraint of each basin belonging to only one group is discarded and instead, each basin undergoes three independent classifications, one relative to each of the major components of the land phase hydrological cycle: interaction of soil water / vegetation and atmosphere; surface flow; and groundwater flow. This insures that there are sufficient members of any single group to have a realistic prospect of deriving an empirical relationship between descriptive variables and hydrological model parameters. The second is in the methodology for objectively selecting which of the basin characteristics are the most appropriate classifiers. There is very rarely hydrological data

that can characterise the response of individual components of the hydrological cycle. In most catchments, only the flow at the outlet is measured. Therefore, the objective test devised in this thesis consists in analysing synthetic data derived by simulation using standard statistical multivariate analysis techniques. This thesis focused on the surface flow component, applying the methodology to identify the best classifiers for flow through river networks.

***At a practical level:***

- A new integrated modelling system for simulating flow in large river networks has been developed that extracted all relevant information on network structure from digital databases of Scotland, and solved a kinematic wave routing model. This is already being used in other research projects at Glasgow University.
- A novel algorithm has been developed for ensuring that river networks derived from a digital terrain model adhered as closely as possible to those on Ordnance Survey maps. This involved locating the source of each river and tracing the path of steepest descent down to the catchment outlet. It is anticipated that this will form the basis of a short journal paper warning hydrologists of the perils of using traditional techniques that employ a discrete threshold in flow accumulation to determine where rivers rise.
- The first extensive test of hydraulic geometry relationships for Scotland has been conducted. From this, a clear and consistent relationship between the median annual maximum flow and channel widths for Scotland was derived.

***At a scientific level,*** significant insights have been made into the relationship between river network geomorphologic structure and stream flow response. These include:

- The importance of channel conveyance in dictating the stream flow response of a network. Whilst this may seem obvious, it has been neglected in previously published theoretical studies of network behaviour; most other models assume a constant water velocity in each reach of the network. However, the features of the hydrograph, namely percentage attenuation and lag in time to peak, scale linearly with both roughness and hydraulic geometry conveyance parameters. If these two variables are plotted for a large number of networks for a given pair of conveyance parameters, then the relative position of the networks on a second plot for a different pair of conveyance parameters does not change.
- The influence of storm duration on the scaling of peak flows with drainage area. Flood peaks in a network scale following a power law with drainage area. This has been reported by previous researchers. However, it is shown in this thesis that the exponent in the power law relationship is a function of storm duration. This is a result that could not be achieved using the simple linear flow routing models used by previous researchers. It is a direct result of the more realistic non-linear kinematic wave model used here. It is of particular significance if real river flow records are to be analysed to look for scaling relationships because it implies that storm duration should be a factor. Thus multifractal scaling laws should include storm duration as one of the variables.
- The identification of 4 network descriptor variables as best classifiers for the surface flow through river network component. A total of 24 potential classifiers were tested. Of these, 7 have been widely used in the past, 4 have been suggested in the literature as being pertinent, and 13 were new, derived specially for this thesis based on knowledge of the physics of fluid flow in rivers. These were whittled down to 4 best classifiers. None of these were classifiers that have been used in previous regionalisation studies. Traditionally, no objective test has been applied and the set of classifiers used reflected the particular prejudices of the individual researcher (Newson, 1978). Therefore the objective method used in this

thesis for arriving at the classifying variables warrants them being selected in future studies in preference to the more traditional classifiers.

- The demonstration that the most powerful classifier is based on the ‘energy index’ spatial distribution along the network, which was derived specially for this thesis.

## 9.2 Future Research

This thesis left significant scope for future research in the context of the ambitious overall research project that piloted a whole new philosophy for classifying river basins and estimating hydrological model parameters in ungauged basins. It also opened up unexpected future research avenues in a variety of different ways.

### *In the context of the overall project*

- Developing similar classification schemes for groundwater and soil-vegetation-atmosphere models will take many more years of research.
- Devising regression equations to transfer hydrological model parameters from gauged to ungauged basins within the same group and then validating them is a further PhD in itself.

### *Unexpected avenues include:*

- The investigation of the influence of storm duration on the scaling of peak flows along a network, observed in synthetic data, in suitably analysed river flow records from nested catchments.
- The potential benefits of integrating the ‘energy index’ distribution in function based hydrological models. Currently, for large scale hydrological modelling, network width function-based models are

commonly used. It should be possible to derive a similar model that uses the 'energy index' distribution instead. On the evidence of this thesis, such a model would better characterise the hydrological response.

## References

- Abbott, M.B. (1978). *Computational Hydraulics: Elements of the theory of free surface flows*. Pitman.
- Abdulla, F.A. and Lettenmaier, D.P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of Hydrology*, 197: 230-257.
- Acreman, M.C. and Sinclair, C.D. (1986). Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *Journal of Hydrology*, 84: 365-380.
- Acreman, M.C. and Wiltshire, S.E. (1987). Identification of regions for regional flood frequency analysis. *EOS* 68(44): 1262 (Abstract).
- Bates, C.B., Rahman, A., Mein, R.G. and Weinmann, P.E. (1998). Climatic and physical factors that influence the homogeneity of regional floods in southeastern Australia. *Water Resources Research*, 34(12): 3369-3381.
- Beven, K.J. (1995). Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models. *Scale Issues In Hydrological Modelling: Hydrological Processes*, 9: 507-525.
- Beven, K.J. (1997). TOPMODEL: A critique. *Hydrological Processes*, 11: 1067-1085.
- Beven, K.J. and Kirkby, M.J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1): 43-69.
- Burn, D.H. (1990a). An appraisal of the 'region of influence' approach to flood frequency analysis. *Hydrological Sciences Journal*, 35(2): 149-165.

- Burn, D.H. (1990b). Evaluation of regional flood frequency analysis with region of influence approach. *Water Resources Research*, 26(10): 2257-2265.
- Burn, D.H. and Boorman, D.B. (1993). Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology*, 143: 429-454.
- Calver, A., Lamb, R., Kay, A.L. and Crewett, J. (2001). The continuous simulation method for river flood frequency estimation. R&D project FD0404 Final Report, CEH Wallingford, UK.
- Chow, V.T. (1959). *Open-channel hydraulics*. McGraw-Hill.
- Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). *Applied hydrology*. McGraw-Hill.
- Dooge, J.C. (1982). The parametrization of hydrological processes. In: *Land Surface Processes in Atmospheric Global Circulation Models*, P.S. Eagleson editor, Cambridge University Press, 243-288.
- Everitt, B.S. and Dunn, G. (1991). *Applied Multivariate Data Analysis*. Edward Arnold.
- Ewen, J. and Parkin, G. (1996). Validation of catchment models for predicting land-use and climate change impacts. 1. Method. *Journal of Hydrology*, 175(1-4): 583-594.
- Fernandez, W., Vogel, R.M. and Sankarasubramanian, A. (2000). Regional calibration of a watershed model. *Hydrological Sciences Journal*, 45(5): 689-707.
- Fetter, C.W (1994). *Applied hydrogeology*, Third edition. Prentice Hall.
- Fisher, R.A. (1936). The use of multiple measurements on taxonomic problems. *Annals of Eugenics*, 7: 179-188.

Franco, J.L. and Chaudhry, F.H. (1998). A comparison of various finite element procedures for watershed routing. *International Conference on Computational Methods in Water Resources, CMWR, Computational Methods in Contamination and Remediation of Water Resources*, 1: 537-544.

Gupta, V.K., Waymire, E. and Rodriguez-Iturbe, I. (1986). On scales, gravity and network structure in basin runoff. In: *Scale Problems in Hydrology*, V.K. Gupta, I. Rodriguez-Iturbe and E.F. Wood editors, Dordrecht, 159-184.

Gupta, V.K. and Dawdy, D.R. (1995). Physical interpretations of regional variations in the scaling exponents of flood quantiles. *Hydrological Processes*, 9: 347-361.

Gupta, V.K., Castro, S.L. and Over, T.M. (1996). On scaling exponents of spatial peak flows from rainfall and river network geometry. *Journal of Hydrology*, 187: 81-104.

Hall, M.J. and Minns, A.W. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal*, 44(5).

Hand, D.J. (1981). *Discrimination and classification*. Wiley series in probability and mathematical statistics, John Wiley & Sons.

Hey, R.D. (1975). *Design Discharge for Natural Channels*. R.D. Hey and T.D. Davies Eds, *Science, Technology and Environmental Management*, Saxon House, Farnborough, England, 73-88.

Hey, R.D. and Thorne, C.R. (1984). Flow processes and river channel morphology. *Proceedings of a meeting of International Geographical Union Commission on Field Experiments in Geomorphology*, 489 – 514.

Hey, R.D. and Thorne, C.R. (1986). Stable Channels with Mobile Gravel Beds. *Journal of Hydraulic Engineering*, 112(8): 671-689.

- Hornberger, G.M., Raffensperger, J.P., Wiberg, P.L. and Eshleman, K.N. (1998). Elements of physical hydrology. Johns Hopkins University Press.
- Horton, R.E. (1932). Drainage-basin characteristics, EOS Trans. AGU, 13: 350-361.
- Horton, R.E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Bull.Geol. Soc. Am., 56: 275-370.
- Hosking, J.R.M. and Wallis, J.R. (1997). Regional frequency analysis: an approach based on L-Moments. Cambridge University Press.
- Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). An appraisal of regional flood frequency procedure in the UK 'Flood Studies Report'. Hydrological Sciences Journal, 30(1): 85-109.
- Huang, H.Q. and Warner, R.F. (1995). The multivariate controls of hydraulic geometry: a causal investigation in terms of boundary shear distribution. Earth Surface Processes and Landforms, 20: 115-130.
- Kirkby, M.J. (1976). Tests of the random model and its application to basin hydrology. Earth Surface Processes and Landforms, 1: 197-212.
- Lacey, G. (1930). Stable channels in alluvium. Proceedings of the Institution of Civil Engineers, 229(1): 259-384.
- Lacey, G. (1939). Regime flow in incoherent alluvium. Central Board Irrigation (India), Simla, 20.
- Leopold, L.B. and Maddock, T. (1953). The hydraulic geometry of stream channels and some physiographic implications. Geological Survey Professional Paper, 252.

- Li, R.M., Simons, D.B. and Stevens, M.A. (1975). Nonlinear Kinematic Wave Approximation for Water Routing. *Water Resources Research*, 11(2): 245-252.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5<sup>th</sup> Berkeley Symposium*, 1: 281-297.
- Maidment, D.R., Olivera, F., Calver, A., Eatherall, A. and Fraczek, W. (1996). Unit hydrograph derived from a spatially distributed velocity field. *Hydrological Processes*, 10(6): 831-844.
- Menabde, M., Veitzer, S., Gupta, V. and Sivapalan, M. (2001). Tests of peak flow scaling in simulated self-similar river networks. *Advances in Water Resources*, 24: 991-999.
- Merz, R. and Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1-4): 95-123.
- Mesa, O.J. and Mifflin, E.R. (1988). On the relative role of hillslope and network geometry in hydrologic response. In: *Scale Problems in Hydrology*, V.K. Gupta, I. Rodriguez-Iturbe and E.F. Wood editors, Dordrecht, 181-190.
- Milligan, G.W. (1980). An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms. *Psychometrica*, 45: 325-342.
- Morris, G.D. and Flavin, R.W. (1990). A Digital Terrain Model for Hydrology. *Proceedings of the 4<sup>th</sup> International Symposium on Spatial Data Handling*, July 23-27, Zurich, 1: 250-262.
- Morrison, J.E. and Smith, J.A. (2001). Scaling Properties of Flood Peaks. *Extremes* 4(1): 5-22.
- Moussa, R. (2003). On morphometric properties of basins, scale effects and hydrological response. *Hydrological Processes*, 17: 33-58.

Naden, P.S. (1992). Spatial variability in flood estimation for large catchments – The exploitation of channel network structure. *Hydrological Sciences Journal*, 37(1): 53-71.

Naden, P., Broadhurst, P., Tauveron, N. and Walker, A. (1999). River routing for global climate models. *Hydrology and Earth System Science*, 3(1): 109-124.

Nathan, R.J. and MacMahon, T.A. (1990). Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology*, 121: 217-238.

Natural Environment Research Council (NERC) (1975). *Flood Studies Report*. Wallingford UK.

Natural Environment Research Council (NERC) (1999). *Flood Estimation Handbook*. Wallingford UK.

Newson, M. (1978). Drainage basin characteristics, their selection, derivation and analysis for flood studies in the British Isles. *Earth Surface Processes*, 3: 227-293.

Nixon, M. (1959). A study of the bank-full discharges of rivers in England and Wales. *Proceedings of the Institution of Civil Engineers*, 12: 151-174.

Post, A.D. and Jakeman, A.J. (1996). Relationships between catchment attributes and response characteristics in small Australian mountain ash catchments. *Hydrological Processes*, 10: 877-892.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1993). *Numerical Recipes in FORTRAN 77 The Art of Scientific Computing*. Cambridge University Press.

Reed, D.W., Bayliss, A.C., Jakob, D., Jones, T.K. and Marshall, D.C.W. (1995). *The use of date information in proposing regions for flood frequency analysis*. Institute of Hydrology, Wallingford, UK.

Reungoat, A.F.J. and Sloan, W.T. (2001). Classification of river networks. First International Conference on River Basin Management, Sep. 11-13 2001, Cardiff, UK. Progress in Water Resource, River Basin Management, WITPress, 23-29.

Reungoat, A.F.J. and Sloan, W.T. (2002). Classifying components of river basins for hydrological model parameter estimation. British Hydrological Society (BHS) Occasional Paper No.13: Continuous river flow simulation: methods, applications and uncertainties. Ed. Ian Littlewood.

Rodriguez-Iturbe, I. and Valdes, J.B. (1979). The geomorphologic structure of hydrologic response. Water Resources Research, 15(6): 1409-1420.

Rodriguez-Iturbe, I. and Rinaldo, A. (1997). Fractal river basins: chance and self-organisation. Cambridge University Press.

Shaw, E.M. (1994). Hydrology in Practice. Chapman & Hall.

Shreve, R.L. (1966). Statistical law of stream numbers. J. Geol., 74: 17-37.

Shreve, R.L. (1969). Stream lengths and basin areas in topologically random channel networks. J.Geol., 74:397-414.

Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrological Processes, 17: 3163-3170.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mondono, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S. and Zehe, E. (2003). IAHS decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48(6): 857-880.

Sloan, W.T. (1999). Upscaling hydrological processes and the development of a large-scale river basin modelling system. PhD Thesis, Department of Civil Engineering, University of Newcastle upon Tyne, UK.

Smith, J.A. (1992). Representation of Basin Scale in Flood Peak Distributions. *Water Resources Research*, 28(11): 2993-2999.

Strahler, A.N. (1952). Hypsometric (area altitude) analysis of erosional topography. *Geol. Soc. Am. Bull.*, 63:1117-1142.

Strahler, A.N. (1957). Quantitative analysis of watershed geomorphology. *EOS Trans. AGU*, 38: 912-920.

Streeter, V.L. (1966). *Fluid Mechanics*. McGraw-Hill.

Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I. (1991). On the extraction of channel networks from digital elevation data. *Hydrological Processes*, 5: 81-100.

Veitzer, S.A. and Gupta, V.K (2001). Statistical self-similarity of width function maxima with implications to floods. *Advances in Water Resources*, 24: 955-965.

Wolock, D.M., Winter, T.C. and McMahon, G. (2003). Delineation of hydrologic setting regions in the United States using geographic information system tools and multivariate statistical analyses. U.S. Geological Survey Open-File Report, 145.

Zrinji, Z. and Burn, D.H. (1994). Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, 153: 1-21.

## Appendix A

List of catchment outlets for which network data were extracted from GIS databases: Gauge Code, Name, Grid References, Catchment Area and outlet flow statistic  $Q_{med}$  are taken from Appendix B, Volume 3 of the Flood Estimation Handbook, (NERC, 1999); Network Length corresponds to the total length of channel calculated by the network data GIS extraction algorithm detailed in Chapter 5.

Code	Name		Grid References		Area km <sup>2</sup>	Net.Length km	Record	Q <sub>med</sub> m <sup>3</sup> /s
	River	Site	X	Y				
w3001	Shin	Lairg	258100	906200	495	993	1950-1955	63
w3002	Carron	Sgodachail	249000	892100	241	354	1974-1992	193
w3003	Oykel	Easter Turmaig	240300	900100	331	603	1978-1992	375
w4001	Conon	Moy Bridge	248200	854700	962	1683	1945-1955	312
w4003	Alness	Alness	265400	869500	201	275	1974-1992	84
w6003	Moriston	Invermoriston	241600	816900	391	827	1930-1943	314
w7002	Findhorn	Forres	301800	858300	782	1196	1958-1991	359
w7003	Lossie	Sheriffmills	319400	862600	216	306	1958-1994	40
w8002	Spey	Kinrara	288100	808200	1012	1703	1951-1994	135
w8003	Spey	Ruthven Bridge	275900	799600	534	914	1951-1972	102
w8004	Avon	Delnashaugh	318600	835200	543	818	1952-1994	225
w8005	Spey	Boat of Garten	294600	819100	1268	2022	1951-1994	158
w8007	Spey	Invertruim	268700	796200	400	695	1952-1994	96
w8008	Tormie	Tromie Bridge	278900	799500	130	260	1952-1988	59
w8009	Dulnain	Balnaa Bridge	297700	824700	272	399	1952-1994	101
w8010	Spey	Grantown	303300	826800	1749	2713	1952-1994	246
w8011	Livet	Minmore	320100	829100	104	149	1981-1994	30
w9001	Deveron	Avochie	353200	846400	442	566	1960-1994	119
w9002	Deveron	Muiresk	370500	849800	955	1112	1960-1994	230
w9003	Isla	Grange	349400	850600	176	220	1969-1994	42
w9004	Bogie	Redcraig	351900	837300	179	230	1981-1994	23
w10002	Ugie	Inverugie	410100	848500	325	281	1972-1994	41
w10003	Ythan	Ellon	394700	830300	523	484	1983-1994	63
w11001	Don	Parkhill	388700	814100	1273	1320	1970-1992	119
w11002	Don	Haughton	375600	820100	787	897	1972-1994	106
w11003	Don	Bridge of Alford	356600	817000	499	617	1974-1994	93
w11004	Urie	Pitcaple	372100	826000	198	181	1988-1994	20
w12001	Dee	Woodend	363500	795600	1370	1808	1929-1994	429
w12002	Dee	Park	379800	798300	1844	2350	1973-1994	572
w12003	Dee	Polhollick	334400	796500	690	956	1976-1994	312
w12005	Muick	Invermuick	336400	794700	110	167	1977-1994	67
w12006	Gaim	Invergaim	335300	797100	150	198	1978-1994	59
w12007	Dee	Mar Lodge	309800	789500	289	430	1982-1994	196
w12008	Feugh	Heugh Head	368700	792800	229	333	1985-1994	138
w13001	Bervie	Inverbervie	382600	773300	123	146	1979-1994	37
w14001	Eden	Kemback	341500	715800	307	351	1967-1992	43
w15007	Tay	Pitnacree	292400	753400	1149	2086	1952-1992	332
w15008	Dean Water	Cookston	334000	747900	177	135	1953-1992	30
w15010	Isla	Wester Cardean	329500	746600	367	440	1972-1992	101
w15013	Almond	Almondbank	306700	725800	175	262	1974-1992	120
w15017	Braan	Ballinloan	297900	740600	197	292	1975-1992	120

## Appendix A (continued)

Code	Name		Grid References		Area km <sup>2</sup>	Net.Length km	Record	Q <sub>med</sub> m <sup>3</sup> /s
	River	Site	X	Y				
w16001	Earn	Kinkell Bridge	293300	716700	591	950	1949-1992	193
w16002	Earn	Aberuchill	275400	721600	177	302	1955-1972	58
w16003	Ruchill Water	Cultybraggan	276400	720400	100	192	1960-1992	165
w16004	Earn	Forteviot Bridge	304300	718400	782	1259	1974-1992	251
w17001	Carron	Headwood	283200	682000	122	209	1968-1992	82
w17002	Leven	Leven	336900	700600	424	589	1968-1972	29
w17005	Avon	Polmonthill	295200	679700	195	266	1971-1992	59
w18001	Allan Water	Kinbuck	279200	705300	161	302	1957-1981	66
w18002	Devon	Glenochil	285800	696000	181	320	1956-1972	41
w18003	Teith	Bridge of Teith	272500	701100	518	1093	1956-1972	183
w18005	Allan Water	Bridge of Allan	278600	698000	210	376	1972-1992	96
w18008	Leny	Anie	258500	709600	190	410	1974-1992	90
w19001	Almond	Craigiehall	316500	675200	369	506	1956-1991	120
w19005	Almond	Almondell	308600	668600	229	347	1962-1992	78
w19006	Water of Leith	Murrayfield	322800	673200	107	135	1962-1991	31
w19007	Esk	Musselburgh	333900	672300	330	451	1962-1991	70
w19008	South Esk	Prestonholm	332500	662300	112	185	1963-1988	19
w19011	North Esk	Dalkeith Palace	333300	667800	137	193	1962-1991	41
w20001	Tyne	East Linton	359100	676800	307	343	1959-1991	49
w20003	Tyne	Spilmersford	345600	668900	161	169	1962-1991	31
w21003	Tweed	Peebles	325700	640000	694	1091	1939-1992	175
w21005	Tweed	Lyne Ford	320600	639700	373	610	1981-1992	124
w21006	Tweed	Boleside	349800	633400	1500	2400	1961-1992	400
w21007	Ettrick Water	Lindean	348600	631500	499	898	1961-1992	233
w21008	Teviot	Orminston Mill	370200	628000	1110	1638	1960-1992	343
w21010	Tweed	Dryburgh	358800	632000	2080	3177	1949-1981	449
w21011	Yarrow Water	Philiphaghaugh	343900	627700	231	408	1962-1981	83
w21012	Teviot	Hawick	352200	615900	323	539	1963-1992	184
w21013	Gala Water	Galashiels	347900	637400	207	290	1963-1992	51
w21015	Leader Water	Earlston	356500	638800	239	321	1966-1992	60
w21016	Eye Water	Eyemouth Mill	394200	663500	119	143	1967-1992	34
w21020	Yarrow Water	Gordon Arms	330900	624700	155	274	1967-1980	52
w21022	Whiteadder Water	Hutton Castle	388100	655000	503	711	1970-1988	117
w21023	Leet Water	Coldstream	383900	639600	113	136	1973-1981	48
w21024	Jed Water	Jedburgh	365500	621400	139	221	1972-1988	59
w21025	Ale Water	Ancrum	363400	624400	174	246	1973-1992	43
w21027	Blackadder Water	Mouth Bridge	382600	653000	159	197	1974-1991	39
w21031	Till	Etal	392700	639600	648	964	1956-1977	81
w21032	Glen	Kirknewton	391900	631000	199	344	1961-1982	42
w21034	Yarrow Water	Craig Douglas	328800	624400	116	208	1969-1973	32
w22007	Wansbeck	Mitford	417500	585800	287	371	1963-1994	95
w23002	Derwent	Eddys Bridge	404100	550800	118	172	1955-1964	42
w23003	North Tyne	Reaverhill	390600	573200	1008	1893	1959-1985	403
w23004	South Tyne	Haydon Bridge	385600	564700	751	1214	1959-1992	416
w23005	North Tyne	Tarset	377600	586100	285	554	1960-1978	214
w23006	South Tyne	Featherstone	367200	561100	322	510	1966-1992	248
w23007	Derwent	Rowlands Gill	416800	558100	242	336	1963-1992	38
w23008	Rede	Rede Bridge	386800	583200	344	615	1968-1992	126
w23010	Tarset Burn	Greenhaugh	378900	587900	96	157	1970-1978	61
w23015	North Tyne	Barrasford	392400	572100	1044	1941	1947-1969	456
w24003	Wear	Stanhope	398400	539100	172	334	1958-1992	119
w24008	Wear	Witton Park	417400	530900	455	782	1974-1992	182
w25002	Tees	Denk Bank	393200	526000	217	469	1959-1973	280
w25018	Tees	Middleton in Teesdale	395000	525000	242	506	1972-1992	181

## Appendix A (continued)

Code	Name River	Site	Grid References		Area km <sup>2</sup>	Net.Length km	Record	Qmed m <sup>3</sup> /s
			X	Y				
w75017	Ellen	Bullgill	309600	538400	96	87	1975-1983	53
w76008	Irthing	Greenholme	348600	558100	335	307	1967-1993	194
w76009	Caldew	Holm Hill	337800	546900	147	106	1968-1993	80
w76010	Petteril	Harraby Green	341200	554500	160	98	1970-1993	25
w77001	Esk	Netherby	339000	571800	842	1553	1961-1993	604
w77002	Esk	Canonbie	339700	575100	495	881	1963-1988	360
w77003	Liddel Water	Rowanbumfoot	341500	575900	319	641	1974-1992	261
w77005	Lyne	Cliff Bridge	341200	566200	191	322	1976-1983	123
w78003	Annan	Brydekirk	319100	570400	925	1671	1967-1992	296
w78005	Kinnel Water	Bridgemuir	309100	584500	229	461	1979-1992	129
w79002	Nith	Friars Carse	292300	585100	799	1613	1957-1992	454
w79004	Scar Water	Capenoch	284500	594000	142	297	1963-1992	148
w79005	Cluden Water	Fiddlers Ford	292800	579500	238	481	1964-1992	109
w79006	Nith	Drumlanrig	285800	599400	471	1011	1967-1992	316
w80001	Urr	Dalbeattie	282200	561000	199	340	1964-1992	102
w81002	Cree	Newton Stewart	241200	565300	368	752	1963-1992	225
w81003	Luce	Airyhemming	218000	559900	171	341	1967-1991	155
w82001	Girvan	Robstone	221700	599700	246	483	1963-1991	87
w82003	Stinchar	Bainowart	210800	583200	341	693	1975-1991	197
w83003	Ayr	Catrine	252500	625900	166	305	1969-1980	128
w83004	Lugar	Langholm	250800	621700	181	375	1973-1992	150
w83005	Irvine	Shewalton	234500	636900	381	587	1973-1992	215
w83006	Ayr	Mainholm	236100	621600	574	973	1976-1992	251
w83802	Irvine	Kilmarnoch	243000	636900	218	353	1913-1987	71
w84001	Kelvin	Killermont	255800	670500	335	444	1947-1992	96
w84003	Clyde	Hazelbank	283500	645200	1093	1530	1955-1993	272
w84004	Clyde	Sills	292700	642400	742	1074	1955-1992	199
w84005	Clyde	Blairston	270400	657900	1704	2321	1955-1993	383
w84007	South Calder Water	Forgewood	275100	658500	93	90	1965-1992	21
w84012	White Cart Water	Hawkhead	249900	662900	227	347	1963-1992	123
w84013	Clyde	Daldowie	267200	661600	1903	2543	1963-1987	392
w84014	Avon Water	Fairholm	275500	651800	266	392	1964-1992	188
w84015	Kelvin	Dryfield	263800	673900	235	310	1947-1987	62
w84017	Black Cart Water	Milliken Park	241100	662000	103	189	1968-1972	28
w84018	Clyde	Tulliford Mill	289100	640400	933	1387	1969-1981	240
w84019	North Calder Water	Calderpark	268100	662500	130	136	1963-1992	39
w85002	Endrick Water	Gaidrew	248500	686600	220	378	1963-1981	119
w86002	Eachaig	Eckford	214000	684300	140	446	1968-1972	80
w93001	Carron	New Kelso	194200	842900	138	404	1979-1992	187
w96001	Halladale	Halladale	289100	956100	205	321	1975-1992	140
w96002	Naver	Apigill	271300	956800	477	916	1978-1992	154
w97002	Thurso	Halkirk	313100	959500	413	557	1972-1992	107

## Appendix B

Values (SI units) for the 24 network descriptor variables defined in Chapter 8.

### 300km total length networks

Code	NUM_Link	AV_LLength	AV_LSlope	MAX_Dist	Drop_MaxD	Dist_MaxW	MAX_Width	MEAN_Dist
3001	381	720.4	0.03	30230.6	215.9	16625	18	16223.5
3003	505	634.8	0.07	26012.3	391.0	14675	27	15003.4
4001	506	648.7	0.14	37325.6	653.0	17150	18	17712.9
4003	292	941.6	0.06	38602.6	611.4	16000	18	18684.2
7002	527	621.4	0.08	26721.5	489.3	21025	27	14780.8
7003	359	853.5	0.04	37613.5	437.2	12075	18	17596.4
8004	461	615.9	0.13	39146.1	912.1	20425	22	19375.4
8010	331	911.6	0.06	36243.4	544.4	23775	17	18590.5
9002	307	946.8	0.04	31947.2	231.5	18875	17	16015.1
10002	170	1654.2	0.01	36962.0	166.0	29325	13	19887.8
10003	215	1285.3	0.02	42106.8	206.3	24525	16	20375.1
11001	197	1374.5	0.03	37307.0	359.2	25625	15	21095.4
12002	402	778.7	0.08	27957.2	476.9	23225	23	16788.3
15007	387	812.9	0.14	31028.6	621.7	25075	23	17461.9
15017	287	1016.1	0.07	29975.0	624.8	16025	20	14687.5
16004	292	1066.4	0.12	28990.9	504.4	15300	19	16160.9
17002	304	989.9	0.04	31857.9	337.9	20275	20	18545.0
18002	473	675.9	0.11	48398.2	535.1	36125	16	26195.9
18003	365	784.6	0.15	24579.4	392.0	11475	25	12472.2
18005	347	947.0	0.05	27557.2	332.1	19800	25	16487.4
20001	297	1076.2	0.05	35239.8	233.7	22825	25	20675.2
21010	345	876.0	0.07	44191.7	524.7	34575	18	26079.9
21022	430	720.3	0.05	27531.5	331.3	19675	24	15315.9
21031	353	920.0	0.06	47672.5	518.4	22300	14	26260.1
23004	430	766.3	0.08	25988.6	505.0	13325	24	13987.9
23007	404	758.4	0.06	44652.4	483.7	32875	18	25136.6
23015	507	617.8	0.07	26332.1	387.1	17375	22	13804.5
24008	534	591.4	0.09	19946.0	416.3	8375	27	11000.1
76008	121	2540.5	0.02	57626.4	447.4	17325	18	25065.4
77001	424	770.7	0.06	27440.1	413.4	21375	26	17437.4
77005	308	1045.5	0.05	37229.8	476.2	27075	21	21967.6
79002	545	582.7	0.10	30068.5	460.1	18825	27	15993.8
80001	463	676.4	0.03	32723.5	308.5	21750	18	17810.6
81002	545	573.5	0.09	26826.5	399.5	14075	25	14042.3
81003	490	667.1	0.03	31388.6	342.8	14950	21	15721.7
82001	441	676.0	0.06	39062.6	435.9	27775	20	19992.9
82003	567	579.0	0.08	37688.4	483.0	27125	23	18371.3
83005	319	1028.4	0.03	30752.8	250.9	17975	27	13878.5
83006	435	757.8	0.05	38316.2	328.8	30575	16	21955.9
84001	299	1091.0	0.06	26483.0	155.6	14500	28	13161.5
84013	379	856.4	0.03	33735.7	234.0	19275	22	19995.1
86002	532	594.7	0.18	25803.1	494.5	23725	24	13772.7
93001	662	463.2	0.18	25183.0	625.5	20725	26	14274.4
96001	539	594.9	0.03	26128.0	377.3	13225	26	12345.5
96002	501	585.8	0.05	21646.0	355.9	14900	31	13089.7
97002	477	685.2	0.03	33767.0	323.9	22825	25	17974.8

## Appendix B – 300km total length networks (continued)

Code	STD_Dist	KURT_Dist	SKEW_Dist	MEAN_Drop	STD_Drop	KURT_Drop	SKEW_Drop	MAX_Drop
3001	7307.6	-0.8228	-0.1020	117.3	67.8	2.4316	1.1965	518.6
3003	6053.2	-0.4552	-0.4415	240.9	121.6	3.6540	1.4483	768.5
4001	9559.3	-1.0127	0.0922	264.9	163.6	-0.4038	0.4838	809.8
4003	8572.6	-0.3483	0.2629	290.9	135.0	-1.0706	-0.0084	611.4
7002	5990.7	-0.8987	-0.1161	309.2	117.0	-0.0575	-0.6082	554.4
7003	9382.3	-0.9313	0.2223	147.5	99.8	-0.9092	0.2587	437.2
8004	9243.1	-0.6904	0.0233	324.5	193.0	0.1788	0.7996	932.3
8010	8660.5	-0.8937	0.0174	238.6	130.1	-0.8316	0.1909	609.2
9002	7944.4	-0.8790	0.1578	82.4	52.2	0.0725	0.7172	300.9
10002	9075.1	-1.0522	-0.1010	56.3	35.5	-0.4727	0.5411	166.0
10003	9779.3	-0.8729	0.0327	77.6	38.5	-0.0201	0.2561	209.1
11001	9328.9	-1.0153	-0.1716	92.5	66.3	0.5355	0.9199	359.2
12002	6612.9	-0.7488	-0.4713	227.4	136.0	-0.8964	0.0843	618.9
15007	7029.9	-0.7888	-0.3069	276.6	153.8	-0.3504	0.5128	814.6
15017	6879.5	-0.8519	0.0245	225.6	111.7	-0.0934	0.5473	624.8
16004	6740.9	-0.7566	-0.2659	220.7	166.7	-1.0541	0.3584	721.5
17002	6786.6	-0.3211	-0.3149	76.9	78.2	0.3347	1.1947	352.8
18002	12879.8	-1.1128	-0.2005	246.9	154.0	-0.8883	0.1358	645.5
18003	5526.4	-0.8160	0.0393	175.5	148.9	-0.2873	0.7442	686.7
18005	6182.9	-0.3821	-0.5961	117.5	91.7	-0.0629	0.8245	439.9
20001	7008.3	0.0434	-0.6940	137.1	91.2	-0.2653	0.6627	450.5
21010	10559.6	-0.4986	-0.4692	198.6	93.4	-0.0253	0.4198	524.7
21022	6029.3	-0.5098	-0.3213	138.2	71.9	-0.3659	0.3808	353.9
21031	12153.2	-0.9209	-0.1298	161.1	129.6	-0.4619	0.7400	542.1
23004	5751.2	-0.5424	-0.1612	248.4	106.2	-0.4309	0.0760	551.8
23007	12253.8	-1.0884	-0.3260	230.1	123.6	-0.7597	0.2029	512.7
23015	6275.8	-0.9101	-0.1777	159.1	86.2	-0.4772	0.4074	402.1
24008	4816.1	-1.0238	0.0051	229.6	103.7	-0.5826	0.0942	482.7
76008	15402.1	-0.8479	0.6154	166.1	100.1	-0.5198	0.2999	503.6
77001	6078.3	-0.1364	-0.7401	146.4	76.8	0.6795	0.7700	457.2
77005	8983.6	-0.7443	-0.5154	148.9	95.7	0.5060	0.9098	476.2
79002	6342.3	-0.4056	-0.3105	227.7	109.8	-0.8384	-0.0542	521.6
80001	8054.6	-0.9548	-0.1254	140.4	63.8	-0.2498	-0.0178	355.7
81002	6293.9	-0.6674	-0.0237	239.7	135.0	-0.4864	0.3246	690.7
81003	7039.2	-0.7737	-0.0379	145.0	62.4	0.3795	0.3258	385.9
82001	8623.6	-0.7453	-0.1934	166.4	93.8	-0.6198	0.1272	501.4
82003	8622.2	-0.8145	0.1425	180.0	100.3	-0.0937	0.6198	527.4
83005	6543.6	-0.6833	0.1450	133.6	78.7	-1.0630	0.0302	333.8
83006	9609.6	-0.8337	-0.4789	176.3	76.7	-0.1766	0.2037	455.6
84001	4955.0	-0.2546	-0.1890	116.2	125.6	0.3547	1.2400	487.5
84013	7336.2	-0.2189	-0.5082	148.4	60.9	0.2308	0.3707	360.8
86002	7181.6	-1.2536	-0.1806	217.4	163.2	-0.9809	0.1906	680.9
93001	5718.9	-0.6717	-0.3043	258.0	172.0	-0.1861	0.5554	814.8
96001	6015.8	-0.7932	0.1352	132.7	65.2	1.8849	1.0714	377.3
96002	4977.4	-0.4837	-0.6050	128.4	77.8	-0.1170	0.5378	426.6
97002	6967.4	-0.5730	-0.2530	77.8	65.4	1.1090	1.3198	323.9

## Appendix B – 300km total length networks (continued)

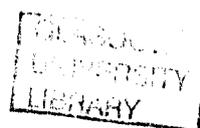
Code	MEAN Eng	STD Eng	KURT Eng	SKEW Eng	MAX Eng	MAX Order	AV Bif	NUM Source
3001	0.0073	0.0028	4.8118	1.4938	0.0230	5	3.76	193
3003	0.0176	0.0096	21.4976	3.6611	0.1102	5	4.11	256
4001	0.0201	0.0212	11.2864	3.0044	0.1635	5	4.43	254
4003	0.0169	0.0090	67.4732	5.8567	0.1793	4	5.31	148
7002	0.0222	0.0120	218.5959	10.8473	0.3033	5	4.45	269
7003	0.0076	0.0038	-0.3589	0.3519	0.0185	5	3.90	181
8004	0.0175	0.0087	30.1850	3.6494	0.1440	5	3.96	234
8010	0.0127	0.0042	1.5755	0.9075	0.0302	5	3.62	167
9002	0.0057	0.0047	15.3383	3.4752	0.0385	5	3.75	155
10002	0.0027	0.0011	1.3892	0.8121	0.0083	5	3.21	86
10003	0.0044	0.0025	2.3115	1.2992	0.0178	4	4.84	108
11001	0.0041	0.0023	9.9915	2.3805	0.0195	4	4.63	99
12002	0.0127	0.0064	3.5593	0.9109	0.0603	5	3.80	202
15007	0.0178	0.0126	3.9600	1.7497	0.0877	5	3.90	196
15017	0.0170	0.0077	1.1010	1.0316	0.0546	5	3.66	146
16004	0.0146	0.0124	3.0743	1.4434	0.0819	4	5.54	148
17002	0.0039	0.0041	11.2770	2.7839	0.0331	5	3.54	154
18002	0.0118	0.0141	11.1980	3.1266	0.1071	5	4.26	238
18003	0.0156	0.0150	3.0059	1.6295	0.0802	5	3.97	186
18005	0.0070	0.0048	1.0981	0.8354	0.0298	4	6.14	177
20001	0.0062	0.0031	0.2368	0.7301	0.0177	5	3.67	150
21010	0.0085	0.0054	21.0179	3.9345	0.0522	5	3.66	173
21022	0.0094	0.0066	88.0276	7.9589	0.1079	5	4.13	216
21031	0.0059	0.0041	2.6809	1.1839	0.0268	5	3.94	177
23004	0.0195	0.0111	63.5152	5.8791	0.2030	5	4.23	216
23007	0.0098	0.0056	43.9477	5.2909	0.0764	5	3.83	204
23015	0.0126	0.0073	11.3920	2.6423	0.0588	5	4.04	255
24008	0.0218	0.0080	1.4751	0.9358	0.0662	5	4.08	270
76008	0.0072	0.0045	3.6508	1.8621	0.0281	4	3.98	61
77001	0.0088	0.0084	196.9802	12.7648	0.1699	5	4.00	214
77005	0.0063	0.0021	0.6542	0.6863	0.0163	5	3.70	155
79002	0.0141	0.0048	0.6869	0.4476	0.0362	5	4.41	278
80001	0.0081	0.0025	4.6968	1.0941	0.0319	5	3.96	235
81002	0.0177	0.0095	2.2252	1.3737	0.0629	6	3.18	275
81003	0.0098	0.0038	22.2136	3.6889	0.0433	5	4.07	247
82001	0.0085	0.0085	124.7743	10.3471	0.1380	5	3.91	223
82003	0.0106	0.0073	72.9977	6.5761	0.1238	5	4.13	288
83005	0.0097	0.0053	-0.1688	0.6289	0.0250	5	3.60	160
83006	0.0090	0.0041	2.4269	1.6586	0.0302	5	4.14	220
84001	0.0082	0.0084	1.3722	1.3813	0.0440	5	3.52	151
84013	0.0081	0.0056	152.0117	10.0402	0.1200	5	4.00	191
86002	0.0201	0.0319	33.3070	5.0433	0.3375	5	4.50	275
93001	0.0187	0.0144	11.9618	2.3509	0.1367	5	4.39	338
96001	0.0121	0.0067	14.5768	3.1138	0.0748	5	4.16	274
96002	0.0099	0.0060	18.8325	2.7894	0.0754	5	4.05	254
97002	0.0038	0.0021	0.7331	1.0731	0.0165	5	4.00	242

## Appendix B (continued)

## 600km total length networks

Code	NUM_Link	AV_LLength	AV_LSlope	MAX_Dist	Drop_MaxD	Dist_MaxW	MAX_Width	MEAN_Dist
3001	1036	636.4	0.06	37811.5	635.0	24475	37	18391.5
3003	934	646.1	0.07	30549.3	641.8	16875	40	16350.1
6003	850	776.5	0.12	36580.0	772.2	23450	31	20875.8
7002	869	691.8	0.08	45734.6	575.3	35750	25	24940.3
8004	862	667.1	0.11	58007.2	1039.9	26975	30	28996.5
9002	619	915.8	0.06	51447.3	438.7	28175	26	28440.5
11001	715	920.1	0.08	69835.6	542.3	46875	22	35163.4
12002	764	717.6	0.12	34979.2	894.0	23325	36	19734.3
15007	752	816.8	0.13	59673.7	734.6	53725	23	32416.3
16004	713	897.8	0.11	35664.3	519.9	18475	37	19645.2
17002	549	1073.8	0.04	46339.6	422.3	28825	30	27029.7
21008	791	799.9	0.06	41889.0	380.0	27275	33	25440.1
21010	830	735.7	0.10	42187.5	377.8	19825	32	24196.7
23004	758	774.9	0.08	36926.2	676.4	10825	24	19515.2
23015	869	714.7	0.06	49337.3	439.0	24750	23	27041.2
24008	1024	619.5	0.09	42357.6	543.0	30650	28	24450.3
77001	975	658.6	0.08	46703.8	428.4	36800	33	28289.0
78003	822	794.1	0.10	47922.5	414.3	34025	37	27950.5
81002	971	670.2	0.07	43037.5	453.6	27125	31	23670.2
82003	939	668.1	0.06	48553.8	506.9	24600	38	24767.7
83005	507	1157.6	0.03	49462.4	277.8	26875	31	26539.8
84013	964	676.6	0.09	41257.6	368.5	22925	35	21792.5
96002	979	672.0	0.05	31904.3	359.5	25775	37	18042.7
97002	687	810.6	0.02	57255.1	399.1	45425	25	31099.3

Code	STD_Dist	KURT_Dist	SKEW_Dist	MEAN_Drop	STD_Drop	KURT_Drop	SKEW_Drop	MAX_Drop
3001	8511.8	-0.6861	-0.1711	115.4	100.2	0.7756	0.8974	635.0
3003	6468.4	-0.4908	-0.2899	225.7	117.0	3.0395	1.2126	773.0
6003	8602.7	-0.8478	-0.2526	254.0	153.9	-0.1675	0.6785	842.1
7002	11636.4	-1.0428	-0.2202	299.4	154.5	-1.0549	-0.0624	640.4
8004	13171.3	-0.7212	-0.0230	350.1	196.5	0.6203	0.8037	1080.1
9002	11633.8	-0.7951	-0.0219	214.7	114.5	-0.4124	0.3579	592.7
11001	18431.3	-1.1116	-0.0674	235.4	131.1	-0.3612	0.4528	682.7
12002	7490.0	-0.4448	-0.3034	285.8	170.8	0.5869	0.8006	945.3
15007	16325.0	-1.2494	-0.2134	371.6	177.7	-0.4076	0.1740	976.1
16004	7239.4	-0.4368	-0.1585	266.6	165.4	-0.9011	0.2143	800.0
17002	10601.4	-0.5494	-0.4769	143.2	81.3	0.3342	0.9670	437.2
21008	9644.0	-0.4801	-0.5315	173.4	77.3	0.2179	0.2831	519.5
21010	8982.2	-0.5486	-0.2604	165.0	111.3	0.4765	0.8878	582.0
23004	9926.3	-1.2864	0.0423	260.7	128.4	-0.6510	0.0911	677.6
23015	12045.5	-0.9096	-0.2138	162.4	86.8	0.2301	0.6979	454.0
24008	10471.8	-0.9114	-0.2224	275.9	132.4	-0.6010	0.0723	609.4
77001	10421.7	-0.3511	-0.5308	206.9	99.4	-0.4264	0.3419	566.6
78003	11215.9	-0.6242	-0.5354	198.8	142.7	0.0540	0.7342	703.0
81002	10089.2	-0.7557	-0.2861	194.7	125.3	0.1574	0.6992	700.9
82003	9677.4	-0.1643	-0.0426	169.5	91.1	0.9703	0.9068	551.3
83005	10417.1	-0.3693	-0.4702	144.1	85.8	-1.1925	0.0462	360.7
84013	9573.6	-0.7324	-0.0690	138.1	79.5	0.2726	0.7258	465.3
96002	7723.7	-0.9245	-0.3274	141.9	98.0	1.9124	1.1554	682.8
97002	14550.9	-0.9885	-0.3962	114.1	71.8	1.2687	1.0696	399.1



## Appendix B – 600km total length networks (continued)

Code	MEAN_Eng	STD_Eng	KURT_Eng	SKEW_Eng	MAX_Eng	MAX_Order	AV_Bif	NUM_Source
3001	0.0074	0.0078	3.5400	1.8014	0.0428	5	5.04	528
3003	0.0145	0.0092	198.7578	10.5138	0.2502	6	3.58	474
6003	0.0139	0.0101	3.1699	1.5515	0.0694	5	4.57	432
7002	0.0126	0.0061	14.6927	2.4781	0.1100	5	4.60	442
8004	0.0126	0.0074	67.1262	6.6978	0.1316	5	4.64	436
9002	0.0074	0.0027	25.2915	2.6292	0.0433	6	3.25	312
11001	0.0075	0.0043	6.8953	2.4190	0.0323	5	4.38	360
12002	0.0147	0.0085	6.3798	1.6515	0.0885	6	3.43	387
15007	0.0159	0.0134	1.5790	1.4908	0.0898	5	4.54	379
16004	0.0139	0.0084	0.6414	0.6340	0.0525	5	4.55	364
17002	0.0057	0.0033	2.5900	1.7853	0.0188	5	4.11	278
21008	0.0072	0.0031	8.8327	1.9391	0.0325	5	4.54	400
21010	0.0071	0.0053	15.4456	2.7376	0.0682	5	4.54	419
23004	0.0146	0.0066	1.2100	0.0907	0.0907	5	4.50	379
23015	0.0068	0.0049	60.0256	5.4850	0.0990	5	4.87	438
24008	0.0115	0.0041	5.9592	1.4692	0.0401	5	4.91	517
77001	0.0076	0.0033	2.0887	1.2786	0.0230	5	4.77	494
78003	0.0066	0.0037	-0.0983	0.5489	0.0217	5	4.63	419
81002	0.0086	0.0063	39.9844	4.1158	0.1293	6	3.54	490
82003	0.0072	0.0044	26.0933	4.0746	0.0539	5	4.68	477
83005	0.0052	0.0023	-1.1041	-0.1494	0.0107	5	4.00	254
84013	0.0071	0.0056	42.9889	5.1426	0.0725	5	4.86	487
96002	0.0085	0.0056	1.5526	1.1403	0.0404	6	3.59	497
97002	0.0037	0.0014	7.6289	2.0967	0.0147	5	4.42	348