Relationship between Juvenile Salmonid Populations and Catchment Features

R&D Technical Report   W2-065/TR

Relationship between Juvenile Salmonid Populations and Catchment Features

R&D Technical Report W2-065/TR

Alexander Coley

Dissemination Status
Internal: Released to Regions
External: Publicly Available

Statement of Use
This technical report demonstrates how GIS can be used effectively to support fisheries management. GIS methods for extracting and analysing catchment data are presented. An example model has been created showing how juvenile salmonid densities are related to catchment features, but the approach could be applied to any in stream features. This report will be of use to Agency staff who use GIS to model in stream ecological or physical features. It will also be of use to fisheries staff, as it offers an insight into the catchment features which control juvenile salmonid production.

Keywords
Salmon, trout, GIS, geographical information system, catchment, hydrological analysis, spatial analysis, habitat, fish, modelling, mapping

CONTENTS

EXECUTIVE SUMMARY

The Environment Agency has a statutory duty to maintain, improve and develop freshwater fisheries. To fulfil this duty, information to describe and understand fish stocks is required. Current salmonid management tools rely upon the measurement of site and catchment features to aid the interpretation of fish survey data.

Manual methods of extracting site and catchment features are inherently time consuming and may be prone to human error. A huge advantage may be gained if such methods could be automated. The use of a computer system and more specifically a Geographical Information System (GIS) allows for the extraction of attributes for any sites of interest on the UK river network. This enables the refinement, and improvement, of the existing salmonid management tools and assists with the development of new tools. With the introduction of a GIS, new data may be obtained allowing for the calculation of attributes that would not currently be feasible to determine using standard Ordnance Survey maps.

This project provides an analysis on new and existing GIS techniques for the measurement of site and catchment features. The GIS provides a data source, a modelling platform and a mapping tool that aim to supply useful improvements to existing fisheries management techniques and have wider usefulness outside the fisheries function. A comparison between data derived from the GIS and existing manual techniques was undertaken. It was shown that in the majority of cases the data was comparable, with successful derivation of the basic map based data using the GIS.

There are a number of significant advantages of the GIS methods over existing manual methods.
1. Mapping and the ability to view distributions spatially
2. Speed of attribute derivation is greatly increased when deriving for a large number of sites.
3. The methodology can be standardised and removes some of the scope for error.
4. The methods allow for the calculation of attributes for every "cell" on the river network not just sampled sites.
5. The methods allow for the application of any regression models within the GIS.

These GIS derived attributes have been shown to be suitable for any Environment Agency methods that currently use or involve parameters that are taken from paper based maps. There is however, a significant limitation that must be overcome: Calculating Site Gradient in a method comparable to the manually derived method is currently not available in a fully automated method.

Example models have been created showing how juvenile salmonid densities are related to GIS derived catchment features. The basic models comprised of topography and distance based attributes, showed a reasonable predictive ability for the national trout fry dataset (non-pristine and pristine sites) and for the national salmon fry dataset. The salmon fry model was less successful in its predictive power than the trout fry model. The models produced using landcover showed that for salmon fry and trout fry most influential parameter was the proportion of urbanised land upstream, although deciduous forest and arable land were also shown to be significant predictors within the models.

1. INTRODUCTION

Current salmonid management tools rely upon the measurement of site and catchment features to aid the interpretation of fish survey data. A number of Environment Agency tools rely on the use of site and catchment features, which are presently derived manually from paper based maps. These methods include HABSCORE (Barnard, Wyatt and Milner, 1995, Wyatt *et al.*, 1995b) and the Fisheries Classification Scheme (Hay *et al.*, 1995, Mainstone *et al.*, 1994, Wyatt and Lacey, 1998). A number of tools in development such as the River Fisheries Habitat Inventory project (Wyatt, 2001) and the Salmon Lifecycle project (Wyatt and Barnard, 1997a) will rely on computer based methods to derive site and catchment features.

The requirement or importance of using map derived or GIS based variables in understanding spatial variability can be clearly seen with the development of the HABSCORE models. For example the model for 0+ trout based on only catchment variables explained 41.3% of the spatial variance as opposed to 63.1% for the full model (Wyatt *et al.* 1995). With the introduction of GIS to the fisheries methodologies some important improvements could be made. For example there could be a reduction in the time required and there could be an increase in the precision and effectiveness of the methods.

Manual methods of extracting site and catchment features are inherently time consuming and may be prone to human error. A huge advantage may be gained if such methods could be automated. The use of a computer system and more specifically a Geographical Information System (GIS) would allow for the extraction of attributes for any sites of interest on the UK river network. This would enable the refinement, and possible improvement, of the above-mentioned methods. With the introduction of a GIS, new data may be obtained allowing for the calculation of attributes that would not currently be feasible to determine using standard Ordnance Survey maps.

This project has many links to the ongoing River Fisheries Habitat Inventory (RFHI) Project (Wyatt and Barnard 1997c and Wyatt 2001). The development of a fisheries habitat inventory is under a two-tier system. The two tiers would be:

1. An assessment of "Reach Suitability" linked to the Fisheries Classification Scheme (FCS) and is based entirely on map-derived variables.
2. An assessment of "Relative Habitat Quality" based on field-derived variables. With juvenile salmonids this would have links to the HABSCORE methodology.

The combination of these two tiers would provide the overall quality of a reach for a particular fish species. The aim of this study was to assess whether GIS can make a significant contribution to the first tier of the RFHI. By providing an effective method for calculating the required map based variables and investigating whether new datasets are likely improve the models.

This project will provide an analysis on new and existing GIS techniques for the measurement of site and catchment features. The GIS will provide a data source, a modelling platform and a mapping tool that aim to supply useful improvements to existing fisheries management techniques and have wider usefulness outside the fisheries function.

1.1 Objectives

1.1.1 Overall

To improve the cost-effectiveness of fisheries management methods through improved measurement of catchment features and to aid interpretation of fish survey data.

1.1.2 Specific

1) To describe the catchment and site features pertaining to Environment Agency salmonid monitoring sites using a Geographical Information System (GIS).
2) To update and extend the existing Salmonid Fisheries Science Group (SFSG) GIS databases of juvenile salmonid survey data.
3) To describe relationships between salmonid distribution, abundance and variability (2) and site characteristics (1), by development of a model.
4) Apply the model (3) to unsampled areas to generate expected densities based on catchment features.
5) Make recommendations for the refinement of existing Environment Agency methods (e.g. impact assessment, spawning targets, HABSCORE, Fisheries Classification Scheme (FCS), Salmon Action Plans, Fisheries Action Plans) based on 3).

1.1.3 Purpose of Report

The purpose of this report is:
- To review and assess new and existing GIS techniques for the measurement of site and catchment features. (Section 2).
- To describe the current and new GIS techniques. (Section 2).
- To compare the GIS techniques to existing techniques. (Section 3).
- To look at the suitability and effectiveness of existing and new catchment data in terms of understanding juvenile salmonid populations. (Section 4).

2. USING GIS TO EXTRACT CATCHMENT AND SITE BASED VARIABLES

2.1 Introduction

There are a number of fisheries and other methods used within the Environment Agency that rely on some level of manually map-derived data. Such methods include HABSCORE (Barnard, Wyatt and Milner, 1995), the Fisheries Classification Scheme (Mainstone *et al.*, 1994), PHABSIM (Johnson *et al.*, 1993), River Habitat Survey (Environment Agency 1996) and RIVPACS (Wright *et al.*, 1994). All of these methods use to a lesser or greater degree the same map based attributes, and these are listed in Table 2.1). These variables included altitude, distance to source, distance to mouth, catchment area and gradient.

Table 2.1   Map based variables used in selected Environment Agency methods

Method	Map Based Attributes	Field / Subjective Measurements that could be Map/GIS based
HABSCORE	Site altitude, site gradient, catchment area, link number, catchment gradient, distance to source, distance to tidal limit, downstream link number	Landuse upstream, access (for migratory fish), conductivity, flow category
Fisheries Classification Scheme	Gradient, distance to source, catchment area	Access (for migratory fish)
River Habitat Survey	Site altitude, solid geology, distance to source, altitude of source, site gradient, drift geology, water quality class, flow regime / category,	Flood plain landuse
RIVPACS	Site altitude, site gradient, flow category, distance to source, water quality	
PHABSIM	Site/reach altitude, catchment area, mean catchment altitude, mean catchment length, mean catchment slope, stream length, reach gradient	Flow regime

Being manually derived variables, a considerable amount of time and effort may be spent on their calculation. There have been recent increases in the availability of datasets and software that are necessary for conducting catchment scale assessments (Johnson and Gage, 1997). Automated systems are faster, they will generally provide more data and present many improvements over existing methods.

A Geographical Information System (GIS) combines layers of information about a place to give a better understanding of that location. These layers of information can be combined depending on the purpose of study to produce maps, perform analysis, inform decisions and so on. Geographical Information Systems have developed significantly since their inception in the early 1960's. By the 1980's Environmental Systems Research Institute (ESRI) had released its ArcView product and other companies had released similar applications. As time has gone on the availability of data has grown but the limits of computing power have put constraints on the magnitude of analysis possible within a GIS. The introduction of spatial analysis tools allowed for the development of hydrological analysis applications and the extraction of catchment based attributes. This allowed for the automated delineation of the upstream catchment and the extraction of a number of catchment features. By mid to late 1990's computing power had increased significantly and the release of ArcView 8 has resulted in further

advancements on the capabilities of a GIS. The ability to integrate spatial data with models has expanded the ability of GIS for analysing and predicting environmental phenomena (Johnson and Gage, 1997). Recent improvements have made the use of GIS not just a possible tool but also an effective tool for performing environmental and hydrological analysis.

This project will apply these recent advancements to automate or improve measurement of catchment features will aid interpretation of fish survey data and improve the cost-effectiveness of fisheries management methods.

2.2 Methods

2.2.1 Geographical Information System

The standard Environment Agency GIS package is ArcView (Environmental Systems Research Institute) this will shortly be updated to the current release, which is ArcView 8.3. For this project ArcView 8.3 was used as it provided a more robust system than earlier versions and prevented methods being developed for a system that would become obsolete within the Environment Agency in the coming years.

ArcView 8 looks very different and in many ways, it behaves differently from the commonly used ArcView 3.x. ArcView is a version of ArcGIS 8 along with ArcEditor, and ArcInfo. Their level of functionality increasing from least functionality in ArcView to most functionality in ArcInfo. ArcInfo has more tools and analysis options than both ArcEditor and ArcView, ArcEditor has more functionality than ArcView, and ArcView has the least functionality of all three.
ArcMap, ArcCatalog, and ArcToolbox are the three application environments that make up all these versions of ArcGIS. ArcView, ArcEditor, and ArcInfo are each made up of these three applications:

1. **ArcMap** is the application for display and editing of features, mapmaking, and analysis.
2. **ArcCatalog** is the application for exploring and managing data and editing metadata or file information.
3. **ArcToolbox** is the application for performing operations such as data conversion and geoprocessing tasks.

ArcView was used for this study. The relationships between the ArcView components used within this project are displayed in Figure 2.1.

Within ArcView there are a number of software extensions that add extra functionality. One such extension is Spatial Analyst, which was used to add specialised tools and functionality to ArcView, such as performing spatial analytical tasks. Spatial Analyst is activated within ArcView and is accessed with the use of the Spatial Analyst toolbar (Figure 2.2). Spatial Analyst gives ArcView the use of Raster Calculator, which is used to build and evaluate equations to perform analysis on spatial grids. More detail on the use of Raster Calculator can be found in Appendix 1.
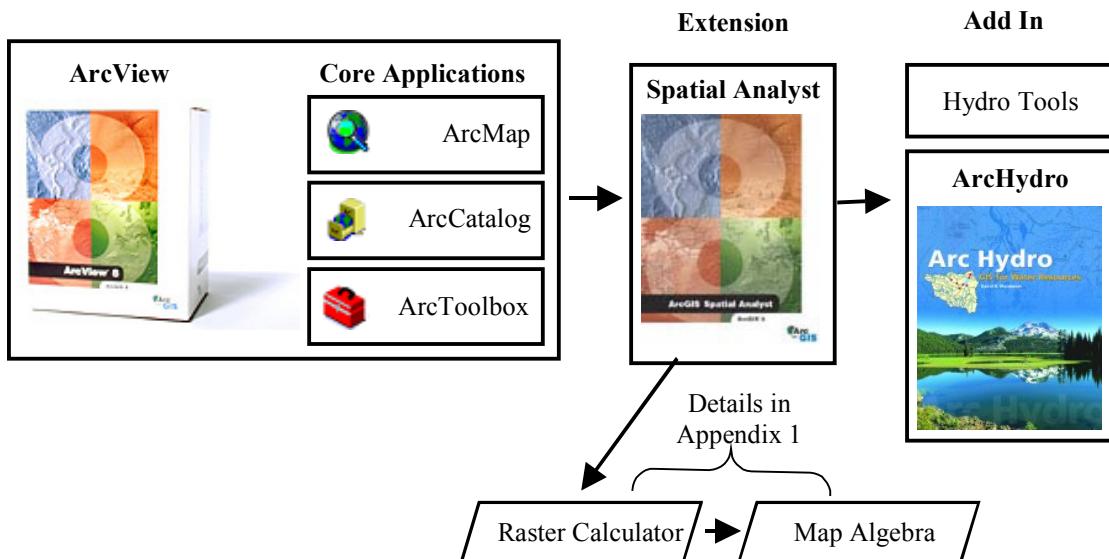
Figure 2.1   Relationship between components within ArcGIS
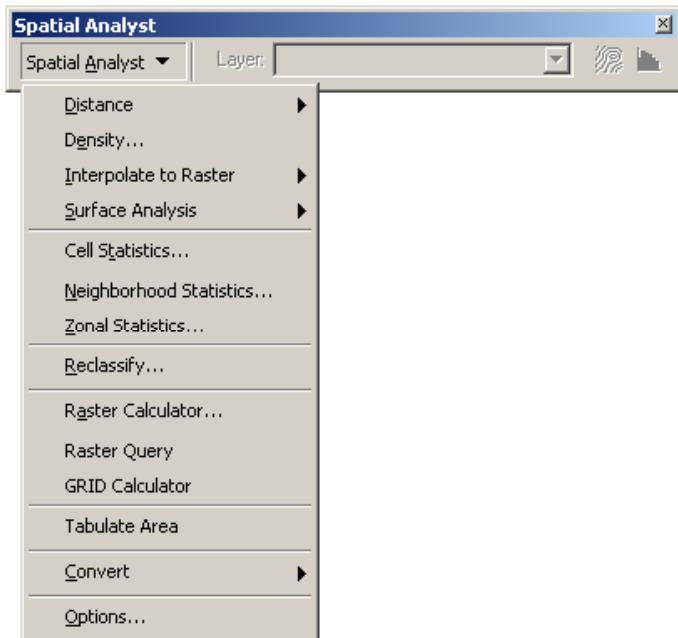
Figure 2.2   Spatial Analyst Toolbar

An additional tool that was used for the project was Arc Hydro tools, which are a toolkit (add-in), developed by ESRI and the GIS in Water Resources Consortium to support the Arc Hydro datamodel (Maidment 2002). Arc Hydro tools require Spatial Analyst and are activated within ArcView after installation and are accessed with the use of the Arc Hydro tools toolbar (Figure 2.3). There is also a basic hydrological analysis add in that comes with the ArcObject Developer Kit samples.
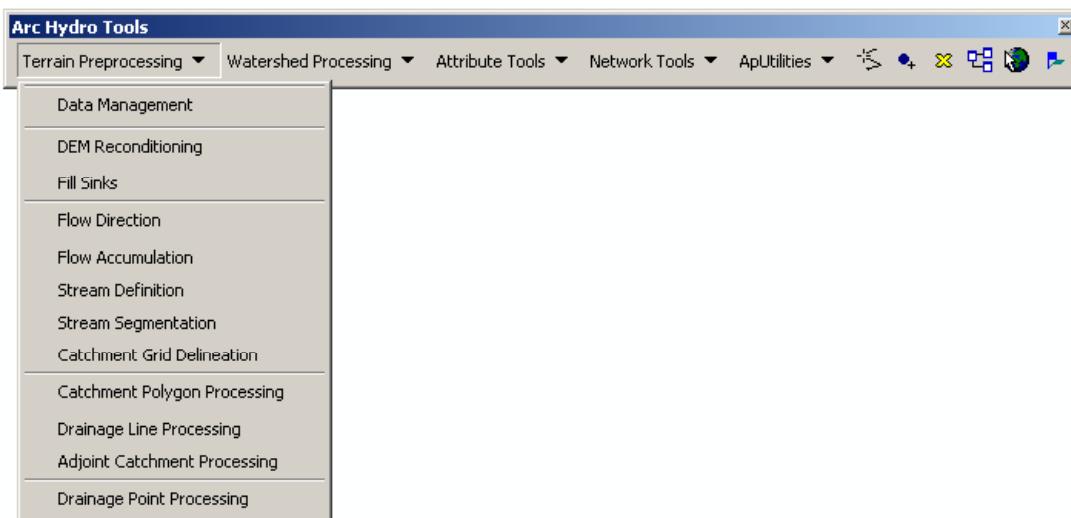
Figure 2.3   Arc Hydro Tools toolbar

2.2.2   Selection of Areas to be GIS Processed

It would be unrealistic to spatially model the whole of England and Wales in terms of restraints of time available for the preparation of GIS data and the need to repeat the methods as they develop. It was decided that a realistic coverage would include all of the Salmon Action Plan rivers (SAP's) plus other catchments that contained high numbers of fully quantitative fish survey sites. The selected area for the GIS processing can be seen in Figure 2.4. A number of catchments that had a large number of survey data were not selected for modelling for example the Tyne catchment. This was due to problems with the data or the quality of the current digital river network required for the modelling process. GIS based analysis is a dynamic process and as both methods and data improve there will be wider scope to increase the GIS processed areas to the whole of England and Wales.

■ Area Selected for GIS Processing
□ Major Hydrological Area
• Sites where fully quantitative fisheries survey data was obtainable
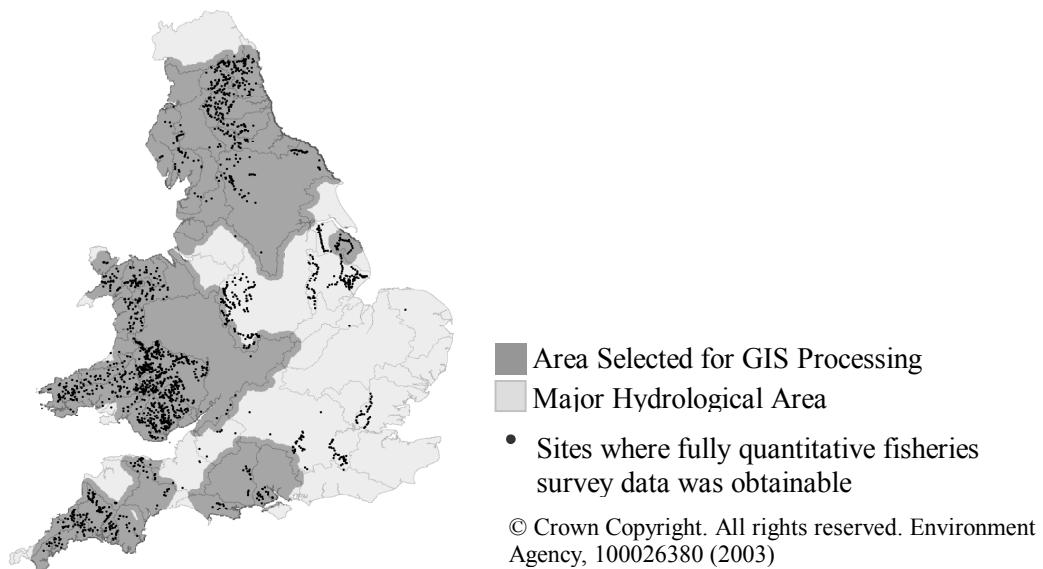
Figure 2.4   Selection of area for GIS processing

Within the modelled area there is a wide range of catchments with varying distributions of salmon and trout populations. Figure 2.5 shows the surveyed densities for the survey

sites used within the study averaged when repeat surveys exist. These have been categorised from high to low as a relative scale to give a quick overview.

Salmon Fry

Salmon Parr

Trout Fry

Trout Parr

Area Selected for Modelling
Major Hydrological Area
© Crown Copyright. All rights reserved.
Environment Agency, 100026380, (2003)

High
Low
None

Fish Densities averaged over surveys.
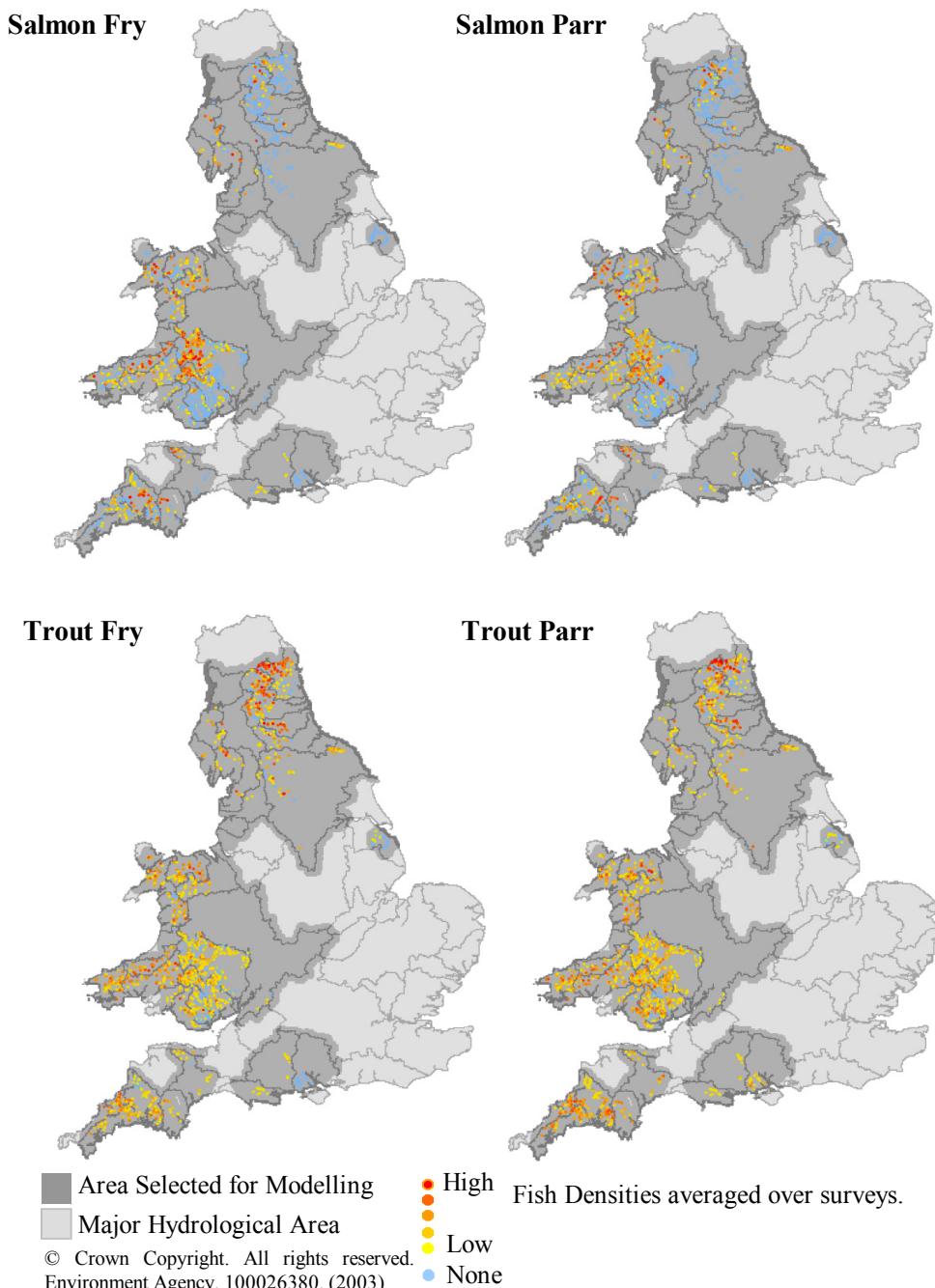
Figure 2.5   Salmon and Trout densities

2.2.3 Data Sources

There were six main datasets used within the GIS to produce the attributes required for the study, these are described below and in Figure 2.6.

A. Elevation Data in the form of a 50m Digital Elevation Model (DEM) grid from Ordnance Survey Land-Form PANORAMA data.

B. Standard Percentage Run-off (SPR) is a dimensionless variable (range 0 - 100 %) that represents the percentage of rainfall that causes the short-term increase in flow at the catchment outlet seen after the storm event (Boorman, Hollis and Lilley, 1995). This was taken from the National Soils Map for England and Wales (NSRI) this derived dataset is a 1km resolution where the 'predominant' soil in each square is given with associated attributes.

C. Rainfall from the Great Britain coverage of rainfall Isohyets (50mm) 1941-1970 average. © CEH, 1999.

D. Landcover from the landcover 2000 dataset. LCM 2000, © Centre for Ecology and Hydrology (CEH) 2002, Fuller *et al* 2002.

E. River layer – 1:50 000 vector river layer, digitised from Ordnance Survey 1:50 000 maps by CEH in 1999.

F. Lakes layer - 1:250 000 vector lakes layer, digitised from Ordnance Survey 1:250 000 maps by CEH in 1999.
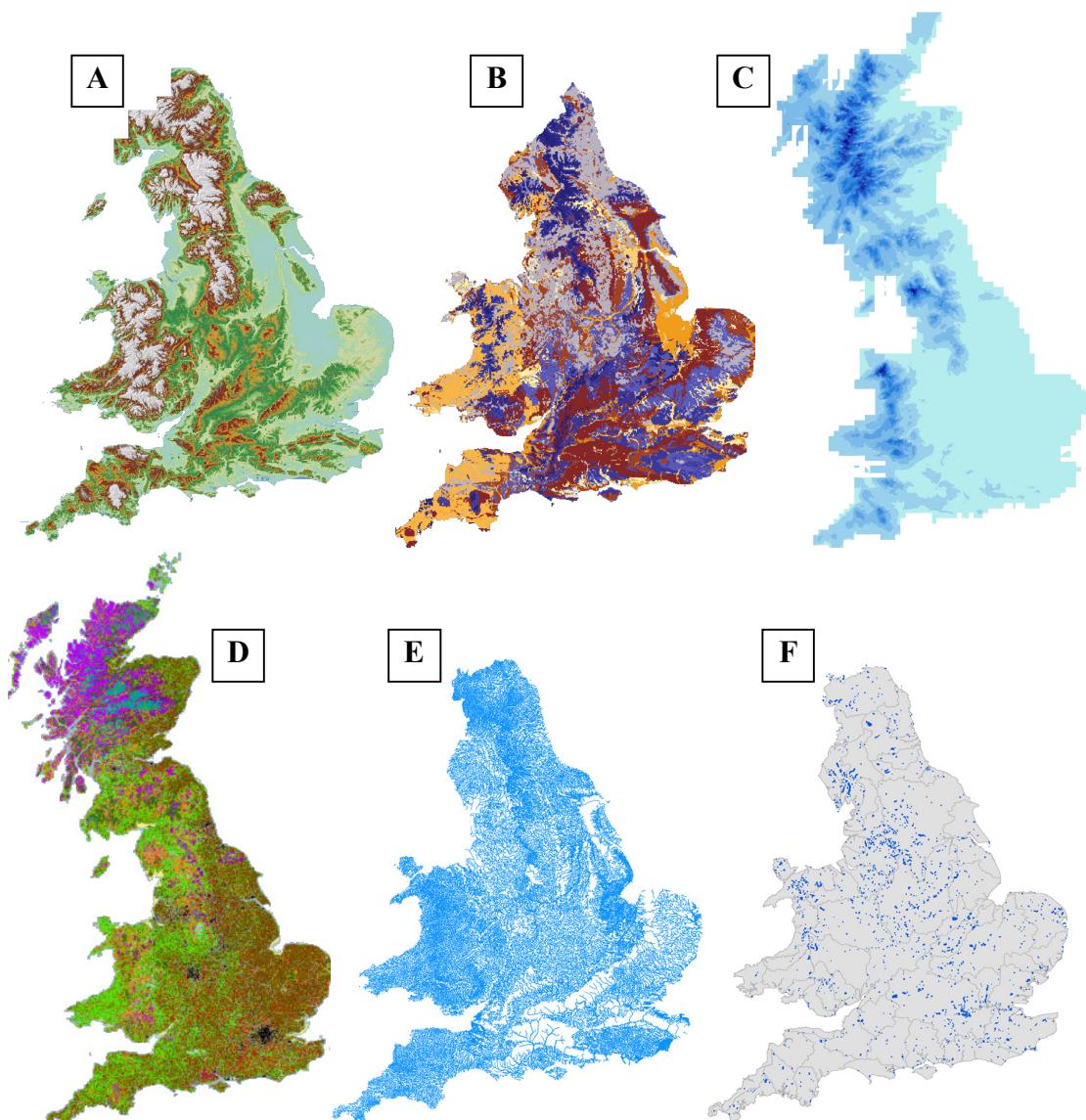
A = 50m DEM, B= SPR from Soils layer, C= Average annual rainfall,
D= Landcover 2000, E= 1:50 000 River Layer, F= 1:250 000 Lakes Layer
© Crown Copyright. All rights reserved. Environment Agency, 100026380 (2003), © Cranfield
University, © Centre for Ecology and Hydrology 1999

Figure 2.6   GIS Data sources

2.2.4   Basic Hydrological Analysis

In order to delineate catchment boundaries and to describe parameters of upstream
catchments using a GIS, we must perform a hydrological analysis. The basis for any
GIS based analysis on the riverine environment requires a digital river network. This
can be in the form of a vectored river centreline, a vectored left and right bank, or a
rasterised river network. Calculating site and catchment based features a river network
derived from the topographic landscape is called hydrological analysis. Hydrological
analysis relies on the automated extraction of topographic parameters from elevation
grids (Jenson and Domingue 1988).   This automated derivation of topographic
watershed data is faster, less subjective and provides more reproducible measurements
than traditional manual techniques (Tribe 1992, Garbrecht and Martz 2000).

The basic GIS hydrological analysis within the ArcView software is a stepped process that is described below and summarised in Figure 2.7.
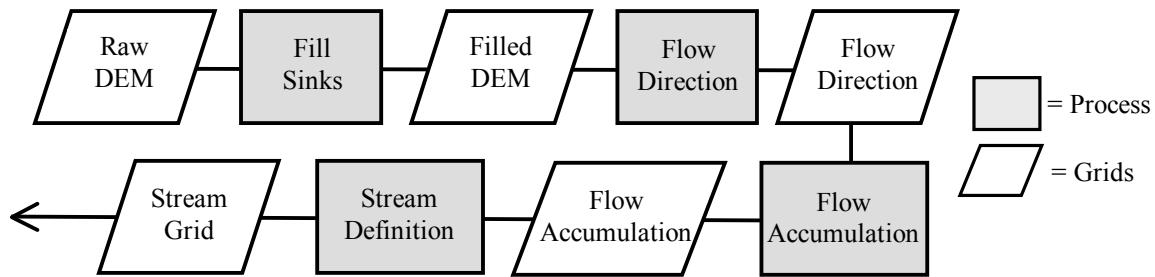
Figure 2.7   Basic Hydrological Processing

Digital Elevation Model

The starting point for the standard hydrological analysis is a digital elevation model (DEM), although other more computer intensive methods may use a triangulated irregular network (TIN) but are much less common. A DEM represents the topography of an area as a digital representation of a continuous variable over a two-dimensional surface. This is achieved with a regular array of values termed a grid as displayed in Figure 2.8. The elevation data used in this study was obtained from the Ordnance Survey as Land-Form PANORAMA Digital Terrain Models (50m horizontal resolution). The DEM was limited to the area being studied (the dark grey area in Figure 2.4) to reduce processing times.

A – Example DEM with River Network superimposed

B – Description of the DEM

Elevation   Surface   DEM

High

Low

Vertical Resolution:
to nearest meter ± 5m

Horizontal Resolution:
50m x 50m
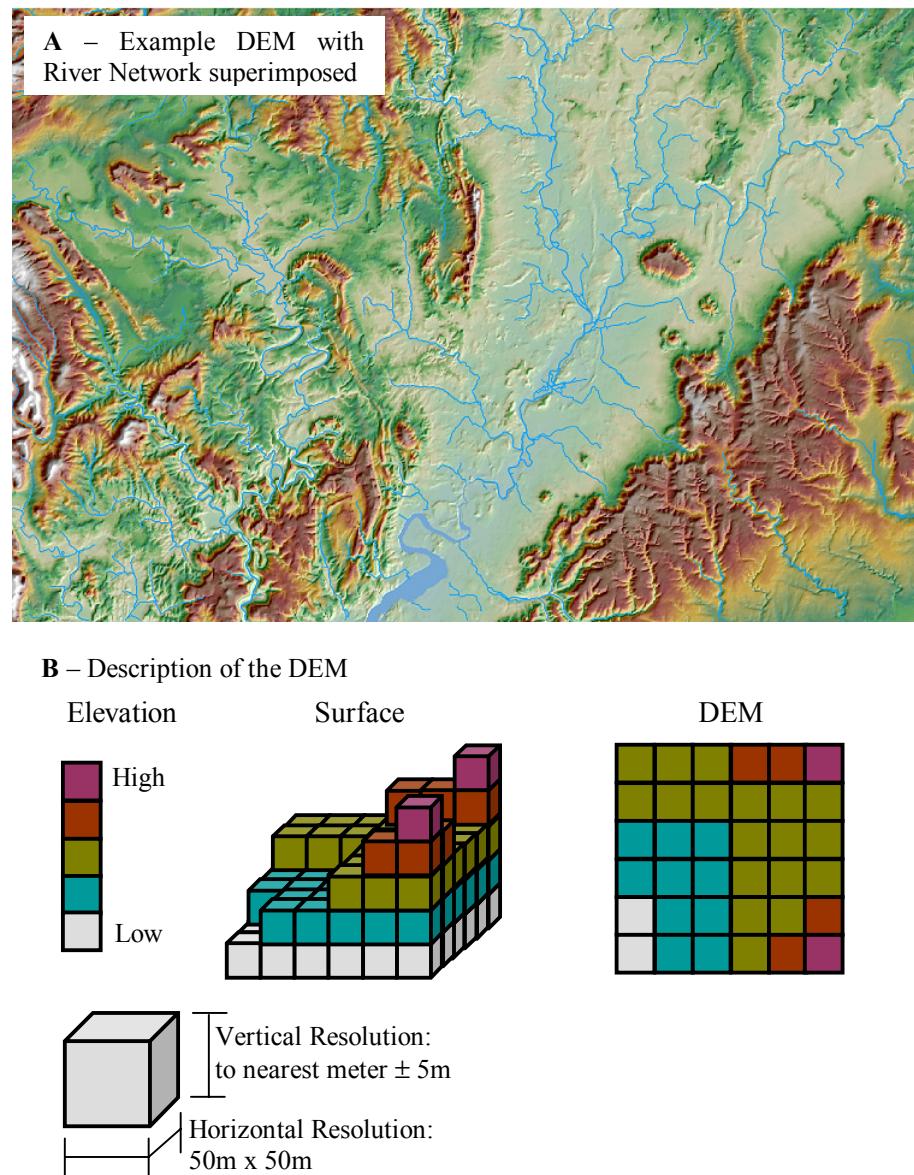
Figure 2.8   The Digital Elevation Model (DEM) raster grid

Filling Sinks

Before performing a hydrological analysis it is necessary to alter the DEM by removing or filling any sinks. A sink is a cell or cells whose flow direction cannot be assigned one of the eight valid values in a flow direction grid (Jenson and Domingue, 1988). This can occur when all neighbouring cells are higher than the processing cell (Figure 2.9), or when two cells flow into each other creating a two-cell loop.

Sinks in elevation data are most commonly due to errors in the data (Jenson and Domingue, 1988). These errors may be due to sampling effects or the rounding of elevations to integer numbers. Naturally occurring sinks in elevation data with a cell size of 10 meters or larger are rare (Mark, 1988) except for glacial areas, and should be considered as errors. As the cell size increases, the number of sinks in a dataset often also increases (ESRI, 2002). The identification and removal of sinks, when trying to create a depressionless DEM, is an iterative process. When a sink is filled, the

boundaries of the filled area may create new sinks, which then need to be filled (ESRI, 2000).
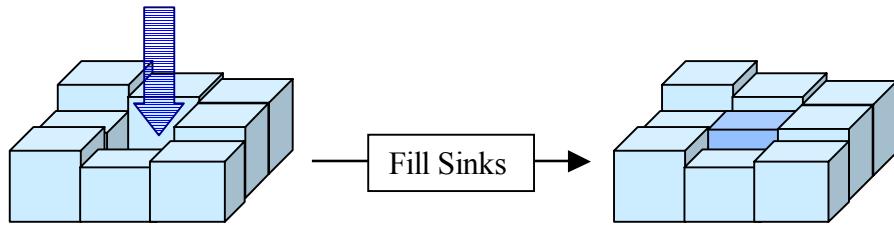
Figure 2.9   Filling of a sink, the central cell being lower than surrounding cells is raised

The removal of sinks to produce a filled DEM is achieved automatically by using either the drop down menu within the ArcHydro extension or within the Hydrological Analysis add-in. Both of these have the function labelled as "Fill Sinks". The concept of filling sinks was introduced by Jenson and Domingue, (1988).

Flow Direction

After the production of the filled DEM grid comes the main processing part of the hydrological processing procedure (Figure 2.7). The production of the Flow Direction grid is the basis of GIS hydrological modelling. Flow direction is based around the Eight-Direction Pour Point Model (O'Callaghan and Mark 1984). A pour point is a location through which water flows out of a cell or region (ESRI 2000). Each cell in the matrix is surrounded by eight neighbouring cells, of which 4 are directly opposed, and 4 diagonally opposed (Figure 2.10). Water in this cell can flow to one and only one of its neighbouring cells, in the direction of the steepest descent. These flow directions are coded with integers as in Figure 2.10 (Maidment 1999). The Eight Direction Pour Point Model is obviously a simplification of true surface water flow, which can be in any direction, not just one of the eight assumed directions. Because the cell matrices employed are very dense compared to the area they cover, a succession of cell-to-cell paths can approximate any flow direction adequately when viewed from a distance, therefore the Eight Direction Pour Point Model is reasonably accurate (Maidment 1999).
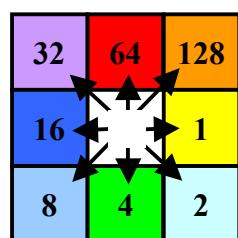
Figure 2.10   Flow Direction – The Eight-Direction Pour Point Model

The Flow Direction Grid is produced either by using the drop down menu within the ArcHydro extension or within the Hydrological Analysis add in or directly within the Raster Calculator using the `Flow Direction` function (Equation 2.1).

(Equation 2.1): `flowdirection([HydroDEM])`

For information about Raster Calculator and Map Algebra see Appendix 1.

Flow Accumulation

The Flow Direction grid is used as the input grid for the `Flow Accumulation` function. Flow Accumulation creates a grid of accumulated flow to each cell, by counting the number of cells upstream of each cell. This function can be enabled either by using the drop down menu within the ArcHydro extension or within the Hydrological Analysis add in or directly within the Raster Calculator using the `Flow Accumulation` function.

(Equation 2.2): `flowaccumulation([FlowDirection])`

Stream Definition

By manipulating the flow accumulation grid a stream network can be derived (Figure 2.11). This produces a linear network of cells representing the river. Equation 2.3 is used within the Raster Calculator:

(Equation 2.3): `setnull((lessthan(3,[FlowAccumulation]), 1)`
        Where `3` = value below which cells will be given the value NoData.
        And `1` = value that the remaining cells will be given.

In the example used, all cells that have more than 3 cells flowing into them are assigned the value 1; all other cells are assigned the value NoData. The resulting stream network can be converted to a line shapefile or be kept as a grid for use in further processing. In this study the threshold of a flow accumulation of 100 was chosen, in order to provide a river network of the greatest detail, so that headwater tributaries are defined. However, this does result in the definition of streams in valleys where the stream might be dry, such as chalk streams. This should not be a problem unless the network is being used to determine an attribute such as distance to source.
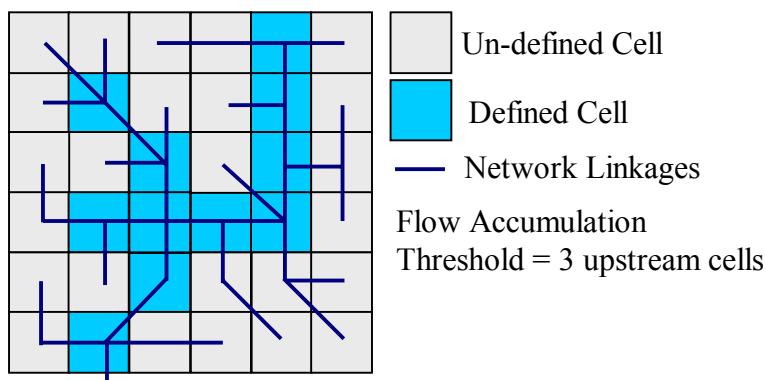

Figure 2.11   Flow Accumulation and Stream Definition

2.2.5 Problems with Basic Hydrological Analysis

The basic/standard hydrological analysis described in Section 2.2.4 results in a derived river network. This river network can be displayed with the "real" river network and some comparisons made. With the basic hydrological analysis a number of common errors occur, these may be significant and render the derivation of catchment features inaccurate.

These problems are at their most extreme in, but not restricted to, low relief or low lying areas. Derived streams can lie some distance from that of the true network or may even link to the wrong river valley (Figure 2.12). Garbrecht and Starks (1995) describes another problem in low relief where numerous parallel streams can be artificially created. Garbrecht and Martz (2000) suggest that these problems will make derived networks unsuitable for calculation of drainage features.

■ Derived river network (basic method)
— River network
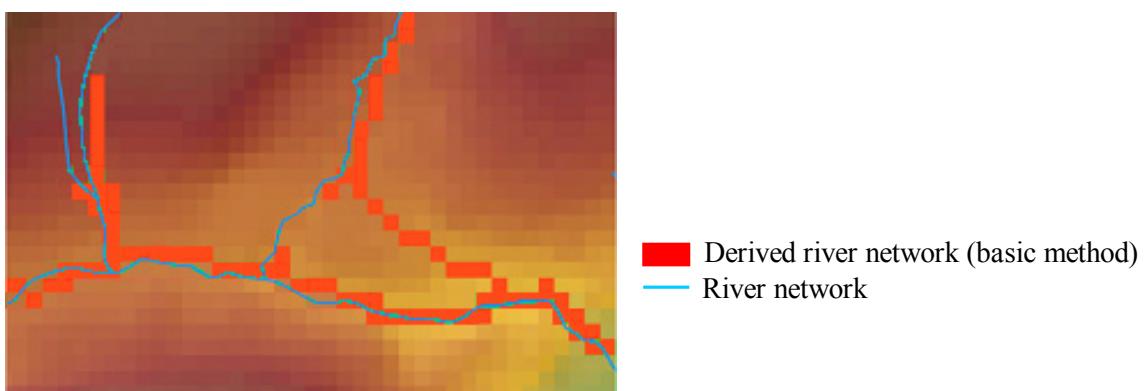
Figure 2.12   Erroneous derivation of river network

Improvements to Basic Hydrological Analysis

The problems that are associated with the standard hydrological analysis have resulted in considerable work in the GIS hydrology field to improve the GIS's quality of derived data from the DEM. Much of the work that has been undertaken has centred on altering or improving the DEM prior to the hydrological analysis, this is termed terrain pre-processing.

The first method of terrain pre-processing was the `fill sinks` algorithm developed by Jenson and Domingue (1988). Since this early development further steps have been made. Hutchinson (1988) described a method of incorporating a digitised river network into the DEM. Maidment (1999) and Saunders (2000) describe a method, where the vectored river network is rasterised and then subtracted (or burned into) the DEM resulting in a trench being created where the river network lies. This method had some definite improvements over the standard hydrological analysis with improved accuracy of stream delineation (Saunders, 2000). However, a number of limitations remain. These include distortions in watershed boundaries (Saunders, 2000) and the introduction of parallel streams (Hellweger, 1997).

Further developments have tried to overcome these problems. One method that provides a processing efficient algorithm is the `AGREE` function developed by Hellweger (1997). This method has been shown to be effective and has been integrated into ESRI's ArcHydro toolkit (Maidment, 2001). This function modifies the DEM by imposing a vectored river network onto them (termed burning/fencing).

The AGREE function is a development on the idea described in Jenson and Domingue (1988), Hutchinson (1988) and Saunders (2000) of integrating a vector river network (as digitised from maps or aerial photos) into the DEM. The AGREE process is more advanced, and will eliminate the parallel stream problem created by the simple burning in process (Hellweger, 1997).

The user needs to enter three reconditioning parameters:

- `Vector buffer`. This is the number of cells around the river network for which the smoothing will occur. A Value of 5 is used within the investigation.
- `Smooth drop/raise`. This is the amount (in vertical units) that the river will be dropped (if the number is positive). This value will be used to interpolate DEM into the buffered area between the boundary of the buffer and the dropped/raised vector feature (ESRI 2002). A Value of 10m is used within the investigation.
- `Sharp drop/raise`. This is the additional amount (in vertical units) that the river will be dropped (if the number is positive). This has the effect of additional burning/fencing on top of the smooth buffer interpolation. It needs to be performed to ensure preserving the linear features used for burning/fencing (ESRI 2002). A Value of 10m used within the investigation.

The result is to alter the DEM so that the derived river channel fits the true position of the river. In Figure 2.13 an example of a single cross section (A to B) with a buffer of 3 cells, a smooth drop of 20m and a sharp drop of 10m is used. Without the AGREE process the derived river network would have been positioned in the wrong place. However, by buffering around the true river position and creating a reshaped river valley (by using a smooth drop of 20m and then a sharp drop of a further 10m) the true river position will be maintained on deriving a river network from the reconditioned DEM.
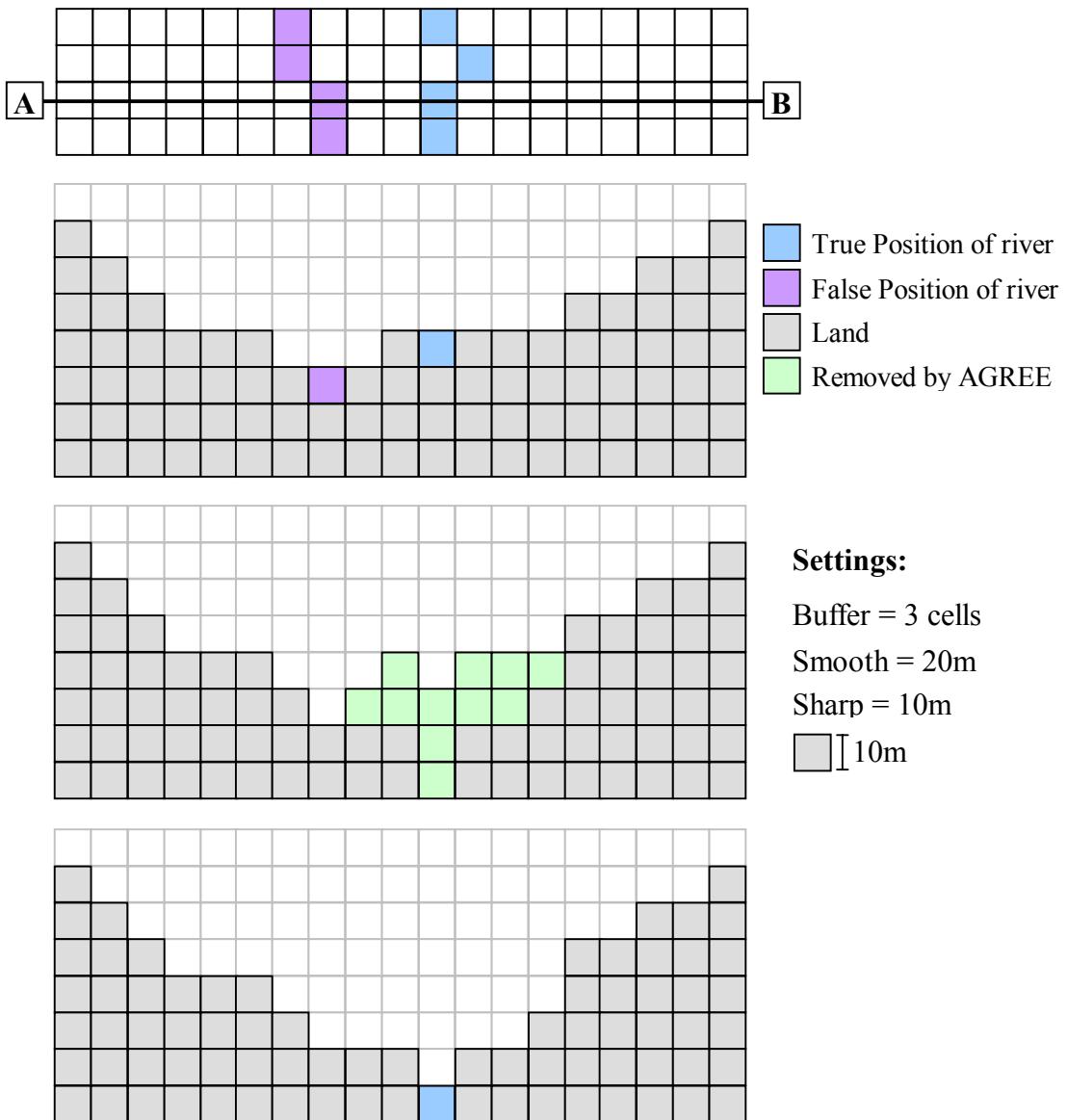
Figure 2.13   Terrain pre-processing, the AGREE methodology

In conclusion, Figure 2.14 highlights the sort of improvements that can be achieved by using the AGREE methodology to recondition the DEM over the standard hydrological analysis, errors such as rivers taking a route into a different sub-catchment can be eliminated or reduced. Saunders (2000) highlights a number of methods and compares their accuracy against a known catchment. Saunders (2000) suggests that the AGREE algorithm was the most processing-efficient of the methods tested that provided satisfactory results. The AGREE method provided only a very limited number of errors determined by comparing delineated catchment boundaries. Although simple "burning in" is a relatively quick process it produced more errors than the AGREE method when compared by Saunders (2002).

The method of applying the AGREE algorithm within terrain pre-processing is described in detail in the next section below.
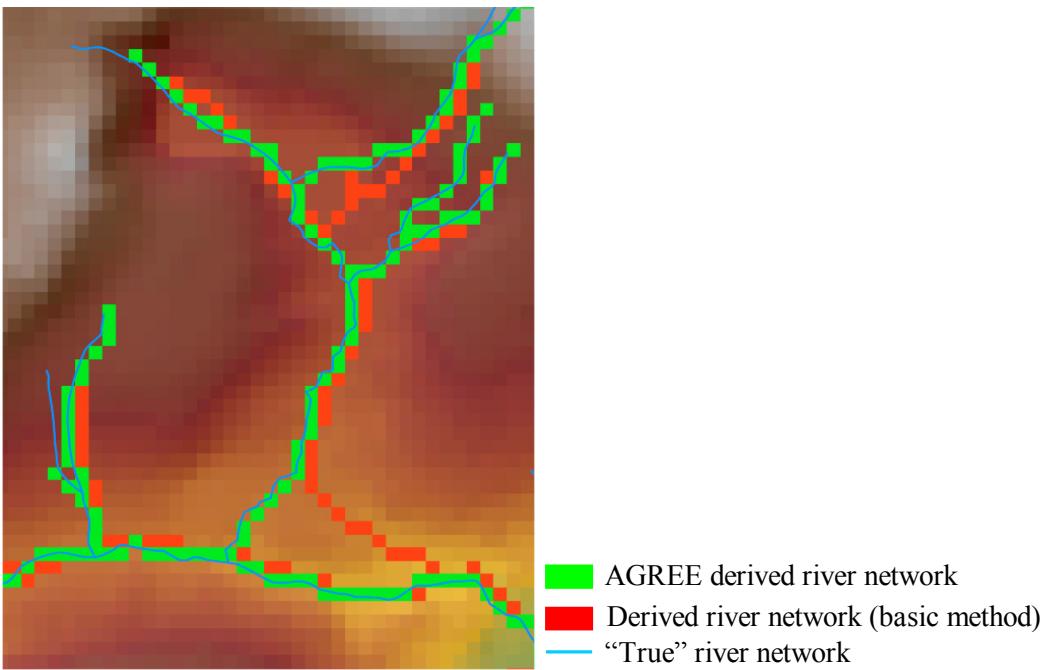
Figure 2.14   Example of improvements to be made with AGREE methodology

Terrain Pre-Processing

Terrain pre-processing is required to alter the DEM making it suitable for hydrological GIS analysis. The process followed (Figure 2.15) is as described in ESRI (2002) as the AGREE methodology followed by identification and filling of any sinks. Following the terrain pre-processing the hydrological analysis is performed within the GIS as the standard method described in Section 2.2.4 above.
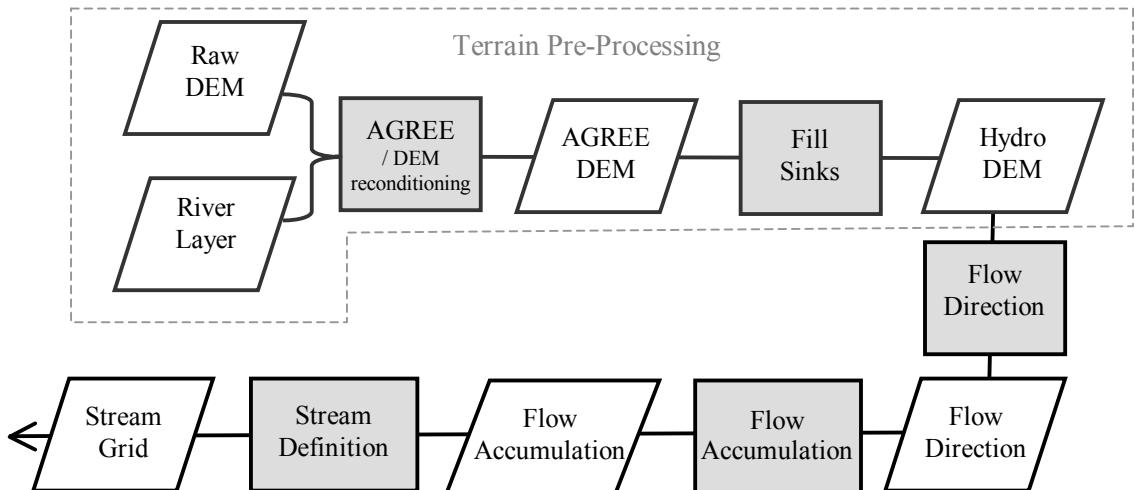
Figure 2.15   Hydrological analysis with terrain pre-processing

The DEM must be in ESRI GRID format. A successful pre-processing is an indication that the underlying DEM does not contain major problems that will prevent further analyses (ESRI, 2002). The AGREE function is implemented from the ArcHydro Terrain Pre-processing menu as the `DEM Reconditioning` function.

There are a number of issues to bear in mind with using the AGREE methodology. It is important that the vector river network used is at a suitable scale for the raster grid being used, an example of the problems that can arise from using inappropriate scales is described in Figure 2.16 that is taken from Saunders (2000). Using the inappropriate scale of rasterisation results in inaccuracies and loss of data. In the investigation the 1:50 000 river network was considered suitable for use with the 50m DEM as the resolutions correspond.

a) Two conceptual grid Layers with stream network superimposed.

b) Rasterisation of the stream network.

c) A gridded and "thinned" stream network.
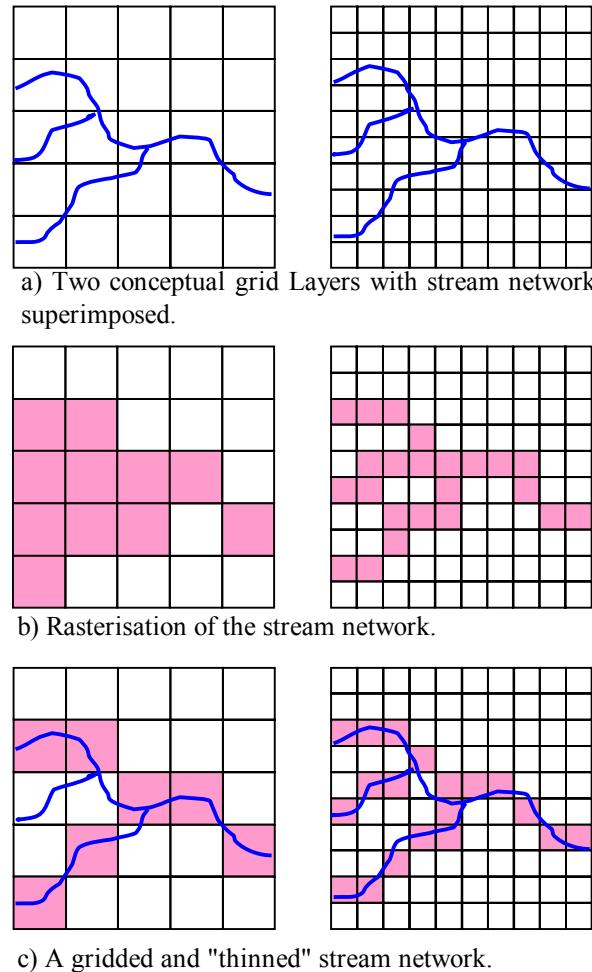
Figure 2.16   Rasterising vector streams into grids of different scales (Saunders, 2000)

It is important to remove artificial drainage, which does not correspond to the DEM from the vector river network. Also the river network used for the DEM reconditioning should not have braided or bifurcated channels but a single main channel. This will prevent the production of numerous mini tributaries in order not to hinder the calculation of attributes.

2.2.6   Calculating Catchment Attributes

After the process of hydrological analysis is performed and the resulting grids produced, catchment based attributes can be calculated. One of the main issues relating to data extraction for sites, is that attributes must be calculated in a way that will enable any model to be applied over the whole model area. To achieve this attributes were calculated for each 50m^2 cell on the derived river network.

Average Altitude Upstream

To calculate the average altitude upstream for each cell on the river network grid the formula in Equation 2.4 was used

(Equation 2.4):
```
(flowaccumulation([FlowDirection],[Altitude]))/[FlowAccumulation]*[Str
eam]
```

This calculates the average altitude upstream of each cell on the derived river network, by using the flow accumulation function to calculate the area upstream of each cell on the network. For an individual cell the value will be the sum of the altitude of cells that are the upstream catchment divided by the number of cells (Figure 2.17).
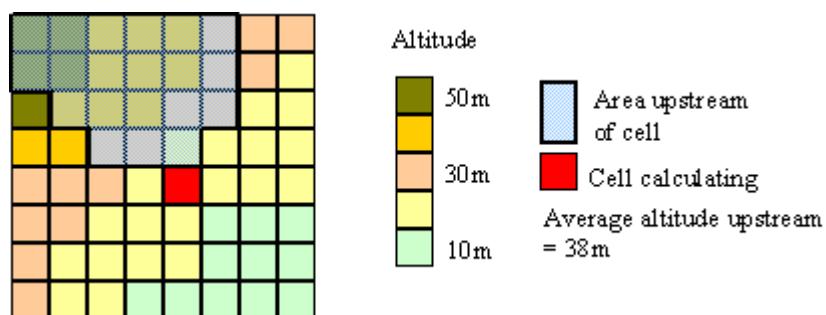
Figure 2.17   Average altitude upstream for a cell on river network

Proportion of a Landcover Type Upstream

To calculate the proportion of a particular landcover type upstream for each cell on the river network grid the formula in Equation 2.5 was used.

(Equation 2.5):
```
(flowaccumulation([FlowDirection],[ArableLand]))/[FlowAccumulation]*[S
tream]
```

Landcover was taken from the Landcover 2000 data set and a calculation performed within Raster Calculator to produce a grid where the particular landcover of interest occurred as a value of one and any other value occurred as NoData. Resulting from the equation would be a river network coloured up by the proportion of the given landcover type upstream of each cell (Figure 2.18).

Landcover types can be grouped together, for example suburban/rural development and continuous urban were combined to produce an urban grid for the calculation.
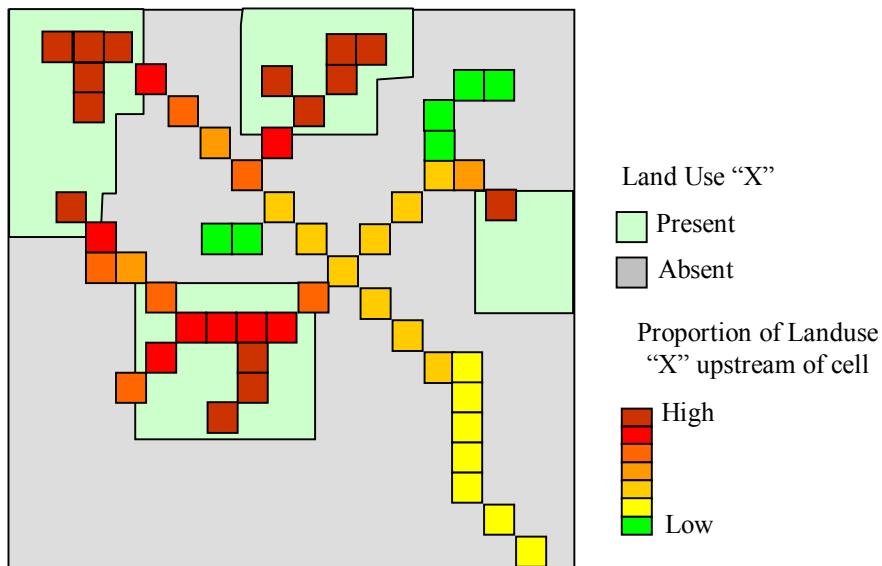
Figure 2.18   Proportion of Landcover type upstream of each cell on river network

Distance to Source

The distance to source is calculated using the `FlowLength` command. This calculates the upstream or downstream distance along a flow path for each cell. Using the upstream function the length of the longest flow path upstream can be calculated.   The function calculates the distance to the most upstream cell, which may not be on the river network. This value may lie some distance further up the catchment from the source of the river and the differences would be at their most extreme in areas with solid geology of a highly porous nature such as chalk. Altering the flow accumulation threshold in order to achieve a delineated river network of the correct detail can negate this effect. The function is described in Figure 2.19. The Map Algebra expression flow length is used to automatically calculate this attribute and is depicted in Equation 2.6.

(Equation 2.6):
```
flowlength([FlowDirection], upstream)
```
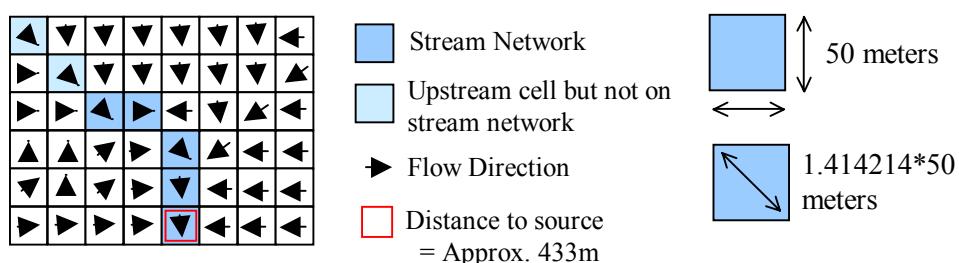

Figure 2.19   Measuring distance to source

Distance to Mouth

The distance to mouth is calculated using the `FlowLength` command. This calculates the upstream or downstream distance along a flow path for each cell. Using the downstream function the length of the longest flow path downstream can be calculated. Equation 2.7 is used to perform this calculation.

(Equation 2.7): `flowlength([FlowDirection], downstream)`

A problem that will commonly occur is that the values will be over estimated. This is because the derivation of the stream network produces sections of straight lines over the areas of zero elevation (Figure 2.20). This can be corrected with the use of a mask over the area of or less than zero meters (Figure 2.21). If the stream network is masked so that these areas are removed and Equation 2.7 re applied the values will be distances to mouth, where the mouth is standardised at a elevation of zero meters above OS Datum.

As new data sets become available it would be possible to recalculate the downstream distances to the tidal limit by using the tidal limit to define the mask described above.

Elevation
High

Low

—— Derived Stream network
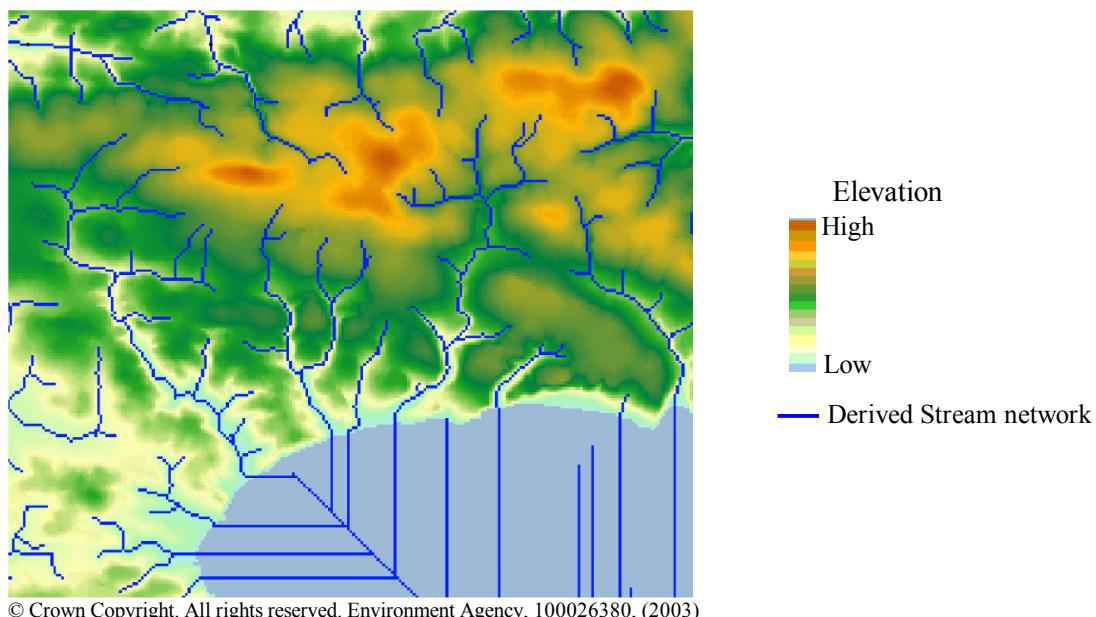
Figure 2.20   Depiction of stream definition once the sea is reached
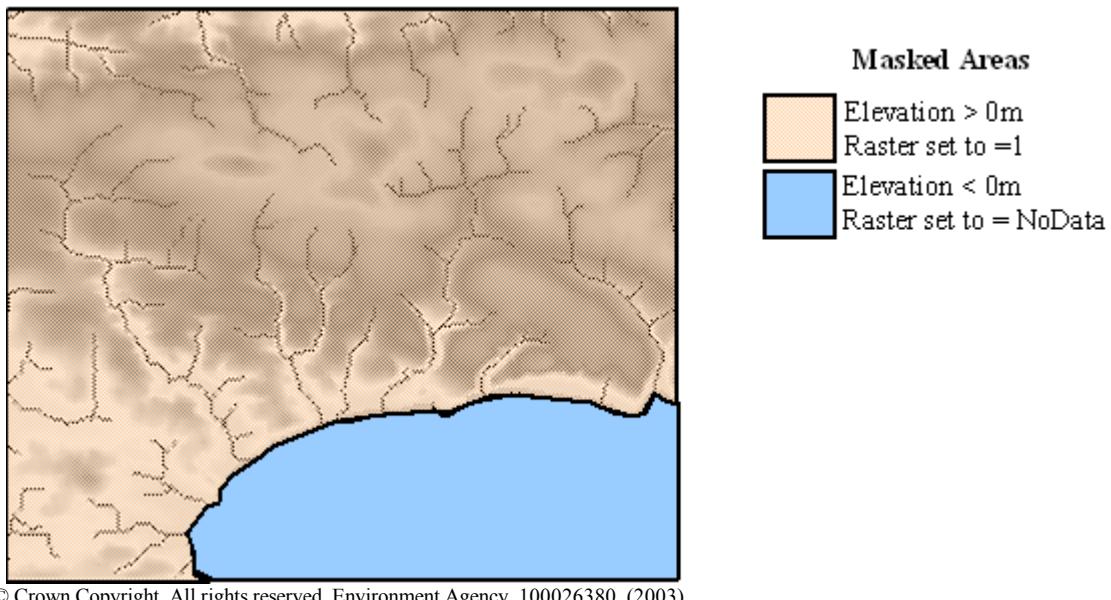
Figure 2.21   Masking of areas to enable calculation of Distance to Mouth

Stream Order

There are two methods of stream ordering used within fisheries; the Strahler method and the Shreve method. In the Strahler method all sections with no tributaries are assigned a stream order of 1 (first-order). When two first-order streams intersect, the downstream section is assigned an order of 2 (second-order). When two second-order sections intersect, the downstream section is assigned an order of 3, and so on. Only when two streams of the same order intersect will the order increase Figure 2.22. For example the Strahler stream order have been used as an analogy for stream size (Argent *et al*. 2003).

In the Shreve method all sections with no tributaries are assigned a stream order of 1 (first-order). When two sections intersect, the downstream section is assigned an order that is the sum of the two intersecting sections Figure 2.22. The Shreve method is commonly referred to as link number and is used in HABSCORE.

The `stream order` function assigns a numeric order to segments of the derived stream network grid representing branches of a linear network. Equation 2.8 describes the Strahler method.

(Equation 2.8): `streamorder([Stream], [FlowDirection], strahler)`
      where `[Stream]` = The raster linear network should be represented as values equal to one on a background of NoData.

Equation 2.9 describes the Shreve method.

(Equation 2.9): `streamorder([Stream], [FlowDirection], shreve)`

Stream order is highly dependent on the threshold value chosen to define the river network, as described in Section 2.2.4.
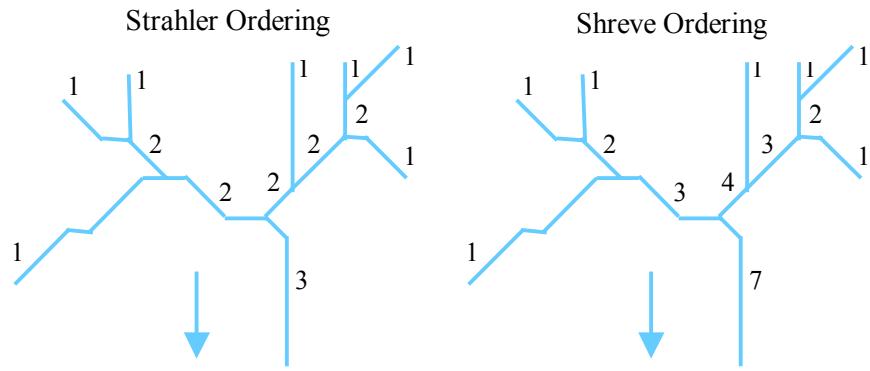
Figure 2.22   Stream ordering methods

Slope

The `slope` function identifies the maximum rate of change in value from each cell to its neighbours (Burrough 1986). An output slope grid can be calculated as percent slope or degrees of slope (Figure 2.24). Within this project slope was calculated in terms of degrees. Using the slope function can do this within ArcGIS in two main ways:

1. Using the built in function from the Spatial Analyst – Surface Analyst menu (Figure 2.23).
2. Using the `Slope` function within Raster Calculator

   (Equation 2.10):
   `slope([DEMGrid], degree)`

Both of these methods will produce the same result.
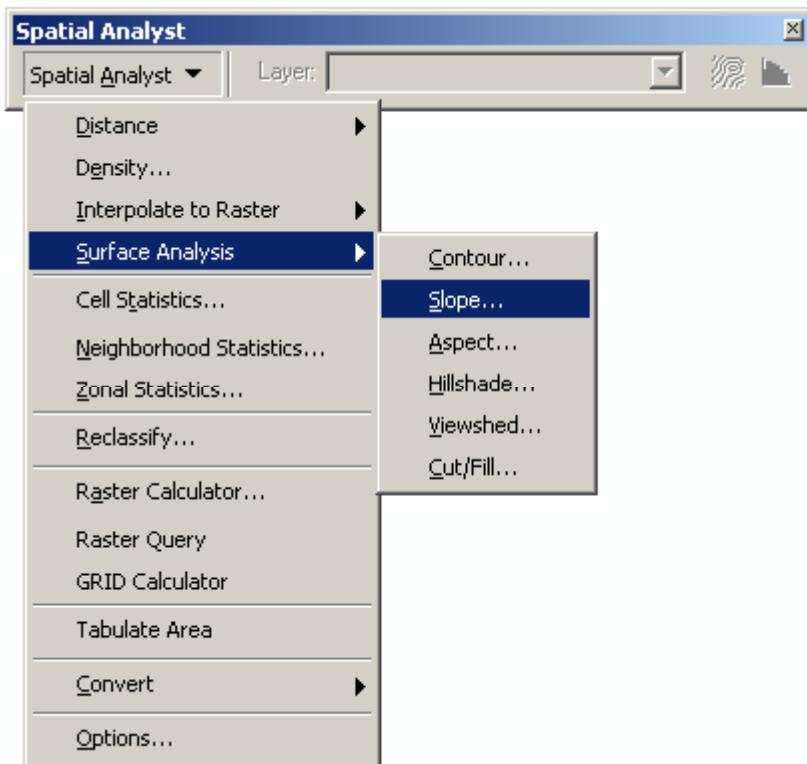
Figure 2.23   Slope Function within Spatial Analyst Toolbar
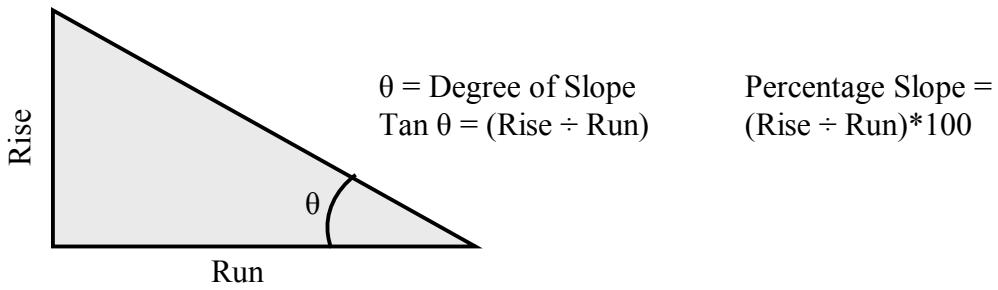
Figure 2.24   Calculating slope

As the slope function identifies the maximum rate of change in value from each cell to its neighbours, the slope calculated may be that of the river valley sides and not of the river itself. To overcome this a modification to the standard slope function can be applied Equation 2.11.

(Equation 2.11): `slope(([DEMGrid]*[Stream100]), degree)`

This cuts the raw elevation data to the river network before performing the slope calculation on the resulting grid. This should force the calculation to take the slope of the river channel rather than the valley sides.

The slope calculations should be performed on the original DEM grid and not the one that has undergone terrain pre-processing as this modifies the DEM in such a way that local slope values may become inaccurate. The terrain pre-processed DEM provides the ability to delineate an accurate river network not accurate height and slope calculations.

Average Slope Upstream

Catchment gradient or average slope upstream was calculated using the same equation as used to calculate average altitude upstream with the altitude grid replaced by the slope grid calculated in Equation 2.11. The calculation is shown in Equation 2.12.

(Equation 2.12):
`(flowaccumulation([FlowDirection],[SlopeGrid]))/[FlowAccumulation]*[Stream]`

2.2.7   Location Auditing

Before data can be extracted from the raster grids produced by the above process for fish survey sites, some auditing of the location details of these sites must be performed. There are a number of errors that are commonly seen, especially in older data sets. These common errors are listed in Table 2.2. Identification of these errors was performed with the help of some automation within the GIS but predominately by manually cross-referencing the sites spatial location against the attribute tables stated location. An example of this would be where a number of sites when plotted in the GIS could be found in the sea. This is obviously an error and was often due to the NGR letters from the opposing 100km grid being given to a set of sites erroneously. This error was easily fixed by altering the NGR 100km grid letters.

Table 2.2　　Common location errors

Error	Description	Methods of Identifying
Wrong Grid Letters	National Grid Reference (NGR) has two leading letters that identify the 100km grid. Errors can arise where two 100km grids occur in the same catchment, and sites are mistakenly given the wrong letters.	1. Sites may appear in the sea. 2. Error code of 99999 or 0 may occur in conversion to GIS. 3. Site may appear in wrong catchment.
Missing Grid Letters	National Grid Reference (NGR) has two leading letters that identify the 100km grid if missing the location will not be identified.	1. Sites may appear in the sea. 2. Error code of 99999 or 0 may occur in conversion to GIS. 3. Site may appear in wrong catchment.
X and Y co-ordinates inverted	The person entering the data can confuse the Eastings and Northings.	1. Site may appear in wrong catchment.
Missing X or Y co-ordinates	The person entering the data can miss data from the Eastings and/or Northings. Often occurs when a leading zero is required.	1. Site may appear in wrong catchment.
Missing NGR	NGR totally missing	1. Site not locatable

2.2.8　Acquiring Site Attributes

To enable the extraction of attributes for the fish survey sites (Figure 2.25), the sites must be moved slightly so that they fall upon the derived river network. This is required because six figure grid references do not give the accuracy of location to enable sites to always fall on the river network. Also the process of deriving a river network may result in the river lying a small distance from the river position used to calculate the sites original position. To overcome this a process of snapping the sites to the derived river network is undertaken.
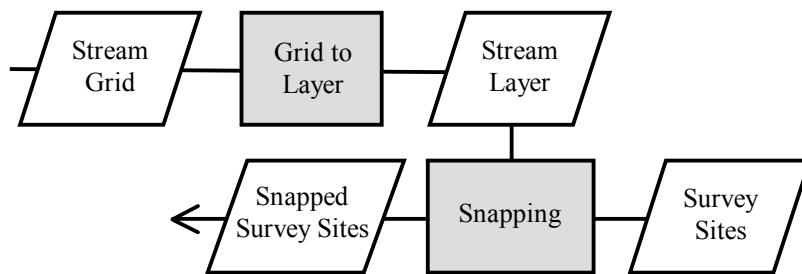
Figure 2.25　Site preparation

Convert Derived River Network to Vector Line

The first stage in snapping the sites to the derived river network is to produce a vector version of the raster grid network.

This can be undertaken by using the `Stream Shape` function that is built into the `Drainage Line Processing` function of ArcHydro tools (Figure 2.3). Another method of reaching the same output is to use the `Stream Shape` function directly within Raster Calculator (Equation 2.13). Stream Shape is optimised to use a `flow direction` grid to aid in vectorising intersecting and adjacent cells (ESRI, 2000).

(Equation 2.13) `streamshape([StreamGrid], [FlowDirection], noweed)`

Where `[StreamGrid]` = the Raster Grid River Network and `noweed` = term that determines no generalisation.

Snapping Sites to Derived River Network

Snapping is a process where site locations as a point shapefile (ESRI point layer) are moved so that they lie directly on the derived river network. Sites are snapped to the closest river if they lie within a defined snapping tolerance (Figure 2.26).
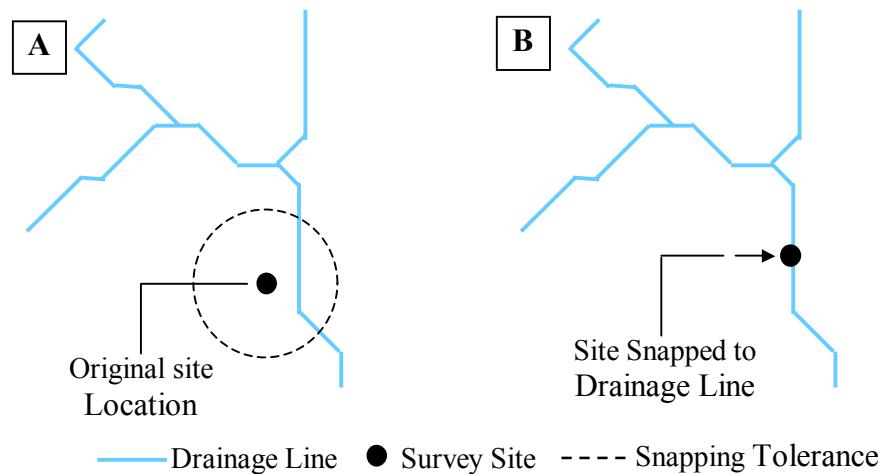
Figure 2.26   Snapping of Survey Site to Derived River Network

Snapping is undertaken by starting an editing session within ArcView's ArcMap. This is achieved by clicking on the Editor Toolbar button. Within the Editor Toolbar snapping environment (where sites will be snapped to) and snapping tolerance can be set (Figure 2.27). In this study the snapping tolerance was set to 200m.

A = Editing Options
B = Snapping Environment
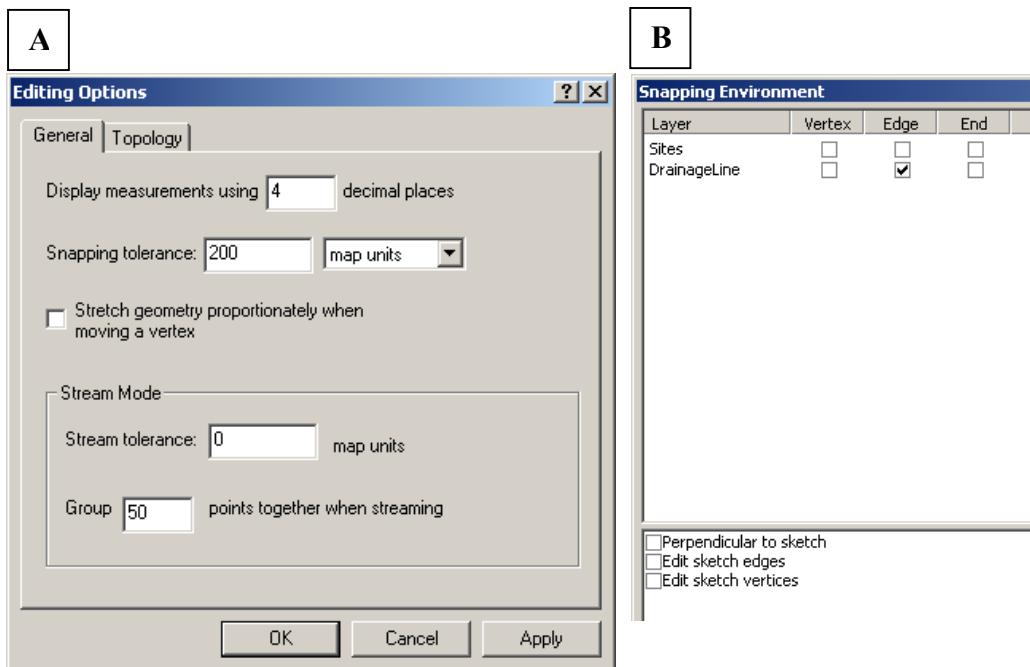
Figure 2.27   Snapping Options within Editing Session

Appendix 2 (Part 1) sets out the process in more detail and describes how to implement a macro that will allow the process of snapping sites to be performed in a batch process of selected sites.

Snapping Errors

There are errors that may occur with the snapping process that should be identified and corrected. These errors result from the sites being snapped to the wrong section of river. Because the original location of the site may not be that accurate the nearest river may be a small tributary while the main river may have been where the site should be located (Figure 2.28). If the snapping process is undertaken in small batches the ability to pickup and correct these errors is simplified.
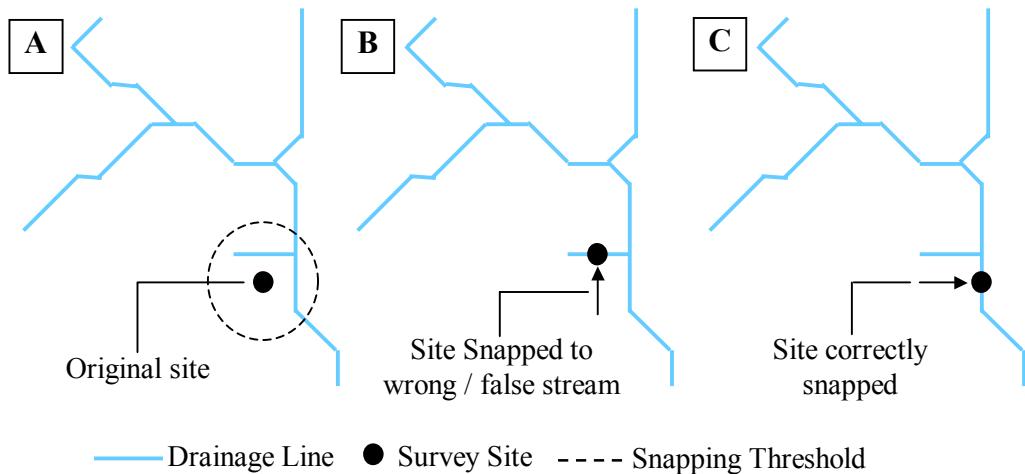
Figure 2.28   Snapping errors

Adding Corrected X and Y Co-ordinates to Attribute Table

Once successful snapping has been achieved the new X and Y co-ordinates can be recorded onto the shapefile's attribute table. Appendix 2 (Part 2) sets out the process in detail and describes how to use a tool within ArcCatalog that will allow X and Y co-ordinates for all snapped sites to be added to the shapefile's attribute table.

Acquiring Attribute Values from Raster Grids

In order to acquire values of the attributes for each site being investigated a method must be applied to extract these values from each of the derived raster grids. Appendix 2 (Part 3) sets out in detail how to add values from a raster grid to a new column in the attribute table for all sites. It also describes how to implement the macro that will allow this to be performed.

2.3   Results

A number of data sets as raster grids were produced by the above process the values from these were extracted for the sites shown in Figure 2.29 and stored in a MS Access database to enable linking with fish data for the same sites. Table 2.3 shows the format of the resulting table with descriptions of all the attributes collected. Data was extracted for some 1989 sites in England and Wales as shown in Figure 2.29.

Table 2.3   GIS Site Data MS Access table

Field Name	Data Type	Description
New_NGR	Text	Corrected National Grid Reference
Old_NGR	Text	Pre-Corrected (original) National Grid Reference
Eastings	Long Integer	X – Coordinate (pre-snapping)
Northings	Long Integer	Y – Coordinate (pre-snapping)
X_sn	Double	X – Coordinate snapped to derived river network
Y_sn	Double	Y – Coordinate snapped to derived river network
SiteAlt	Integer	Altitude of Site from 50m Raw DEM
AvAltUpStr	Double	Average Alt upstream of site from 50mRaw DEM
P_UrbUpStr	Double	Proportion upstream of site that is Urban
P_GrasUpSt	Double	Proportion upstream of site that is Grassland
P_DecUpStr	Double	Proportion u/s of site that is Deciduous Woodland
P_ConiUpSt	Double	Proportion u/s of site that is Coniferous Woodland
P_ArablUpS	Double	Proportion upstream of site that is Arable land
AvSPRUpStr	Double	Average SPR value Upstream of site
SiteFlowAc	Long Integer	Flow Accumulation value of site
SiteLndUse	Integer	Landcover at site from Landcover 2000
SiteAspect	Double	Aspect of site calculated from 50m Raw DEM
SiteRain	Integer	Average Annual Rainfall at site Isohyets (50mm)
AvSlopeUpS	Double	Average Slope of catchment Upstream of site
AvRivSlope	Double	Average Slope of river network Upstream of site
P_LakesUpS	Double	Proportion upstream of site that is lakes
Strahler	Long Integer	Stream Order at site – Strahler method 0=NoData
Shreve	Long Integer	Stream Order at site – Shreve method
P_up_asp_s	Double	Proportion upstream of site that has a Southerly Aspect (175-185°)
P_up_asp_n	Double	Proportion upstream of site that has a Northerly Aspect (355-005°)
DistSource	Double	Flow Length from virtual source (m)
DistMouth	Double	Flow Length to mouth (m)– defined by 0m elevation
Position	Double	Position =(DistSource) / (DistSource+DistMouth)
CatchArea	Double	Catchment Area = $(FlowAccumulation*50^2)/1000^2$
SiteSlope	Double	Slope at site calculated from 50 m DEM
D_slope	Double	Slope at site calculated from 50 m raw DEM cut to river network
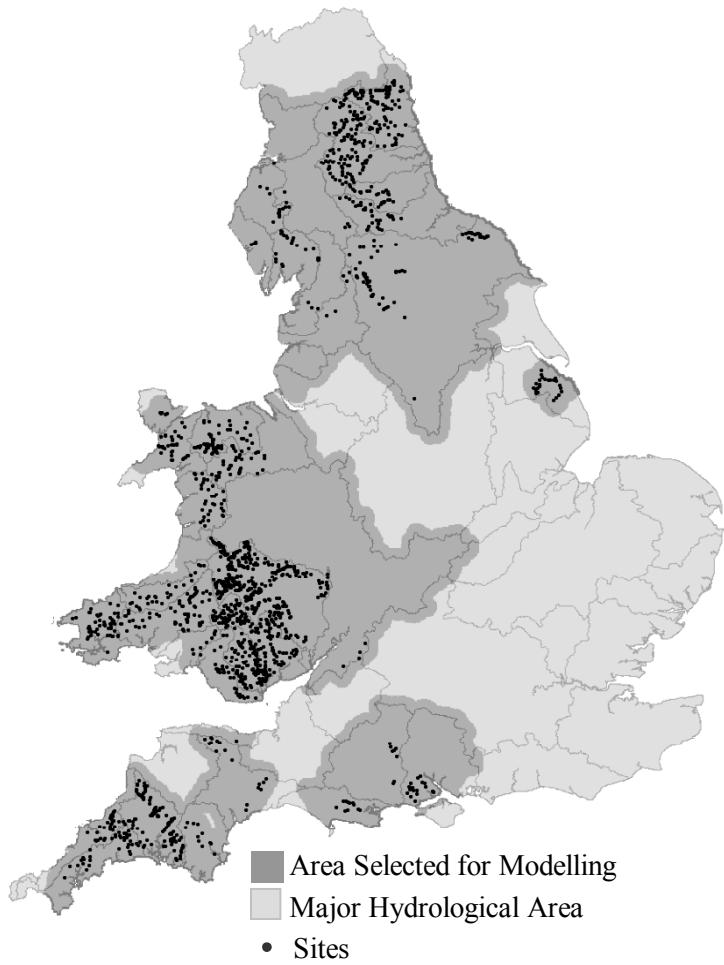
Figure 2.29   Sites for which attributes were extracted

2.4   Discussion

The methods of extracting attributes for sites and river networks from the GIS were generally successful. The methods used for calculating average or proportions upstream of a cell on the river network can be tailored for numerous other datasets as required. The main problem identified with the methods is that of the quality of raw data used. The methods themselves appear to be robust and successful, however, quality issues with some of the datasets reduces the confidence in the data extracted.

2.4.1   Data Issues

There are a number of issues that have been identified relating to the quality of some of the raw data sets. These quality issues may be inherent to the data or only result from its use in a certain way. Each of the data sets in question will be dealt with in turn.

River Layer

The vectored river layer dataset was originally produced by CEH in 1998 by a process of manually digitising from the raster Ordnance Survey 1:50 000 maps. Being manually derived the data is subject to human errors and early versions of the data had errors that

included missing areas (Figure 2.30). A proportion of these errors may have been corrected in recent updates, but the dataset used may still contain some errors.
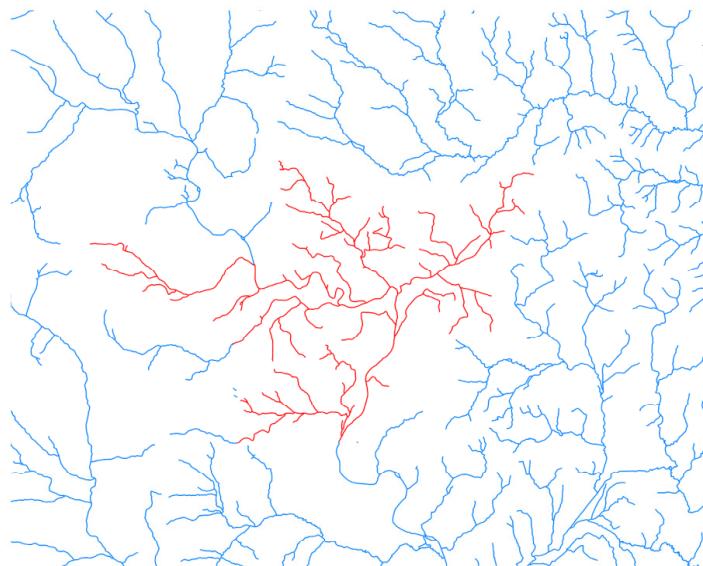
River Usk example, missing areas in original dataset highlighted in red.
© Crown Copyright. All rights reserved. Environment Agency, 100026380, (2003),
© Centre for Ecology and Hydrology 1999

Figure 2.30   Issues with Rivers at 1:50 000 dataset

Another major problem with the dataset is the inability to automatically select out the main river channel and remove features such as ditches or braided channels (Figure 2.31). The problem is that no suitable attributes exist to allow for automatic selection of specific types of river. The result is that a fair amount of time must be spent manually cleaning up the network prior to using it as an input into the AGREE terrain pre-processing. This is required because the AGREE terrain pre-processing does not cope well with hanging sections of river, loops and excessive areas of artificial ditches that do not augment the drainage network.

A = Braided Channel, bifurcated or artificial channel.

B = Artificial drainage ditches.

© Centre for Ecology and Hydrology 1999
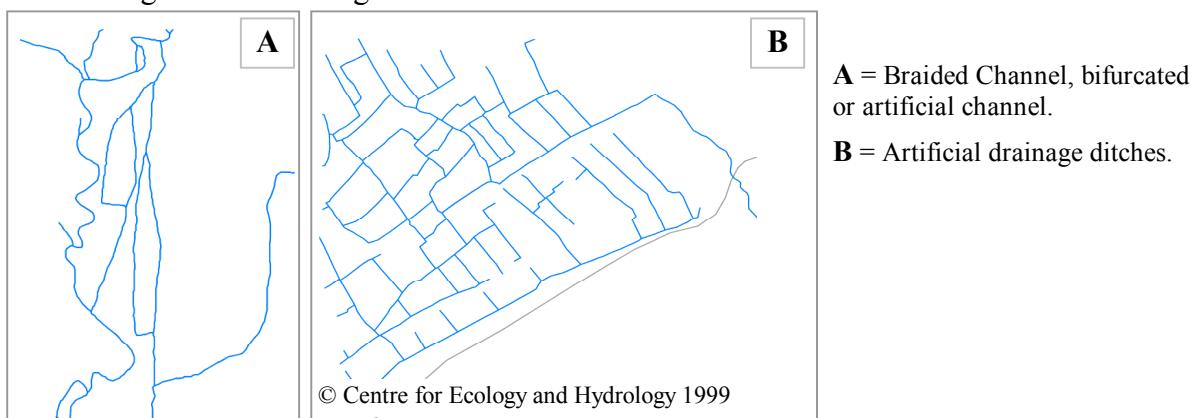
Figure 2.31   Network connectivity issues with Rivers at 1:50,000 dataset

There may be other sources of these data and there are projects underway within the Ordnance Survey to produce an accurate river layer, with useful attributes. This may provide a more reliable dataset to use as an input to the hydrological analysis process.

Land Cover

The Landcover Map 2000 (LCM2000) was produced from satellite data and calibrated using field based measurements. The LCM2000 was derived from a computer classification of satellite images, obtained mainly from Landsat satellites. Field survey data was used to provide calibration; some 569 sample squares were visited in 1998-1999, across the full range UK environments (Haines-Young *et al* 2000 and Fuller *et al* 2002).

Although there have been considerable improvements since the 1990 Landcover Map, some issues with the quality of the data have been identified. The top-level classification (the most general) appears to be reasonable but the more detailed levels of classification are prone to errors when looking at the whole country. An example of this is depicted in Figure 2.32 where the general classification of Arable is split into Arable Cereals and Arable Horticultural. An edge effect can be seen when the Arable Cereals is highlighted. This occurs where there is low confidence in the particular classifications and may be the result of a mismatch in the satellite image tiles being taken from different dates.

This problem would suggest that using the more detailed classifications should only be undertaken with caution but the more general classification should be suitable for use. Another major difficulty with landcover data is that it may not reflect the land use in a way that is meaningful to the effect on a fish population. Landcover gives much less direct information on the actual type of activity that might impact a fish population. True land use might give us information like park, golf course, pasture and so on, but landcover would use a general term for all of these as "improved grass". This will result in a reduction of the value of this data in terms of picking up a possible influencing factor.

© Centre for Ecology and Hydrology 1999
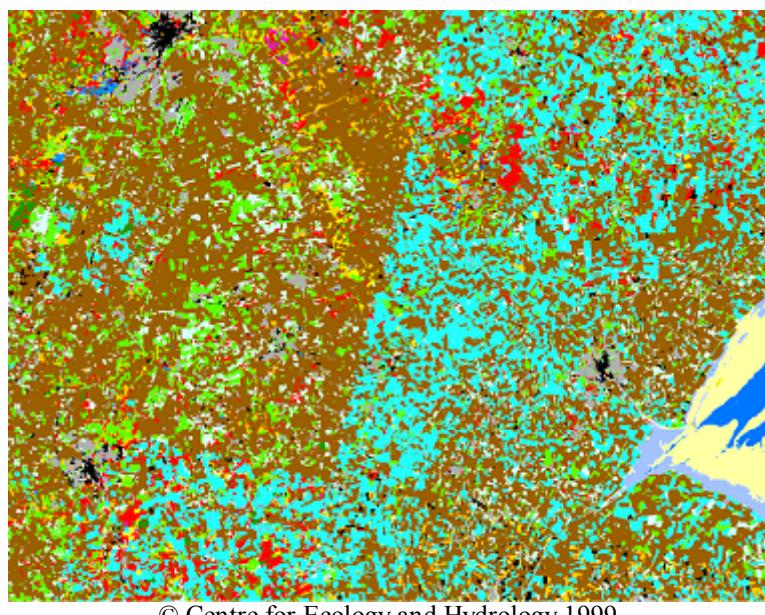
Figure 2.32   Issues with Landcover 2000 (LCM2000) dataset. Arable Cereals class highlighted light blue to demonstrate edge effect

Digital Elevation Model

The elevation data within the DEM used (Ordnance Survey Land-Form PANORAMA) has a vertical accuracy of only ± 5.0m (Ordnance Survey 1997), this would generally not be a problem in high relief areas but improvements to this vertical accuracy would be of help in the problematical low relief areas. The importance of vertical accuracy is shown in low relief areas where this level of inaccuracy can result in the derived river network flowing many kilometres away from the true location or tributaries linking into the wrong main river. The reduced vertical accuracy of the DEM can also result in a lower level of confidence in other derived attributes such as; altitude, average altitude upstream slope and average slope upstream.

There is scope to improve on the current DEM data sets for example; a new data set is due for delivery within the Environment Agency in mid 2003. These dataset from Intermap technologies known as NEXTMap Britain brings a more current, and accurate digital elevation and image data of England, Wales and Scotland the vertical accuracy for these data is ±50cm in some areas and ±1m in the other areas. This improved data set would provide for better accuracy in extracted data and improved performance in GIS based hydrological modelling.

Slope

The ratio between DEM horizontal and vertical resolution can have a significant effect on computed surface parameters (Garbrecht and Martz 2000). This may be a particular problem because; slope is calculated as the difference in elevation between adjacent cells divided by the distance between them. As the DEM is generally in an integer format the computed slope can only take on a limited number of discrete values (Garbrecht and Martz 2000). This may be inadequate to represent slope values in low relief areas but may not be a significant problem in mountainous or high relief areas (Garbrecht and Martz 2000). Improvements may be made with the introduction of the improved resolution DEM's that are likely to become available in the future (See Digital Elevation Model section above). Because slope is calculated between adjacent cells (50m in this case) small scale changes and noise may mask broader changes.

There have been studies that have attempted to calculate upstream river gradient using the elevation at the site and at the source (Broadbridge, 2002). However, the difficulty with this method is that it is both time consuming and not readily applied to every cell on the derived river network as an automated method.

A possible improvement to slope or gradient calculations would be to use the raw DEM values. These would be used for set distances upstream and down-stream of a river section to calculate gradient for that section. This method would calculate river gradient rather than the noisy topographical gradient that results from the slope algorithm. Garbrecht and Martz (2000) compare the results of a number of methods for calculating slope. They show that there can be varying differences between the methods depending on whether the slope is closer to flat or steep. They conclude that the most appropriate method depends entirely on the application.

2.4.2 New Data Sources

There are numerous other sources of data that are available for the UK. However, a number of these are insufficient in their coverage or had not been compiled into full data sets at the time of the project. The methods described above could be applied to new data sources as they become available to provide site and catchment based attributes.

In the short term a number of other datasets will become available within the Environment Agency. Examples of these are listed in Table 2.4.

Table 2.4   Examples of other data sources

Layer	Information	Present Availability	Future Availability
Solid Geology	BGS Solid Geology for the UK 1:250 000 Scale	Patchy coverage available for UK, Environment Agency hold full UK coverage of 1:650 000 scale but licensing issues limit use.	1:250 000 scale data from BGS for UK Solid Geology should be available in the short term.
Drift Geology	BGS Drift Geology for the UK 1:250 000 Scale	Patchy coverage available for UK, Environment Agency hold full UK coverage of 1:650 000 scale but licensing issues limit use.	1:250 000 scale data from BGS for UK Drift Geology should be available in the short term.
Soils	National Soils Map for England and Wales., derived dataset is a 1km grid	Available for England and Wales but there are licensing issues. Available as 100km tiles in vector format. Time consuming to convert to grids at present.	Licensing issues should be resolved. Time required for converting to useful formats for modelling.
Sediment Risk to Water-courses	Prediction of sediment delivery to water-courses from land project (McHugh *et al.* 2002)	Data received by Environment Agency, but requires processing before delivery to the national GIS data drive.	Will be available for internal Environment Agency use by April 2003.
Tidal Limit	Foreshore layer derived from OS BoundaryLine dataset at 1:10 000 scale	Available but very large dataset so requires considerable processing to convert to a useful mask grid.	Time required for converting to useful format for modelling.

2.5   Conclusions

GIS provides the ability to automate and improve measurement of catchment features that will aid interpretation of fish survey data and improve the cost-effectiveness of fisheries management methods. However, good data auditing is required. This is of particular significance in relation to location data, which is prone to errors when taken from older data sources.

GIS provides the ability to calculate attributes that are easy to calculate by hand in an automated way but also offers the ability to calculate variables that would be either very time consuming or impossible to calculate without using a GIS.

Attributes can be calculated in a systematic way and the calculation of new attributes can be achieved by using new or improved datasets, as they become available. The advantage of such a system is threefold:
1. By having raster datasets attributes can be extracted for any 50m cell on the river network.

2. Any relationships discovered during analysis can easily be applied to the whole river network.
3. Any dataset alterations or improvements can easily be applied.

There are a number of issues that must be borne in mind. Data is not infallible and neither is the GIS methodology. There are quality issues with many data sources; these are however, generally quite small and often not quantified. The GIS method of deriving a river network is not perfect, but improvements are possible over time and may be applied as modifications to recalculate the desired attributes. Manual methods would not be expected to be perfect either and comparing the GIS methods to manual methods is discussed in Section 3.

2.5.1 Recommendations

As with all studies of this nature there are improvements that can be made. In the future there will be improved and additional datasets. The method should be reapplied to these so that more up to date data exists for use within fisheries and the wider community. The GIS and more specifically hydrological analysis are continually being developed and improvements are likely as time goes by.

Possible Method Improvements

One improvement that would be ideal for further work would be to improve the derived river network. Although the present method derives a suitable river network because of the flow accumulation threshold being a fixed value for the whole area being processed the network may depart from reality by being overly detailed in some areas and under detailed in others. The method that should to be investigated would be to use a combination of the derived sub-catchments and a threshold value set for each of these sub-catchments determined by the (Standard Percentage Runoff) SPR value for that sub-catchment. Initial investigations have shown a good relationship between the SPR value and the detail of the river network (Figure 2.33). This is to be expected as the SPR value relates to the amount of rainfall that will end up as run-off and hence result in rivers.

☐ Defined River – Note 3
☐ Note 1   ☐ High SPR
☐ Note 2   ☐ Low SPR

Note 1: Defined river where a single `Flow Accumulation Threshold` could result in poor detail in sub-catchment.

Note 2: Defined river where a `Flow Accumulation Threshold` could result in too much detail in sub-catchment.

Note 3: Defined river where SPR has been used to determine different `Flow Accumulation Threshold` values for each sub catchment.

Figure 2.33   Using SPR to vary detail in defined river networks
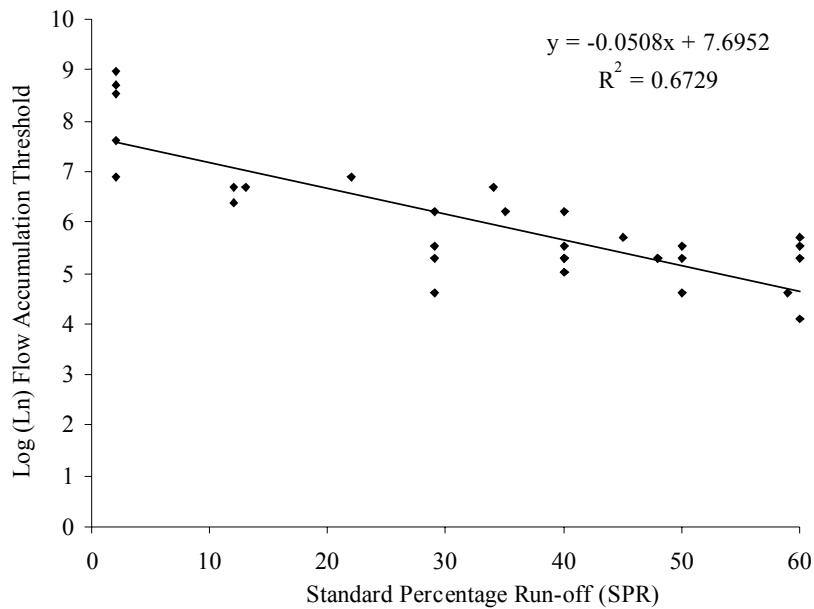
Figure 2.34   Example of relationship between "Flow Accumulation Threshold" that produces correct detail in river network and SPR value in that catchment

Data Delivery

The methods could be applied consistently over the whole of England and Wales once the improvements to the raw datasets are in place. The production of a set of attribute layers that could be available to fisheries scientists via the national GIS data drives should be a valuable tool.

3. COMPARING GIS AND MANUALY DERIVED VARIABLES

3.1 Introduction

The commonly used map based variables in existing Environment Agency methods such as HABSCORE (Barnard, Wyatt and Milner, 1995), the Fisheries Classification Scheme (Mainstone *et al*., 1994), PHABSIM (Johnson *et al*., 1993), River Habitat Survey (Environment Agency 1996) and RIVPACS (Wright *et al*., 1994) include altitude, distance to source, distance to mouth, catchment area and gradient.

The majority of these attributes are relatively easy to calculate within a GIS. However, to make comparisons with previous work and to enable such methods to be used directly with these methods it is useful to look at the relationships between the GIS derived and the manually derived variables.

This section investigates and describes the relationship between the GIS and the manually derived variables.

3.2 Methods

There are a number of comparable GIS and manual variables; site altitude, distance to source, distance to mouth, catchment area, site gradient, and catchment gradient. The manually derived values used for the comparison come from 976 sites that have map based variables from HABSCORE surveys. The HABSCORE data set was chosen because it consisted of a large number of sites and the data was available in a Microsoft Access database.

Within Microsoft Access a query was created to link any sites that had both GIS derived habitat data and data derived manually from paper maps (the sites selected are displayed in Figure 3.1 below). In all, 976 sites were used in the comparison and the data imported into MS Excel and the relationships between different derived features were investigated.
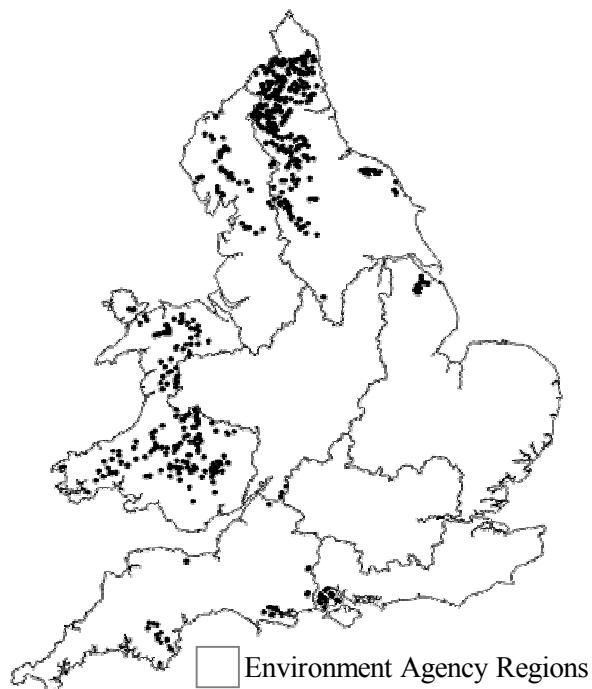
□ Environment Agency Regions

Figure 3.1   The 976 sites used in the GIS derived data against manually derived data, investigation

Catchment area can be calculated from the flow accumulation value, which equates to the number of cells upstream of the site (Equation 3.1).

(Equation 3.1): `Catchment Area (m`^{2}`)= Flow Accumulation * 50m`^{2}
    Flow Accumulation * 50m^2 (because the grid resolution used was 50m x 50m)

River width is extremely difficult to accurately derive from the relatively coarse scale of DEM available. For some studies an analogy for stream size may be such as the Strahler stream order which has been used by Argent *et al*. (2003). However, estimating width would be more useful. Catchment area has been used as a basis for estimating width, as it allows for a step change when large sub-catchments combine (Wyatt, 2001). An attempt to predict river width using a number of the GIS derived variables was undertaken in Minitab.

3.3 Results

3.3.1 Site Altitude

Within GIS, one the most simple attributes to derive for a site would be that of altitude, if you have an accurate DEM as is available in the UK. The altitude for the sites is given in metres above Ordnance Datum (Newlyn). The data set used has a vertical resolution of ± 5m and a horizontal resolution of 50 m. The manually derived data would be taken from the contours on paper based maps. As can be seen in Figure 3.2 there is a good correlation as would be expected. The graph suggests that the two methods are comparable. The outliers highlighted in Figure 3.2 are most likely to be due to location errors. A number of these outliers were investigated to find their source. In the lower part of the graph highlighted, as blue squares are a set of sites where the data collected manually was wrong. The site on the graph highlighted, as red diamonds are a set of sites where the sites were snapped to the wrong river section and hence removable. There was insufficient detail in the site forms to pick this up at the initial auditing stage. Figure 3.3 depicts the sites across the county showing how the identified errors are spread. They are spread across the county indicating that the problem is not due to a particular error in the methods for one particular area of the country.
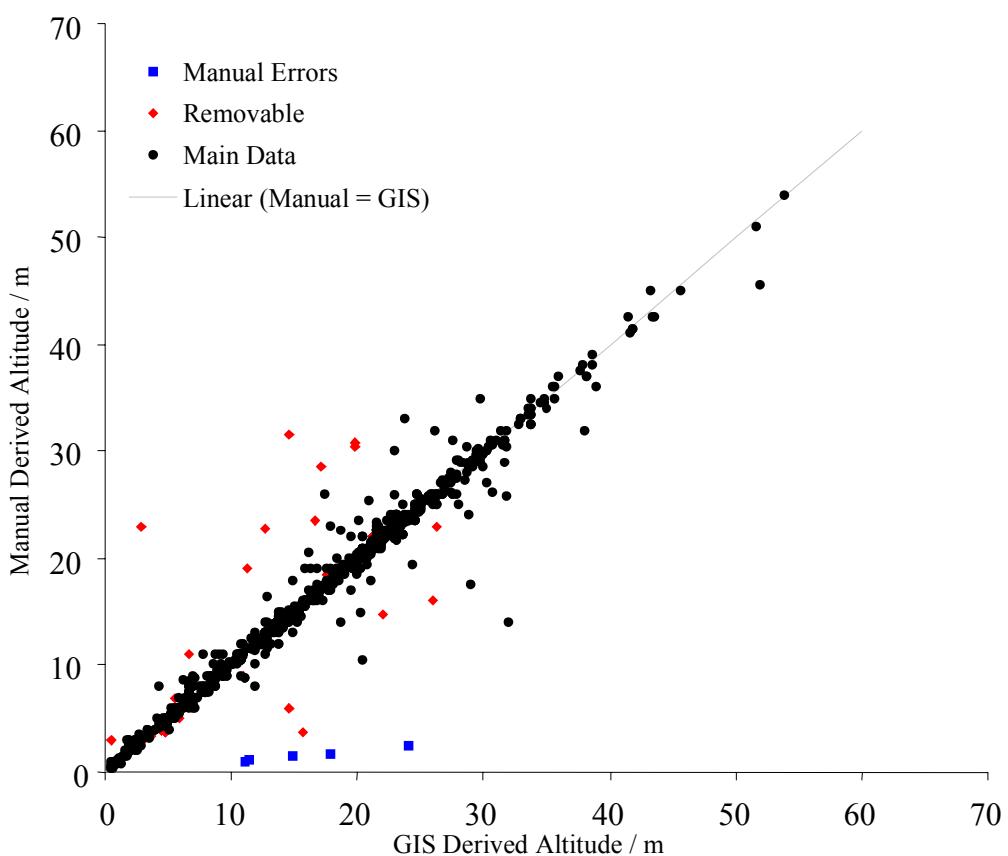
Figure 3.2   Comparing GIS and manually derived altitude data
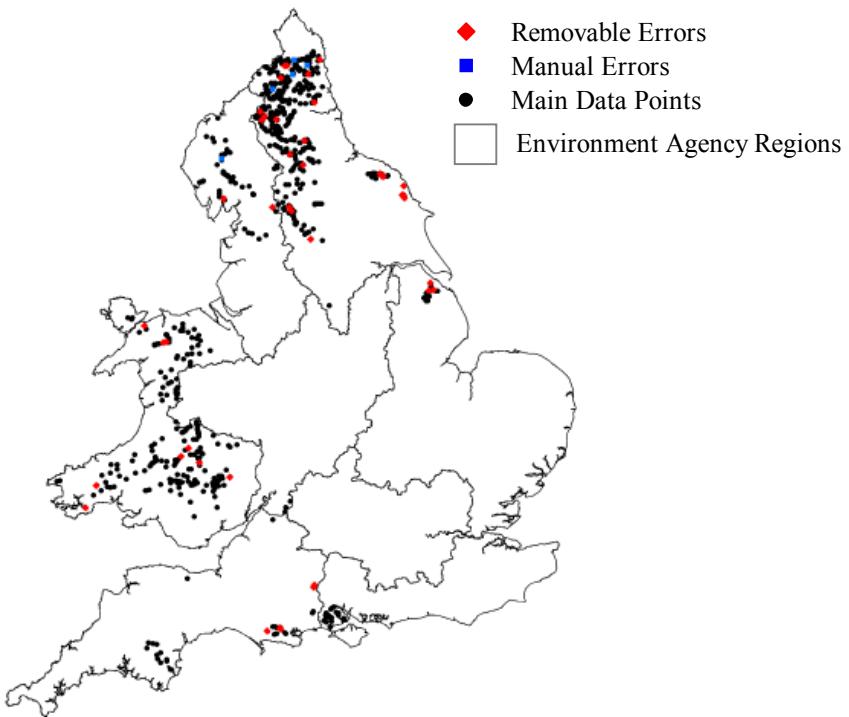
Figure 3.3   Identified errors for the 976 sites used in the GIS derived data against manually derived data, investigation

3.3.2   Distance to Source

Within GIS, an analogy to the distance to source was calculated as distance to the top of the catchment.   This was undertaken, as an easy to calculate attribute that did not require source point data that in not currently available for the whole of England and Wales.   The manually derived data would be taken by tracing the river network on paper based maps. As can be seen in Figure 3.4 there is a good correlation and would suggest that the two are comparable.   There are a number of outliers that are depicted in Figure 3.4.   The errors had been investigated as described in Section 3.3.1 and are highlighted in Figure 3.4 as the red diamonds.   Those highlighted are due to errors in the snapping process.   The initial auditing, due to lack of site details existing did not pick these errors up.
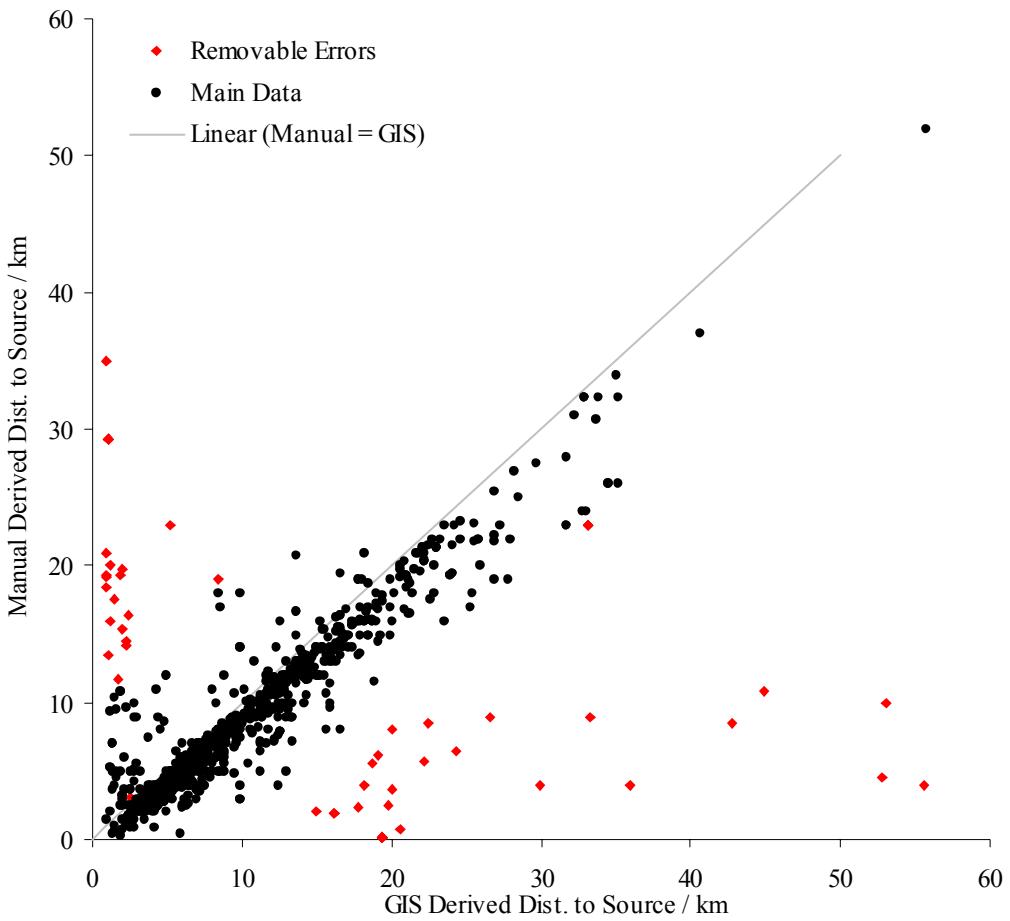
Figure 3.4   Comparing GIS and manually derived distance to source data

3.3.3   Distance to Mouth

Within GIS, the distance to mouth was calculated as distance from a point on the river network to the point where the elevation value became 0 m (above Ordnance Datum). The manually derived data would be taken by measuring the river network on paper-based maps. As can be seen in Figure 3.5 there is a good correlation and would suggest that the two are comparable. The furthest outliers highlighted by the red box in Figure 3.5 are explained by as errors identified in Section 3.3.1 and could be removed from the graph. There are a number of parallel lines that lay parallel to the X=Y line that are highlighted in Figure 3.5, this can be explained as using zero metres elevation as the cut off value will cause some discrepancies where this lies some distance from the 'mouth' identified in the manual method. Figure 3.6 shows an occurrence of this on an example estuary. This discrepancy will result in all the sites on a certain river being the same value away from the GIS derived distance. This can be seen where a number of different rivers have been highlighted, they all run parallel to the directly comparable X-Y line but offset by a set amount that differs for each river / catchment. As new data sets become available it would be possible to recalculate the downstream distances to the tidal limit. Distance to 0 m may be as good as distance to tidal limit in terms of describing the relationship with fish populations.
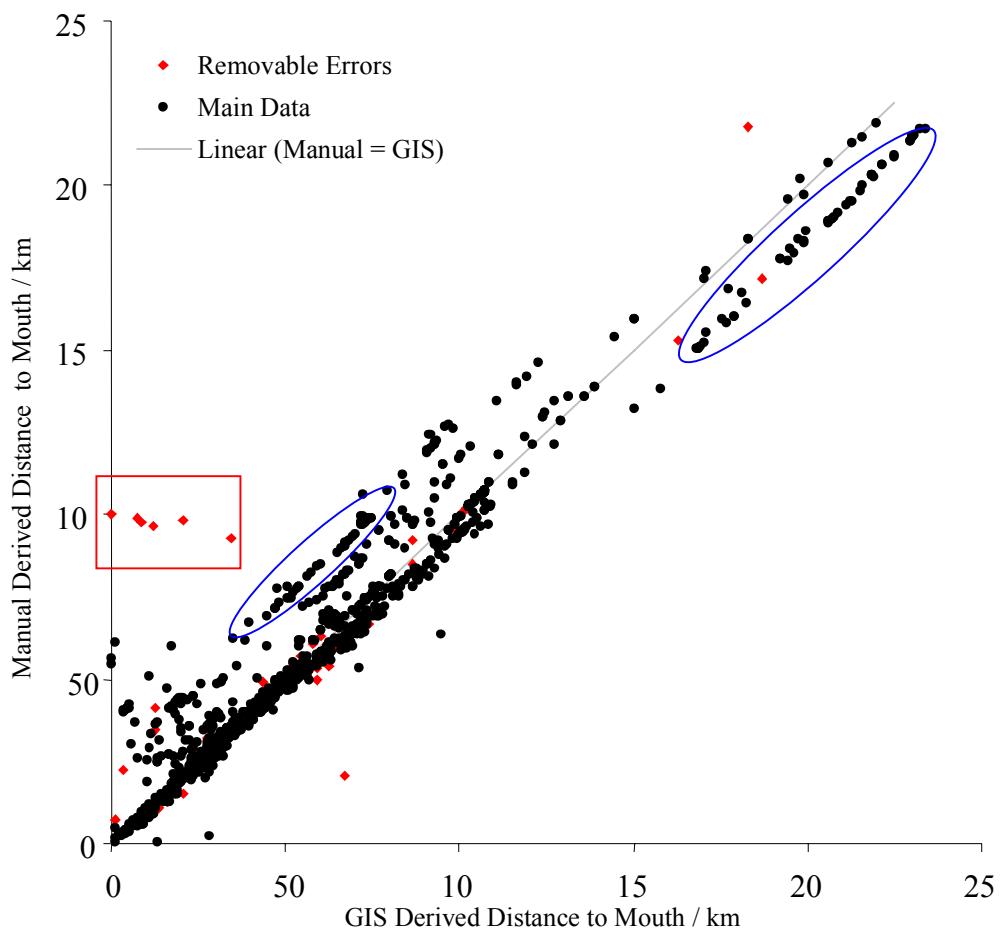
Figure 3.5   Comparing GIS and manually derived distance to mouth data
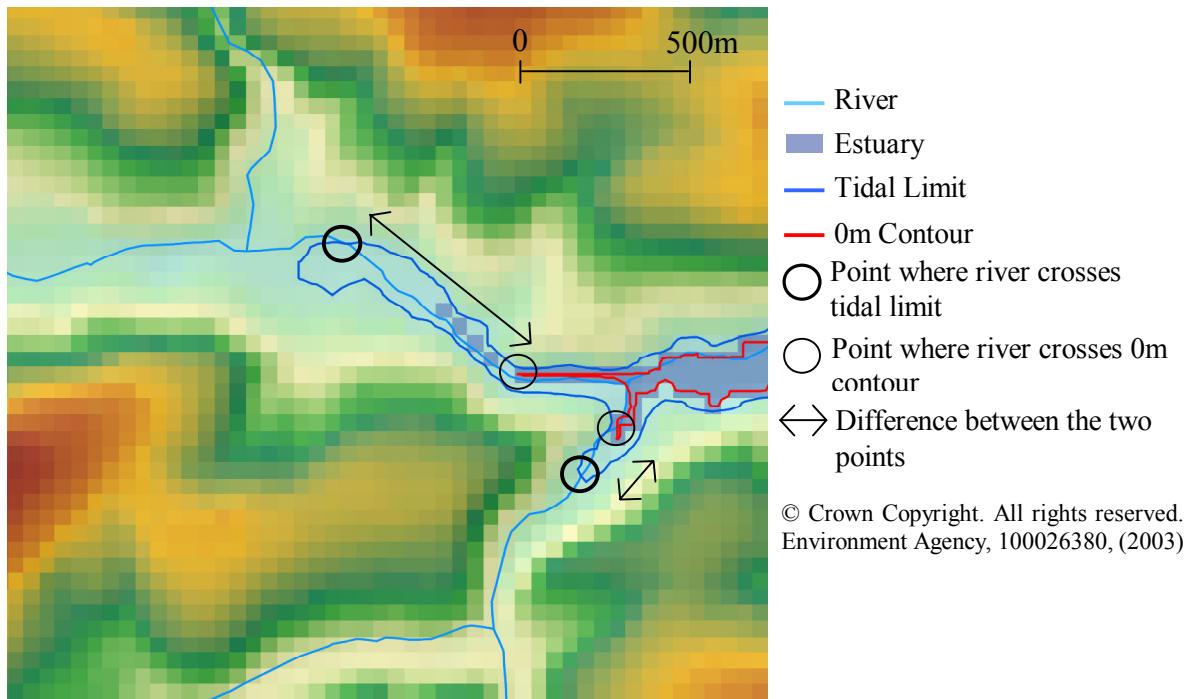
Figure 3.6   Position of tidal limit and elevation at 0 metres

3.3.4   Catchment Area

The GIS derived upstream catchment area is easy to calculate for a cell or site by multiplying the `Flow Accumulation` value for that cell by 50 m². The manual method is time consuming and relies on using grid squares to calculate by tracing an assumed catchment onto gridded paper. Only 222 sites had both GIS derived and manually derived Catchment Area data. However, as can be seen in Figure 3.7 there is a good correlation and would suggest that the two are comparable. A few discrepancies can be seen, these are highlighted in Figure 3.7 as A and B, and the majority of these can be removed as are identified as errors in Section 3.3.1. Some of the remaining discrepancies could easily be explained as the manual method can be inaccurate as it relies on a objective decision where the catchment boundary is and tracing this to gridded paper to calculate the area.
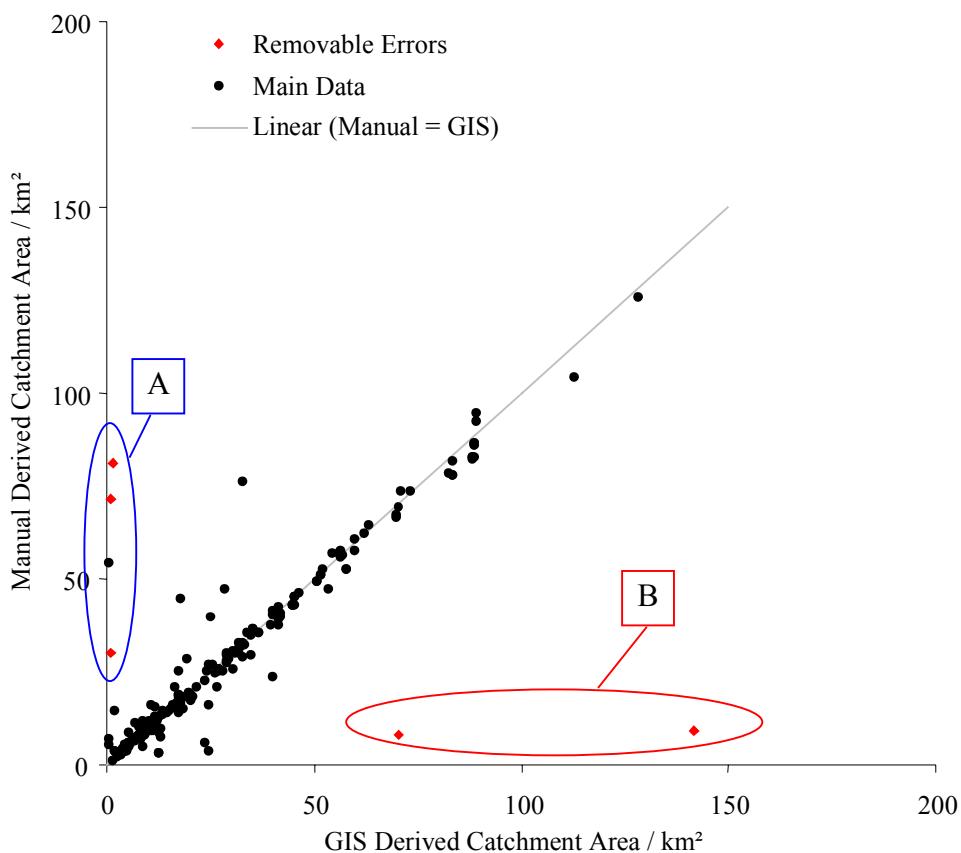
Figure 3.7   Comparing GIS and manually derived catchment area data

3.3.5   Site Gradient

One of the most difficult attributes to calculate in a way that is comparable to the manually derived value is that of site gradient. In GIS, the slope function identifies the rate of maximum change in z value from each cell as either a percentage or angle of slope in degrees. Angle of slope from the GIS method can be converted to gradient (m km^{-1}) as used in the manual method within MS Excel using Equation 3.2.

(Equation 3.2): `Gradient = ((TAN((GIS Slope)*(2*(PI()/360))))*1000`

There appears to be little correlation between the GIS derived and the manually derived methods (Figure 3.8). This is not too surprising for four reasons:

1. The manual method is subject to a large amount of variability. This is due to the arbitrary selection of altitude at points away from the site.
2. The GIS method of determining slope for each cell in the grid highlights the maximum slope that may be the valley side rather than the river slope. Figure 3.9 describes this where the gradient in the A-B axis has a greater change in altitude than that of the slope of the river (axis C-D).
3. GIS derived slope is at its worst at low gradients. With the integer based DEM the GIS will have a minimum drop of $\pm$ 1m this will result in low gradients being either 0 or 1/50.
4. As manual gradients effectively take an average over a wider area they may be less susceptible to the smaller scale variation and the lower likelihood of a very large change in height between to adjacent 50 m cells. This would explain the high number of slope values clustered around zero (Figure 3.8 Box A).
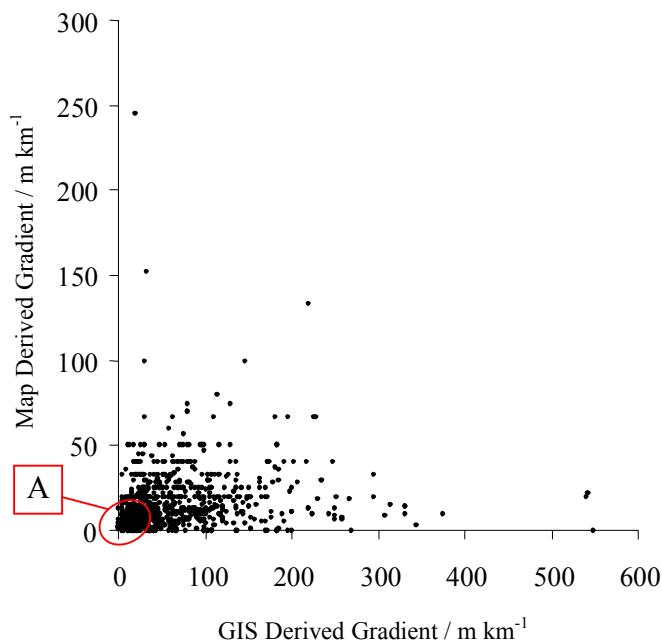
Figure 3.8   Comparing GIS and manually derived site data
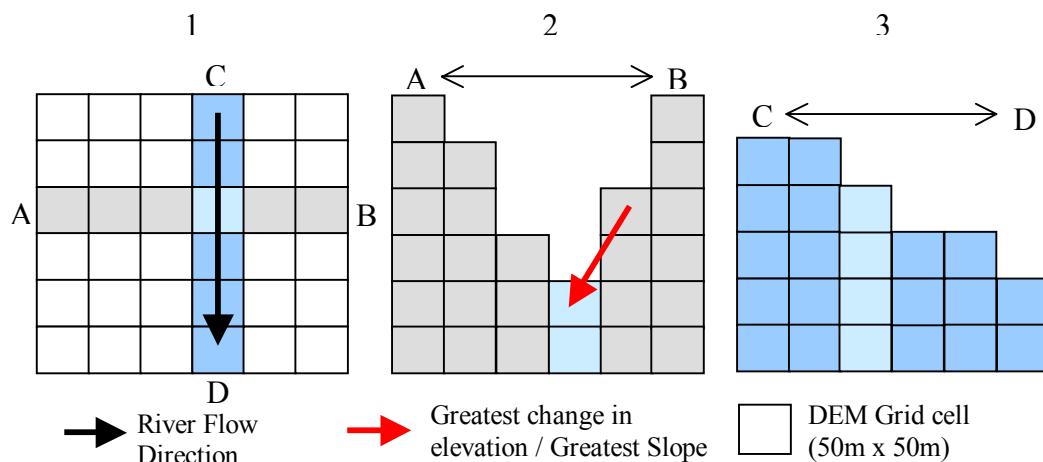
Figure 3.9   GIS derived slope, reasons for over estimation

As an attempt to improve on the calculation of the gradient the DEM was cut to the river network and the slope function calculated on this grid (Section 2.2.6 Slope). This forced the calculation to determine the slope of the river rather than that of the valley sides. The result of this was again compared to the manually derived values as seen in Figure 3.10. There still appears to be little correlation between the two. The data from the two axis were logged (Figure 3.11). The vertical striping that is seen (Figure 3.11 Box A) occurs because the problem of poor vertical resolution producing only a limited number of possible slope values. As can be seen this problem is most marked in the low relief areas. The horizontal striping that is seen (Figure 3.11 Box B) occurs because rounding up or down of the manual derrived data produces only a limited number of possible slope values and may be of particular problem where there are close together contors and measurment accuracy becomes more difficult.

Using the current GIS methods it is not possible to calculate a value that is directly comparible to that of the manual derrived site gradient. The GIS derrived gradient may, however, be as good. If a comparible GIS method to the manual derrived site gradient is required further work would need to be undertaken.
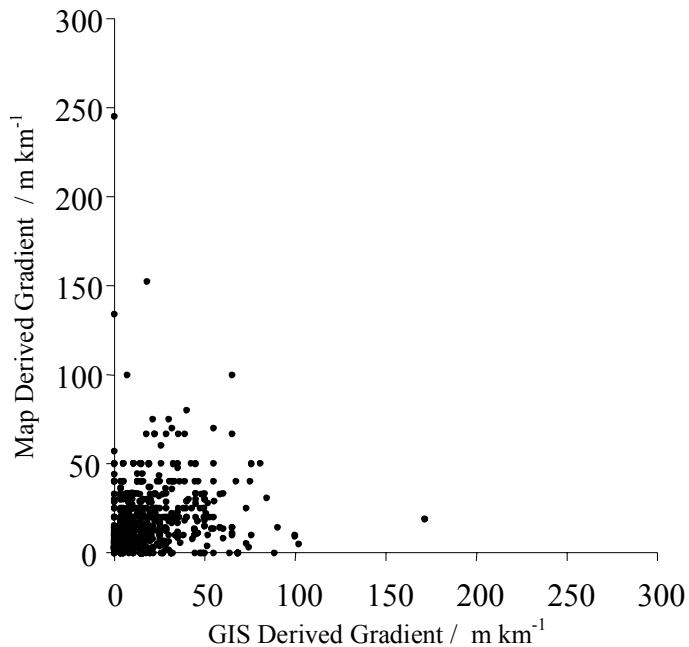
Figure 3.10 Comparing GIS and manually derived gradient with attempted improvements
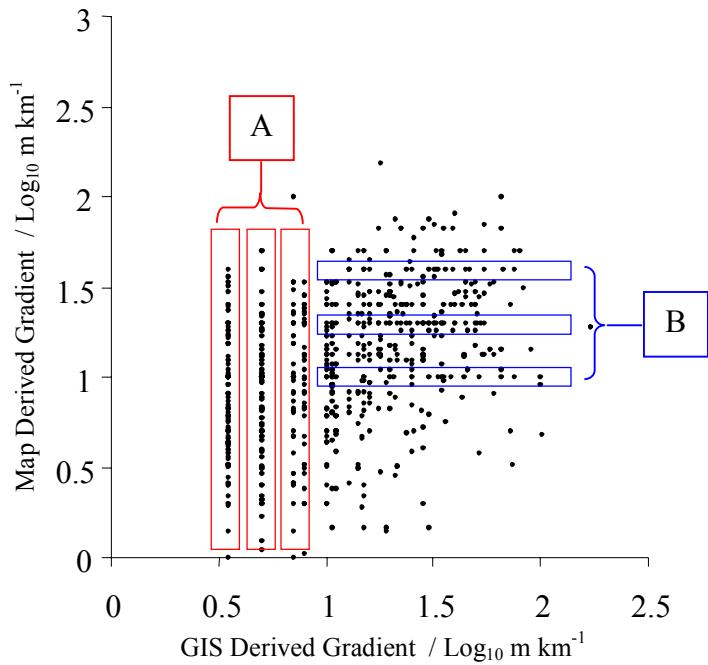
Figure 3.11   Comparing GIS and manually derived gradient data (Log relationship)

3.3.6   Catchment Gradient

Another difficult attribute to calculate that is comparable to the manually derived value is that of catchment gradient. Within the HABSCORE methodology this is defined as the 10:85 slope and is time consuming to calculate. Taking the altitudes at 10% and 85% of the distance upstream of a site and calculating the gradient from this calculate the 10:85 slope. Within the GIS, the slope function identifies the rate of maximum change in z value from each cell. The average slope value upstream of a cell can be calculated and used as a value for catchment gradient. This can be converted to gradient (m km^{-1}) as used in the manual method within MS Excel using Equation 3.3.

(Equation 3.3): `Gradient = ((TAN((GIS Average Slope Upstream) * (2 *(PI()/360))))*1000)`

There appears to be little correlation between the GIS derived and the manually derived methods (Figure 3.12). This is not surprising, as the methods are so different. The GIS derived average slope upstream may, however, be as useful in describing fish populations as the 10:85 slope is within the HABSCORE models.
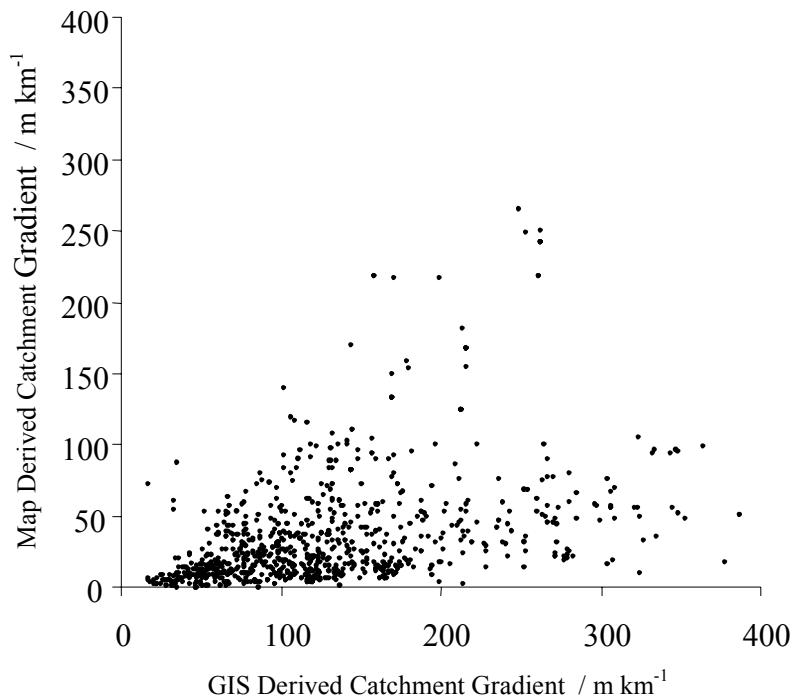
Figure 3.12   Comparing GIS and manually derived catchment gradient data

3.3.7   Stream Width

At present stream width cannot be derived directly from the GIS. Working a raster grid based system does not provide the ability to directly acquire stream widths and the information is not currently available in vector form.

As catchment area has been used as a basis for estimating width (Wyatt, 2001) this was used as the initial regression test within Minitab against the field derived stream width from the fish survey data. Table 3.1 displays the regression results and Equation 3.3 displays the equation of the resulting fit. This relationship is displayed in Figure 3.13 and graphically displays the regression output that suggests that only 29.5% of the variability can be explained by the use of this variable. There is a sharp cut off at the smaller widths caused by the constant used in the equation and the lack of sites with very low catchment areas.

(Equation 3.3): `Field Width = (4.11+(0.0351*[Catchment Area]))`

Table 3.1: Minitab Regression output

```
Predictor              Coef      SE Coef         T        P
Constant             4.1088       0.1057     38.89    0.000
Catchment Area      0.035093     0.001736     20.21    0.000

S = 2.697            R-Sq = 29.5%        R-Sq(adj) = 29.4%
PRESS = 7208.65      R-Sq(pred) = 28.41%
```

Analysis of Variance
```
Source            DF          SS         MS         F        P
Regression         1      2970.6     2970.6    408.43    0.000
Residual Error   976      7098.6        7.3
  Lack of Fit    774      6825.0        8.8      6.51    0.000
  Pure Error     202       273.6        1.4
Total            977     10069.2
```
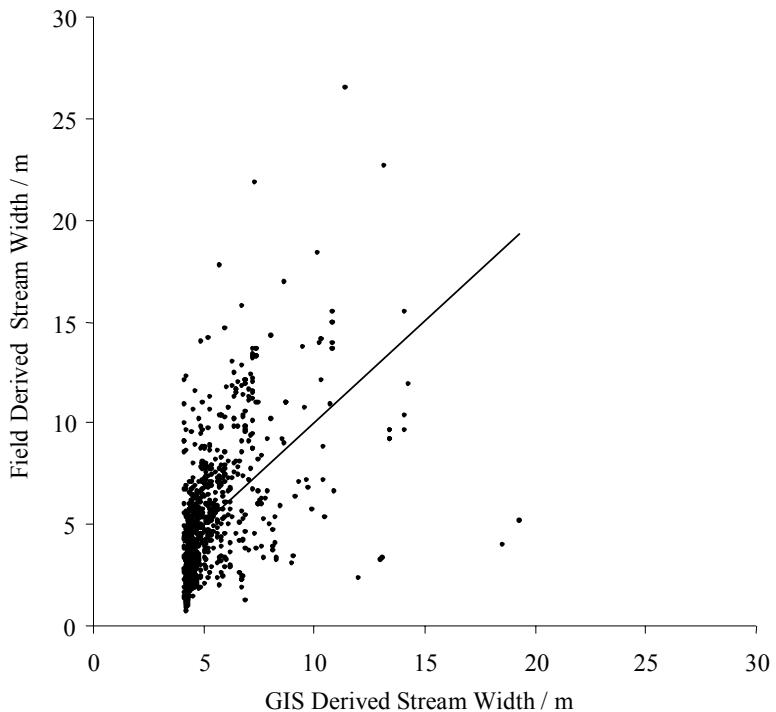

Figure 3.13   Relationship between field derived and GIS derived stream widths

As an attempt to improve the estimation of stream width some other variables were added to the model. Of the resulting estimations the best (in terms of percentage of variability explained) included distance to source and average rainfall upstream. Table 3.2 displays the regression results and Equation 3.4 displays the equation of the resulting fit.

This relationship is displayed in Figure 3.14 and graphically displays the regression output that suggests that 39.1% of the variability can be explained by the use of these variables. Understandably this level of variability would mean that accurate prediction of stream width on an England and Wales scale is difficult. A method has been suggested by Wyatt (2001) that would use geostatistical techniques on a catchment scale to interpolate field-based measurements throughout the catchment. At present this method is time consuming to apply. Ordnance Survey are likely to produce a new river layer within two years that will contain surveyed river widths above one metre for all of

the UK, this should provide a data set with accurate and standardised data for the whole of the river network in a standardised form. Potentially this will negate the need to model river width to predict values.

(Equation 3.4): `Field Width = (0.043 + (0.00691*[Catchment Area]) +(0.000230*[Distance to Source]) + (0.00219*[Site Rainfall Average]))`

Table 3.2: Minitab Regression output

```
Predictor                   Coef     SE Coef          T        P
Constant                  0.0435      0.3508       0.12    0.901
Catchment Area          0.006908    0.003874       1.78    0.075
Distance to Source    0.00022989  0.00002536       9.06    0.000
Site Rainfall Average   0.0021927   0.0002288       9.58    0.000

S = 2.508       R-Sq = 39.1%     R-Sq(adj) = 39.0%
```

Analysis of Variance
```
Source            DF          SS          MS         F        P
Regression         3      3940.9      1313.6    208.78    0.000
Residual Error   974      6128.3         6.3
Total            977     10069.2

Source                   DF      Seq SS
Catchment Area            1      2970.6
Distance to Source        1       392.7
Site Rainfall Average     1       577.7
```
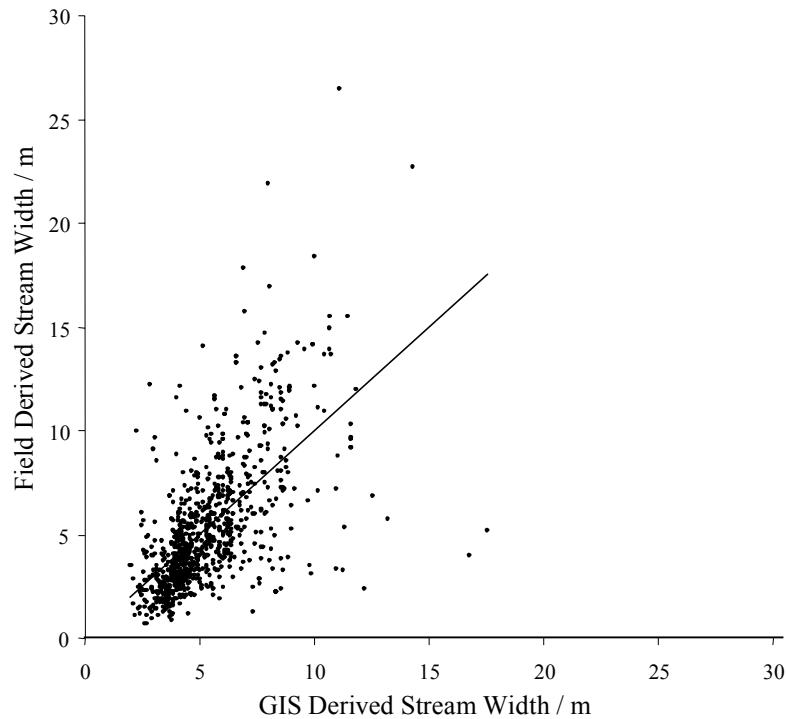

Figure 3.14   Relationship between field derived and GIS derived stream widths

3.4 Conclusions

GIS provides the ability to automate and improve measurement of certain catchment features. GIS based measurements are comparable with the majority of those derived manually from paper maps. Large differences between GIS derived measurements and the manually derived measurements can be avoided if sufficient auditing of location data occurs. Location data is prone to errors when taken from older data sources, especially when grid references are taken from paper maps by hand.

The GIS provides the ability to calculate site altitude, distance to source, distance to mouth and catchment area in an effective way that is comparable to that of manual methods. The GIS, however, provides for this to be undertaken in an automated, time saving manner that is more precise. The other main advantage of the GIS methodology is that the attributes can be calculated for every cell on the river network enabling the application of models to the whole river network.

GIS also provides the opportunity to calculate variables that are not possible or are very time consuming to calculate manually.

3.4.1 Recommendations

There is a requirement to refine the site slope methodology, or find a modification to the method that will provide a meaningful value in terms of fish populations.

The distance to mouth attribute was calculated as distance to zero elevation. This could be re-calculated using the tidal limit layer as distance to tidal limit.

The methodology should be re-run with the improved quality data sets that will become available over the next year. These should include a better vertical resolution DEM and an improved quality river network. Together these datasets should provide the ability to improve or refine the derived river network and the associated attributes.

4. RELATIONSHIP BETWEEN JUVENILE SALMONID POPULATIONS AND CATCHMENT FEATURES

4.1 Introduction

Previous work that has looked at the relationship between juvenile salmonid populations and habitat features such as HABSCORE (Wyatt *et al* 1995 and Barnard *et al* 1995) and the Fisheries Classification scheme (Hay *et al*., 1995, Mainstone *et al*., 1994, and Wyatt and Lacey, 1998), have relied on manually derived map-based variables. These methods have produced statistical models that have related the map-based variables to the juvenile salmonid populations.

These have included: Distance to source, catchment area, site altitude, site gradient, stream order, catchment gradient and distance to tidal limit.

The common method of investigating the relationships between populations and habitat variables is by using multiple regression analysis. The advantage of such a method is the production of an equation that describes the relationship between the variables and the fish population. This equation can be converted into a format that can be used within the GIS to produce maps describing the predicted distribution from the model.

4.2 Methods

4.2.1 Data Sources

The fish data were collected together from four main sources:
1. The original HABSCORE dataset (Wales),
2. The second HABSCORE survey dataset (England and Wales),
3. Any subsequent HABSCORE survey data,
4. Fully Quantitative surveys from the National Survey database.

The three HABSCORE data sets differ only slightly. The original HABSCORE data set had a number of different fields that had to be converted to be the same as the other data sets. The second HABSCORE data set was the main HABSCORE model development survey data and the remaining HABSCORE data set was any subsequent HABSCORE survey data.

The data were altered to be in the same format in terms of fish density and same table structure within MS Access, and then merged. The structure of the resulting table is shown in Table 3.1. The data was stored in MS Access to allow for the storage and querying of the data.

Table 4.1   Fish Data table structure

Field Name	Data Type	Description	Field Length
FSDate	Date/Time	Fish Survey Date	N/A
Old_NGR	Text	Original National Grid Reference ¤	15
New_NGR	Text	Corrected National Grid Reference	15
Year	Long Integer	Year	N/A
DaysFrom	Long Integer	Days from 1st Jan on year of survey	N/A
SiteName	Text	Site Name	100
RiverName	Text	River Name	100
SiteCode	Text	Site Code	100
CatchName	Text	Catchment Name	100
Species	Text	Species - Salmon / Trout	100
AgeClass	Text	Age Class - Fry / Parr / >20cm / <20cm	25
Density	Double	Density per 100m^2	N/A
PopEst	Double	Population Estimate	N/A
VarPopEst	Double	Variance of Population Estimate	N/A
Source	Text	Source of Data – ND[Q]/HD/OW	5

ND[Q] = National Data – Full Quantitative Surveys.
HD = Second and subsequent HABSCORE surveys.
OW= Original Welsh HABSCORE survey.
¤ = NGR's from original survey sheets

There were 1989 sites with from 1 up to a maximum of 16 surveys. The sites are predominantly located in the main salmonid regions with the exception of the Midlands region where data for fully quantitative juvenile salmonid surveys was not available. Figure 4.1 shows the sites used in the modelling process and the area used within the GIS for modelling.

■ Area Selected for Modelling
□ Major Hydrological Area
• Sites
© Crown Copyright. All rights reserved.
Environment Agency, 100026380, (2003)
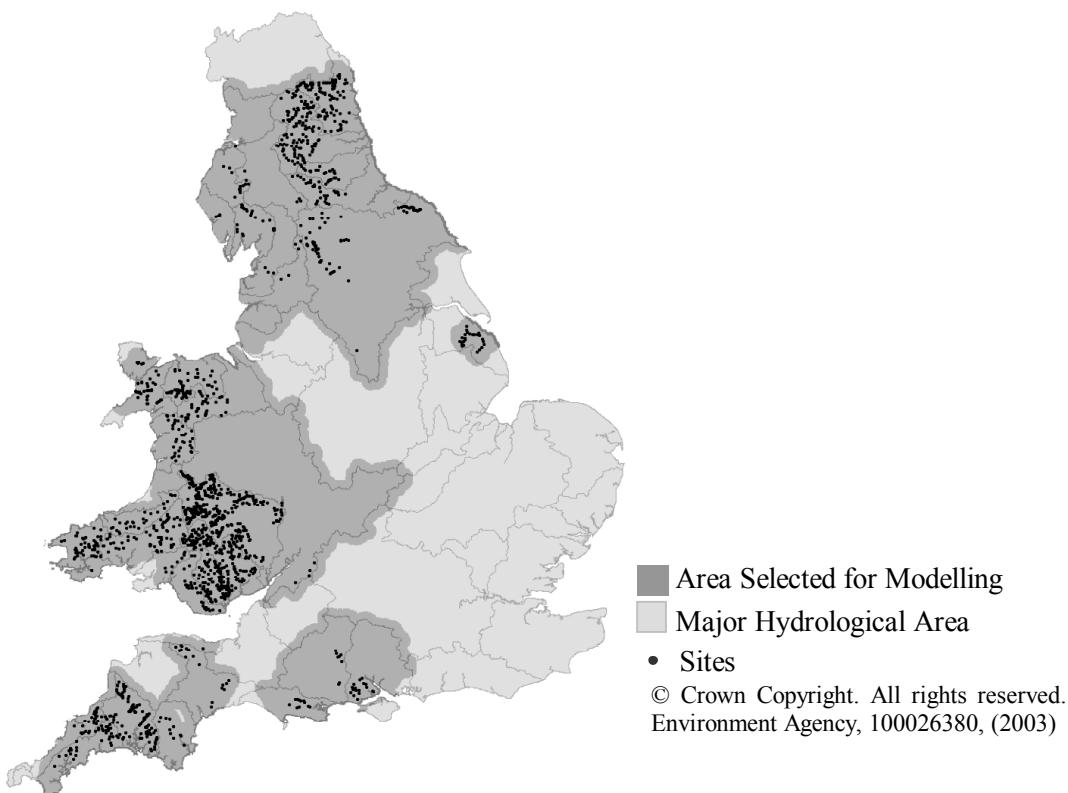
Figure 4.1   Sites used within the GIS and subsequent relationship modelling

The fish survey data are predominantly from 1985 to 1998 but cover the period from 1974 to 2001 (Figure 4.2). This period covers the initial HABSCORE development survey in Wales and the subsequent development surveys for HABSCORE in England and Wales. This does not necessarily represent the number of fully quantitative surveys undertaken in a given year but rather the availability of the data in a usable electronic format.

Of the 1989 sites the majority were surveyed only once. Of the sites with repeat surveys the majority was 2 surveys (1 repeat). There were a number of sites with numerous repeat surveys with the maximum being of 16 surveys, which was due to a number of repeat surveys (Figure 4.3) being undertaken on the same year (pseudo-replication). There may be a small level of duplication of data from different sources. However, this will be removed by the averaging of density data for each site. However, there are sites that have had 10 – 15 separate surveys undertaken in the 26 years between 1975 and 2001. The graph does not necessarily portray a drop off in quantitative surveys in the periods 1990-1991, 1994-1996, and 1999-2001 but reflects the data that was collected by a national trawl to area fisheries teams.
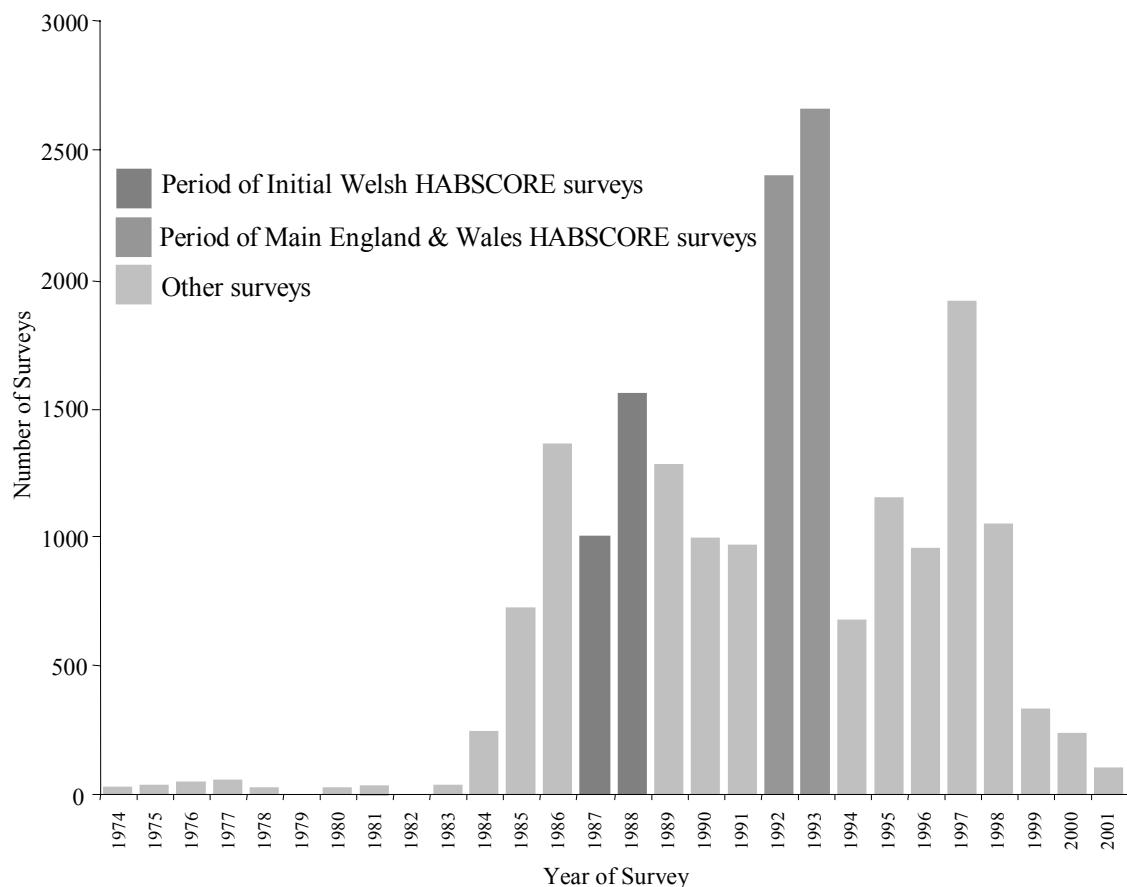
Figure 4.2   Number of fully quantitative surveys undertaken in each year (of surveys held within database)
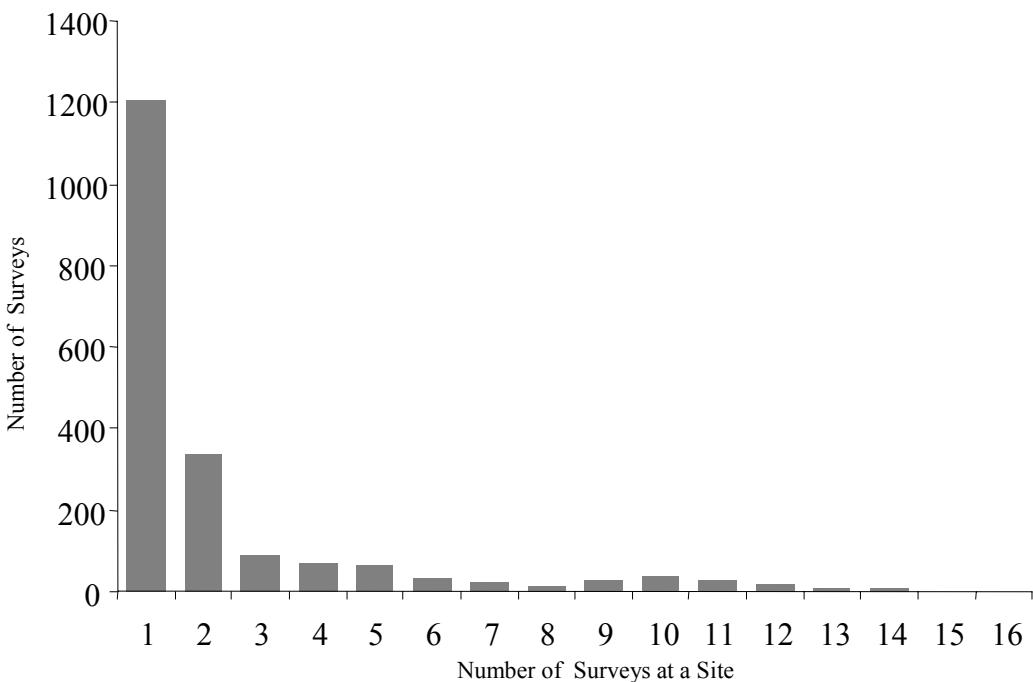
Figure 4.3   Example of distribution of number of sites that have had repeat surveys (trout fry)

Fish Data

Variation in fisheries survey data can be categorised as spatial, temporal or as random. Each category poses different issues. In this study the spatial variation is of interest because we are dealing with spatial data sets to explain the spatial variation. The temporal and random variation masks the relationships that we are looking for. We cannot do much about the random variation but the temporal variation may be reduced somewhat by having multiple surveys for the same site.

Averaging the densities over each survey at a site will reduce the overall temporal variation of the data. However, the majority of sites have only one survey so this reduction of temporal variation is not equal over all sites. As can be seen in Figure 4.4 there is a much greater range of densities at the sites where only one survey has been undertaken. This will be due both to there being a larger number of sites and hence diversity of values and more spatial variation but also due to increased temporal variation.
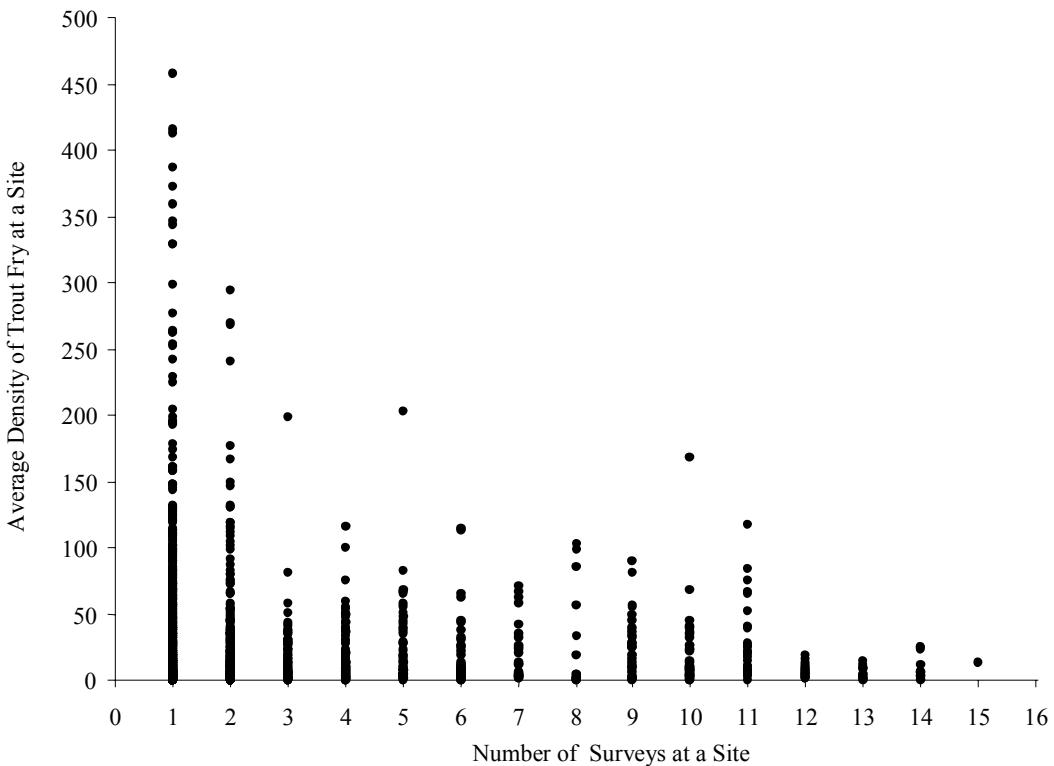
Figure 4.4   Example of distribution of fish densities verses number of surveys undertaken at a site

If survey data has been derived from counts some difficulties present themselves when deciding how best to deal with the survey data. Log density is often a desired transformation of fisheries data. Fisheries data does not follow a normal distribution in relation to a fixed variable value and will generally be skewed to the lower end (close to zero); a log transformation brings the data closer to a normal distribution, in order to fulfil the assumptions of regression analysis. But this poses a particular problem, dealing with zero densities. Log transformations cannot be undertaken on values of zero. Log (Density+1) is commonly used as a quick method of dealing with the zero values. This method is not ideal and would not be appropriate for a more detailed study. Adding this arbitrary constant will result in the value being modelled to varying significance at different ends of the data's range. A further complication is that the number of fish will not change in direct proportion to the site area (Wyatt *et al*. 1995). There are a number of alternative approaches such as discarding observations with no fish. However, many alternative techniques have their own disadvantages. The aim of this section is to look at the suitability and effectiveness of the GIS techniques for understanding juvenile salmonid populations and not a detailed study repeating a HABSCORE type model or duplicating existing work as part of the River Fisheries Habitat Inventory project.

For the trout fry, log (density+1) was used for simplicity. However, for salmon fry a method was required to select and remove those sites that would be inaccessible to adult salmon. The easiest way to achieve this was to remove the zero densities, assuming that the presence of juvenile salmon means that access is sufficient for adult salmon. Because the zeros were removed the use of log (density+1) was not required and so for salmon fry log density was used.

4.2.2 Linking GIS and Fish Data

Data was extracted for each of the sites from the GIS using the methods described in Section 2 (Using GIS to extract catchment and site based variables). These data were linked to the fish density/survey data using the query functions of MS Access to produce a set of tables that could be used in either MS Excel or Minitab for analysis.

4.2.3 Analysis

Initial work was undertaken in MS Excel to plot simple relationships between two variables and to add extra fields as transformations of the data. The majority of the analysis was undertaken within Minitab (Version 13.1 for Windows) using two main techniques.

1. Best Subsets Regression to help in the identification of suitable variables for the model.
2. Regression function to perform multiple regressions. This was for the refinement and testing of the final models.

Best Subsets is an efficient way to select a group of "best subsets" for further analysis by selecting the smallest subset that fulfils certain statistical criteria (Minitab, 2000). This was used to select those site and catchment variables that were most likely to explain the variation in juvenile salmon or trout densities.

Multiple regressions of the selected site and catchment variables were undertaken to further refine the models and add or remove variables if appropriate. A number of transformations of the data were examined where appropriate.

4.2.4 GIS Modelling

The models produced from the multiple regressions that explain the variation in the juvenile salmon or trout densities were transposed to allow them to be used within Spatial Analysis' Raster calculator.

This application of the equation within the GIS allows for the production of a river network attributed by the predicted fish density following the model. This was undertaken and a coloured river network produced for a number of models.

4.3 Results

This section is split into five main sub-sections:
1. Explaining Variability in Fish Populations with Slope Variables
2. Single Variable Relationships,
3. Trout Fry Multiple Variable Relationships,
4. Salmon Fry Multiple Variable Relationships and
5. Application of Models within the GIS.

The purpose is to show the stages of model development and the application within the GIS to provide a visual interpretation of the models. Initially slope was looked at to compare the usefulness of the GIS derived slope, which was shown in Sections 3.3.5 and 3.3.6 to bear little comparability to that of manually derived gradient measurements.

Trout fry and salmon fry were chosen to provide examples for the production of models and the methodology for applying these models within the GIS.

4.3.1 Explaining Variability in Fish Populations with Slope Variables

Section 3.3.5 shows that the current GIS method of calculating slope does not produce values that are comparable to the site or catchment gradient calculated manually using map contours. However, the GIS derived slope may still explain the variation in fish densities as well or better than the manual methods.

To investigate this a number of regressions were performed and the results compared. As in the Fisheries Classification Scheme (FCS) distance to source or catchment area was used as an analogy for stream size. This removes the variability within fish density caused by different stream sizes. Separate regressions were undertaken for site slope both GIS and manually derived and for catchment gradient (or average gradient upstream) both GIS and manually derived. The regressions used Log (base 10) density (+1) values of salmon fry data from the 976 sites shown in Figure 3.1.

Site Slope

The regressions were performed within Minitab and the resulting outputs are displayed in Table 3.3 for the GIS derived method and Table 3.4 for the manually derived method.

The P value for the GIS method suggests that this is not a good model and does not help to explain the variability within the fish density data. The P value for the manual method supports the rejection of the null hypothesis that the coefficients in the model are zero. This model would appear to be plausible, although the R^2 value suggests that this model explains only a small amount of the variability in the fish density.

Comparing the GIS and the manual method suggests that the current GIS site slope is not adequate to explain some of the variability in fish populations. Further work is required to reproduce site gradient within the GIS in a similar way to the manual method that is automated for the whole river network.

The poor performance of GIS derived site slope compared to manual methods in terms of explaining variability in fish populations may be because of a number of reasons. For example numerous zero densities may mask any small relationships. Also, at larger spatial scales the greater diversity of data would result in more difficult model production and result in a greater level of unexplained variability. However, the same sites and data were used for the GIS and the manual models. Although site slope has not been found to be of use for England and Wales as a whole this does not preclude relationships being found at the catchment scale, where the variability of other factors may be reduced. The manual method appears to be more relevant to fish ecology than the current GIS derived values. This indicates that further work is required to extract automatically comparable attributes to the manually derived ones.

Table 4.2   Minitab Regression output GIS Derived Site Slope

```
The regression equation is:
Log10(Density+1)  =  0.420  +  0.0819  Log10(Site  Slope+1)  +  0.00233
Catchment Area


Predictor                 Coef       SE Coef       T        P       VIF
Constant                  0.41998    0.06001       7.00     0.000
Log10(SiteSlope+1)        0.08193    0.08604       0.95     0.341    1.1
Catchment Area            0.0023281  0.0005003     4.65     0.000    1.1


S = 0.7414            R-Sq = 2.3%          R-Sq(adj) = 2.1%
PRESS = 511.879       R-Sq(pred) = 1.65%
```

Analysis of Variance
```
Source            DF        SS          MS          F         P
Regression         2     11.9601      5.9800      10.88     0.000
Residual Error   925    508.4949      0.5497
  Lack of Fit    758    505.8941      0.6674      42.85     0.000
  Pure Error     167      2.6008      0.0156
Total            927    520.4550
```

Table 4.3   Minitab Regression output Manually Derived Site Gradient

```
The regression equation is:
Log10(Density+1)  =  0.796  -  0.265  Log10(Site  Gradient)  +  0.00126
Catchment Area


Predictor                 Coef       SE Coef       T        P       VIF
Constant                  0.79618    0.08512       9.35     0.000
Log10(Site Gradient)     -0.26545    0.06822      -3.89     0.000    1.3
Catchment Area            0.0012591  0.0005523     2.28     0.023    1.3


S = 0.7418            R-Sq = 3.9%          R-Sq(adj) = 3.7%
PRESS = 492.833       R-Sq(pred) = 3.34%
```

Analysis of Variance
```
Source            DF        SS          MS          F         P
Regression         2     20.074      10.037      18.24     0.000
Residual Error   890    489.771       0.550
Total            892    509.845
```

Catchment Slope

The regressions were performed within Minitab and the resulting outputs are displayed in Table 3.5 for the GIS derived method and Table 3.6 for the manually derived method.

The P value for the GIS method suggests that this is a plausible model and will help to explain the variability within the fish density data. The P value for the manual method also supports the rejection of the null hypothesis that the coefficients in the model are zero. This model would also appear to be plausible. The R^2 is improved for the GIS method, and would suggest that the GIS method models the variability in the fish density better than the manual method. However, the GIS method shows a very low description of the variability. This is as expected as many other attributes would contribute to the spatial variation of the data and only two variables would not be expected to explain a high proportion of the total variation.

The relationship between the GIS and manually derived catchment gradient as described in Section 3.3.6 suggests that the GIS method might not be appropriate. However, comparing the GIS and the manual method suggests that the GIS catchment slope may be a suitable attribute to explain some of the variability in fish populations.

Table 4.4   Minitab Regression output GIS Derived Catchment Gradient

```
The regression equation is:
Log10(Density+1)= 0.155 + 0.0420 Av Slope U/S + 0.00249 Catchment Area


Predictor           Coef        SE Coef      T        P        VIF
Constant            0.15503      0.05386      2.88     0.004
Average Slope U/S   0.041996     0.006045     6.95     0.000    1.0
Catchment Area      0.0024894    0.0004716    5.28     0.000    1.0


S = 0.7232              R-Sq = 7.1%         R-Sq(adj) = 6.9%
PRESS = 486.774         R-Sq(pred) = 6.47%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	36.707	18.353	35.09	0.000
Residual Error	925	483.748	0.523		
Lack of Fit	758	481.148	0.635	40.76	0.000
Pure Error	167	2.601	0.016		
Total	927	520.455			

Table 4.5   Minitab Regression output Manually Derived Catchment Gradient

```
The regression equation is:
Log10(Density+1)  =  0.574  -  0.00266  Catchment  Gradient  +  0.00168
CatchArea


Predictor            Coef         SE Coef      T        P        VIF
Constant             0.57412      0.04409      13.02    0.000
Catchment Gradient  -0.0026580    0.0008260    -3.22    0.001    1.1
Catchment Area       0.0016810    0.0005055    3.33     0.001    1.1


S = 0.7377              R-Sq = 3.3%         R-Sq(adj) = 3.1%
PRESS = 506.389         R-Sq(pred) = 2.70%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	17.0967	8.5483	15.71	0.000
Residual Error	925	503.3583	0.5442		
Total	927	520.4550			

4.3.2   Single Variable Relationships

In order to get an idea of the likely relationships between the fish densities and the catchment variables derived from ArcView some matrix plots were produced in Minitab. Figure 4.5 and 4.6 depict these plots for trout fry and salmon fry. The plots also display some of the inter-attribute relationships that may exist.
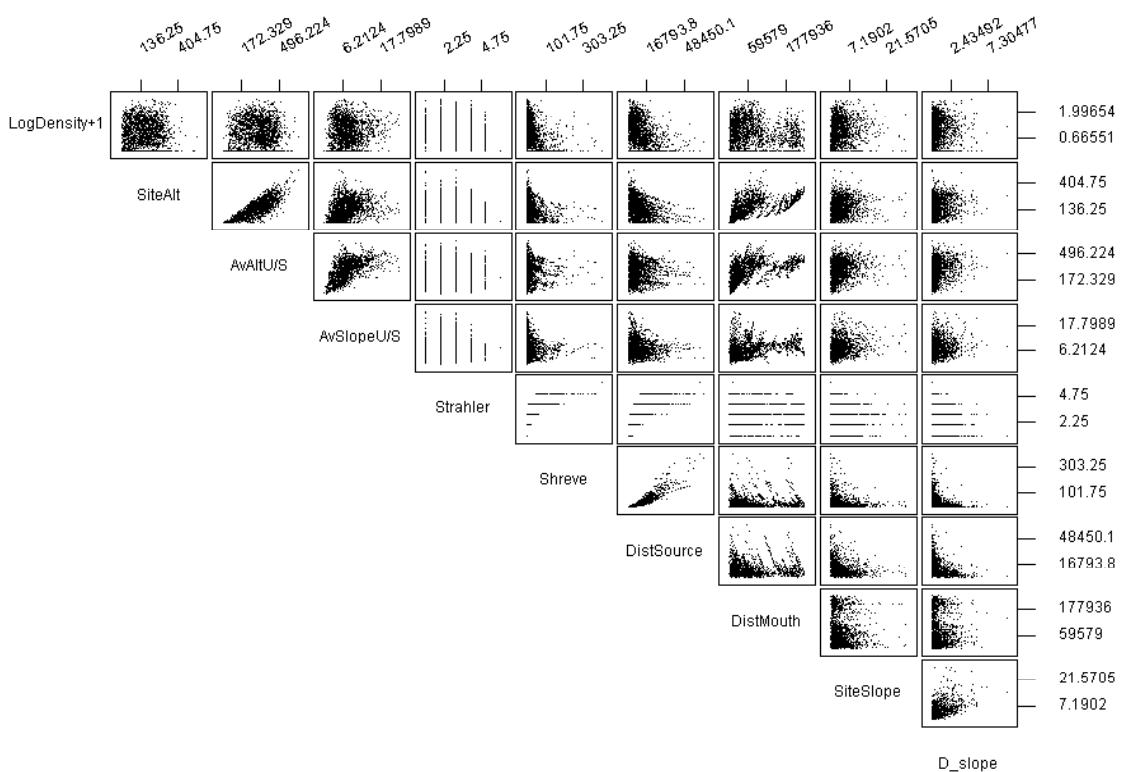
Figure 4.5 **Example of single attribute relationships, Matrix plot of Log (Density+1) trout fry verses selected GIS derived variables**
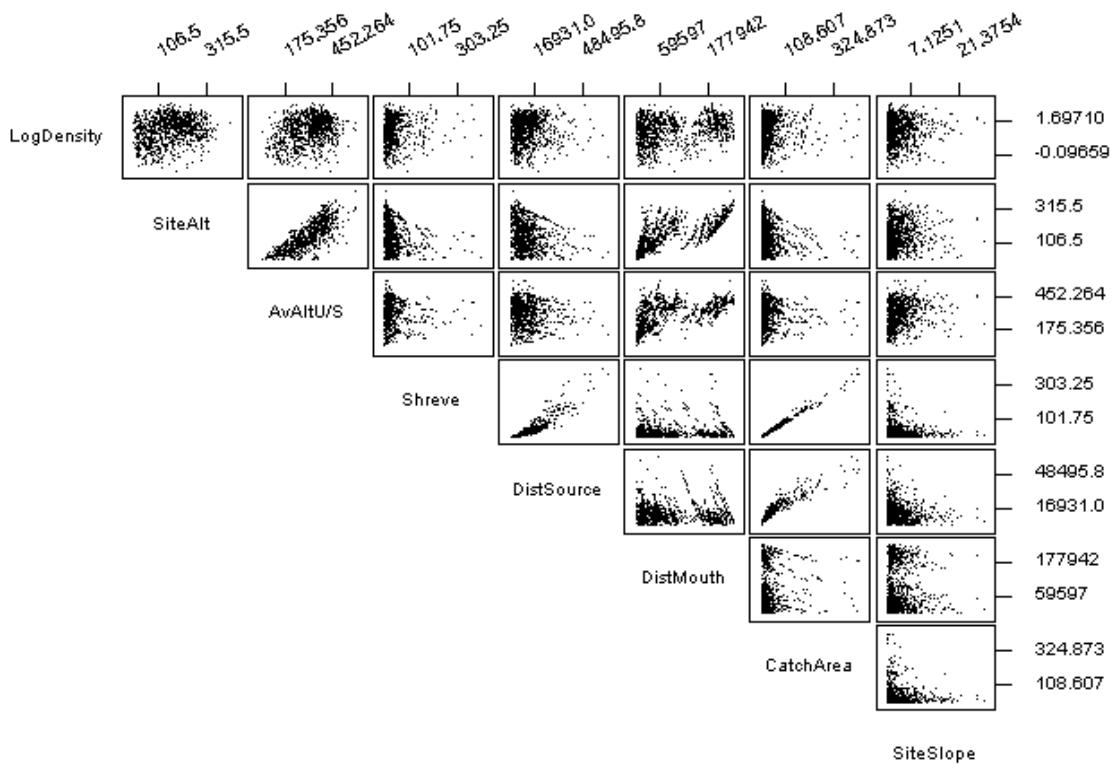

Figure 4.6 **Example of single attribute relationships, Matrix plot of Log Density salmon fry verses selected GIS derived variables**

The two figures (4.5 and 4.6) show that there are no clear single variable relationships between either salmon or trout fry densities and the catchment or site variables. However, they clearly show that there are relationships between a number of the variables as should be expected.

These two points lead to the requirement for multivariable relationships to be investigated. Care must be taken in choosing variables because of the inter-relationship between a number of them.

4.3.3 Trout Fry Multivariable Relationships

Trout fry data in the form of log (density + 1) was analysed in relation to the GIS derived variables. Initially the best subsets regression was used as an efficient way to select a group of "best subsets" for further analysis. Initially the idea was to come up with a model based on the sort of variables that may be used in current relationship models such as HABSCORE. Table 4.2 displays the Minitab output of the best subsets regression performed. Primarily it suggested a number of variables for further analysis these were:
1. Site Altitude,
2. Average Altitude upstream,
3. Average Slope Upstream,
4. Site Slope
5. The two distance attributes,
6. The two distance attributes in combination as "Position" (Table 2.3).

Table 4.6   Best Subsets Regression Minitab output for trout fry

```
Response is Log (Density+1)

Vars   R-Sq     R-Sq(adj)      C-p         S    A B C D E F G H I J

  3    13.6        13.5       46.0    0.61108    X         X X
  3    13.3        13.2       53.4    0.61220              X X       X
  4    14.6        14.4       24.7    0.60769    X         X X       X
  4    14.2        14.0       35.1    0.60927  X X         X X
  5    15.0        14.8       16.3    0.60626  X X         X X       X
  5    14.8        14.5       23.1    0.60729    X         X X X     X
  6    15.4        15.1       10.2    0.60516  X X         X X X     X
  6    15.2        14.9       15.6    0.60600  X X X       X X       X
```

```
A= Site Altitude
B= Average Altitude U/S
C= Average Slope U/S (Catchment Gradient)
D= Stream Order Strahler
E= Stream Order Shreve
F= Distance to Source
G= Distance to Mouth
H= Position
I= Catchment Area
J= Site Slope
```

Multiple regressions of the selected site and catchment variables were undertaken to further refine the models and add or remove variables when appropriate. A number of transformations of the data were also looked such as log10 and X^2 terms, where X^2

terms rely on the removing the mean value to prevent correlation between the X and X^2 terms in the regression analysis.

Initial analysis suggested that site slope could be removed, as it was not a significant predictor as determined by the P-values derived from regression analysis in Minitab. Further analysis suggested that either catchment area or stream order (Shreve) provided a better predictor of variation than did either or both of the distance attributes. Figure 4.7 shows why catchment area was chosen as it could provide a better description of the variation at the lower end of the scale.

Equation 4.1 displays the final model produced. This is termed the trout fry "basic model" as no new attributes (for example landcover) were included. Table 4.3 displays the Minitab output of this regression and Figure 4.8 displays the residual plot outputs from Minitab. These residual plot outputs show how well the model fits the assumptions of a regression model. The plot that shows some concern is that of the residuals verses fitted, this suggests on initial viewing that the variance is increasing on the x-axis. However, the line created by the zero densities may exaggerate this. These may not have such a large effect on the model itself.
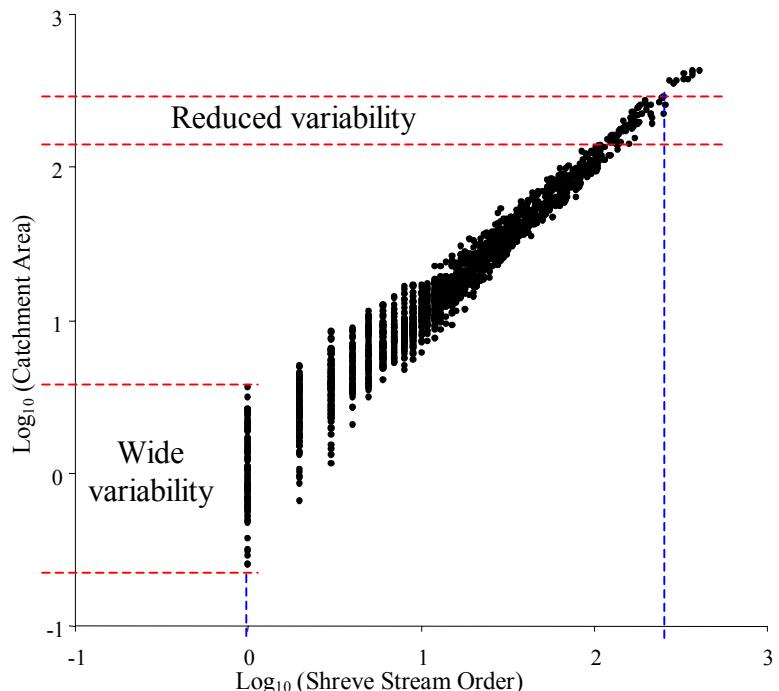

Figure 4.7   Choosing Catchment Area over Stream Order (Shreve) to represent stream size

(Equation 4.1): $\delta = 1.65 - 0.486\ A - 0.239\ A^2 - 0.00297\ B + 0.00236\ C - 0.000007\ C^2 - 0.394\ D - 0.696\ D^2$

Where:

δ = Log (Density +1) Trout Fry

A = Log_{10} Catchment Area - $\bar{x}$ ($\bar{x}$ = 1.183)

B = Site Altitude

C = Average Altitude U/S - $\bar{x}$ ($\bar{x}$ = 300.8)

D = Log_{10} Average Slope U/S - $\bar{x}$ ($\bar{x}$ = 0.284)

Table 4.7   **Multiple Regression Minitab output for trout fry, Basic Model**

```
Predictor         Coef           SE Coef           T        P
Constant          1.64615        0.06328           26.01    0.000
A                 -0.48552       0.02824           -17.19   0.000
A²                -0.23868       0.03058           -7.80    0.000
B                 -0.0029726     0.0003642         -8.16    0.000
C                 0.0023630      0.0003321          7.12    0.000
C²                -0.00000698    0.00000089        -7.82    0.000
D                 -0.3936        0.1037            -3.80    0.000
D²                -0.6963        0.1887            -3.69    0.000

S = 0.5970        R-Sq = 17.7%         R-Sq(adj) = 17.4%
```

Analysis of Variance

```
Source           DF          SS        MS          F        P
Regression        7       150.937    21.562       60.50    0.000
Residual Error  1969      701.798     0.356
  Lack of Fit   1925      696.944     0.362        3.28     0.000
  Pure Error      44        4.854     0.110
Total           1976      852.736
```

Residuals Versus the Fitted Values

(response is Log (Density +1))

Normal Probability Plot of the Residuals

(response is Log (Density + 1))

Histogram of the Residuals

(response is Log (Density +1))
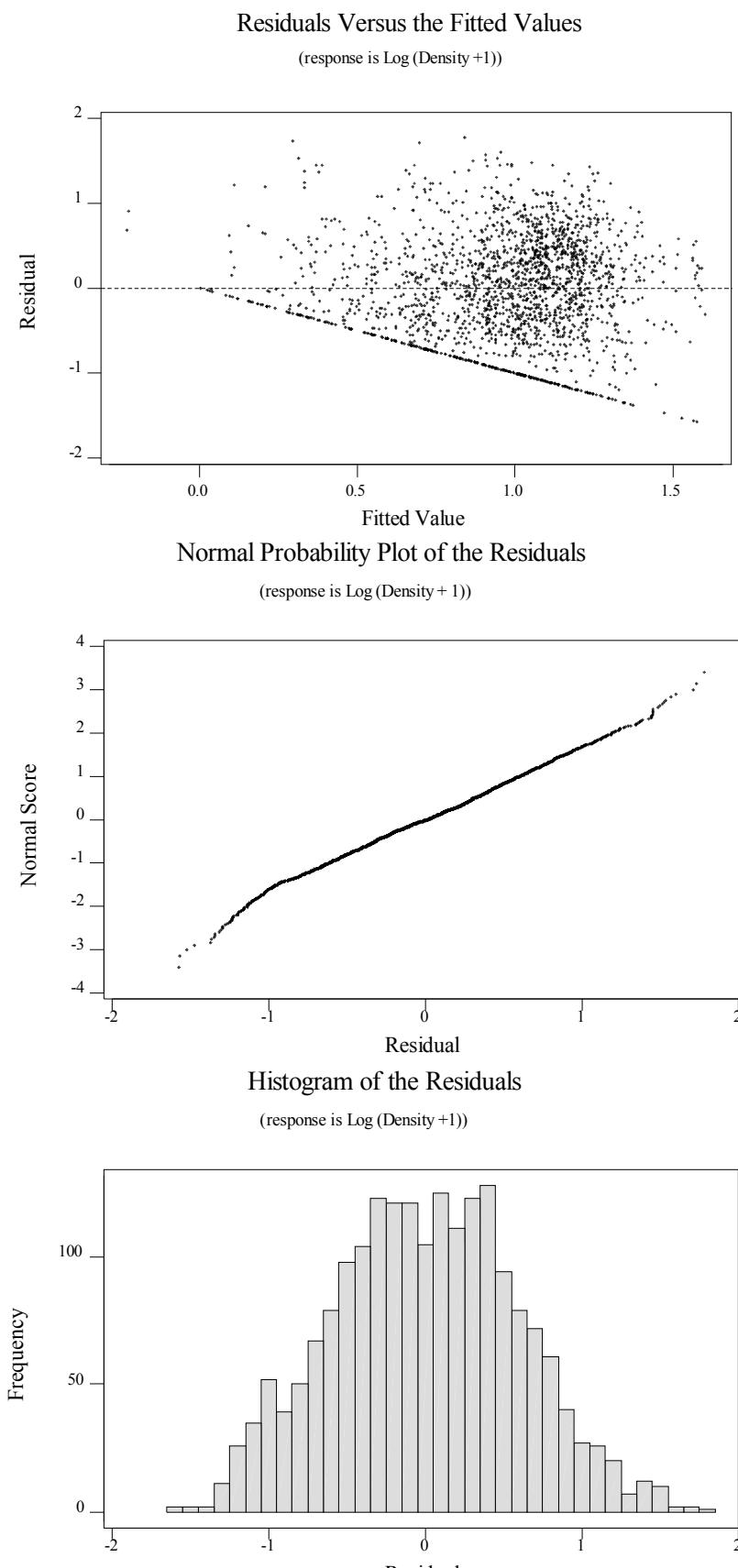
Figure 4.8   Multiple Regression Minitab residual plots for trout fry, Basic Model

Figure 4.9 describes how each individual attribute acts within the model if all the other attributes are at their average value and remain constant. Figures 4.10 to 4.12 try to describe some of the interactions between the variables within the model. All of these figures display trends on fish population rather than absolute densities as they are only picking out the influence of each part of the model. The coloured zones represent the fish density as predicted by the interaction between the two components of the models in the particular graph. The colours from red to green indicate categories of low to high densities respectively. The black dots represent the spread of the real data used to construct the models. These indicate the zone of the predicted densities that are realistic, in terms of combinations of attribute values that actually exist.
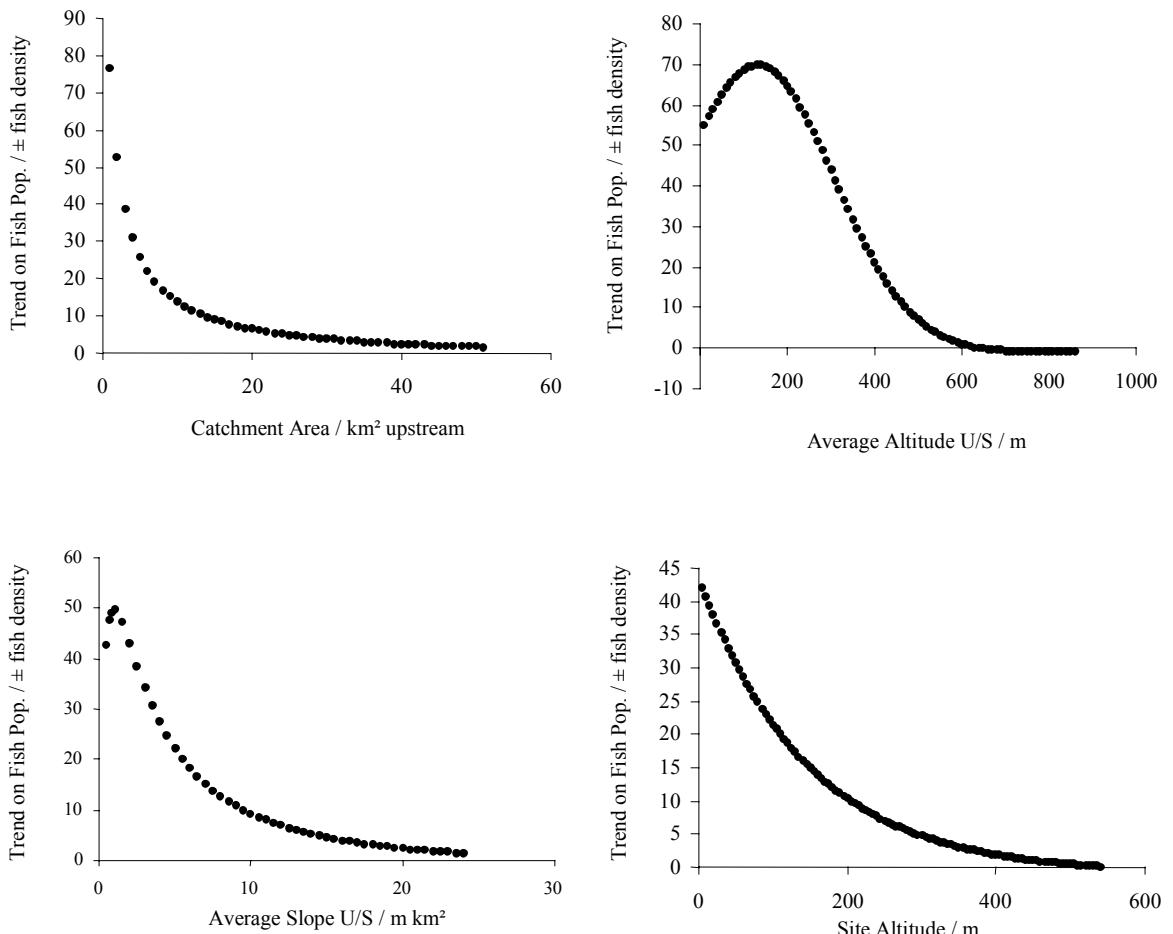

**Figure 4.9   Relationship between selected catchment features from the model and the trend on trout fry populations**
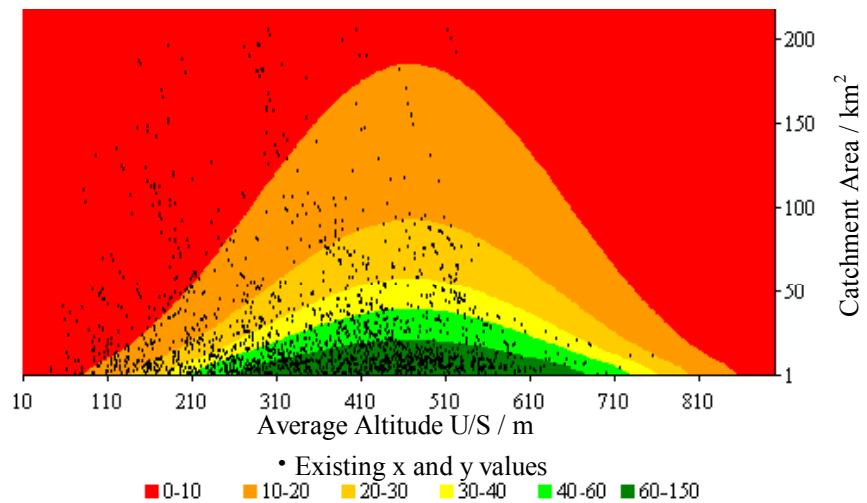
Figure 4.10   Tendency on trout fry density from Average Altitude Upstream and Catchment area within the model
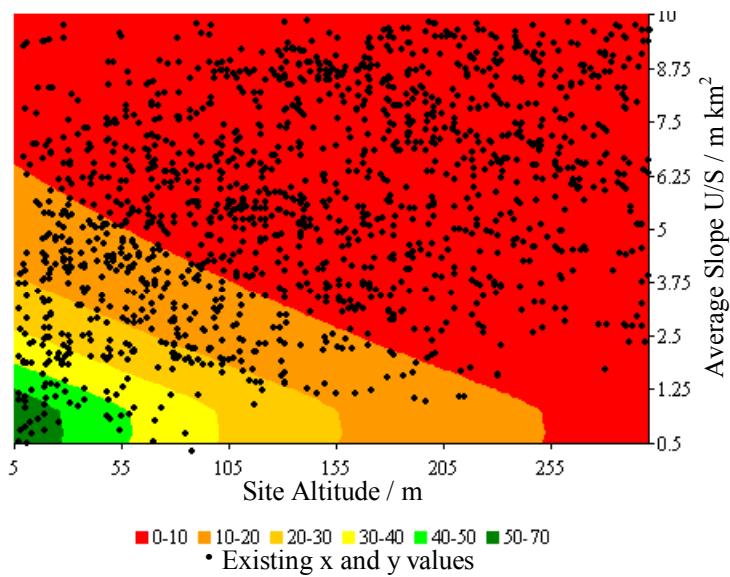

Figure 4.11   Tendency on trout fry density from Site Altitude and Average Slope Upstream within the model
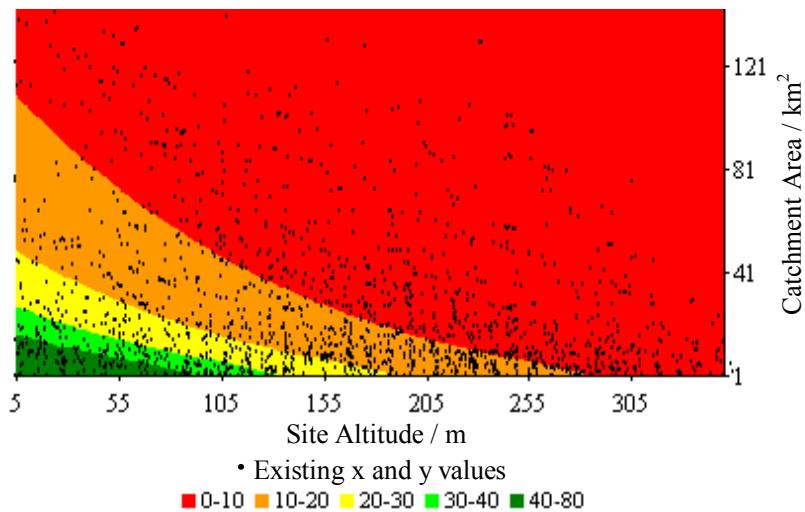

Figure 4.12   Tendency on trout fry density from Site Altitude and Catchment Area within the model

Multivariable Relationships with Landcover Variables

With the onset of GIS methods the production of new attributes has become possible. One such suite of new variables that may be of use in describing and understanding fish populations is that of land cover. A number of landcover upstream variables were tested in addition to the basic model to see if they helped in explaining the variability of the trout fry density data. There were three landcover variables that were found to be of use in describing the variation. These were:
1. Proportion Deciduous Forested Land Upstream
2. Proportion Urbanised Land Upstream
3. Proportion Arable Land Upstream

Equation 4.2 and Table 4.4 display the regression model resulting from this investigation. An extra 5.9% of the variation was explained by the addition of these landcover variables. Figure 4.13 displays the residual plot outputs from Minitab. These residual plot outputs show how well the model fits the assumptions of a regression model. The plot that shows some concern is again that of the residuals verses fitted, which suggests on initial viewing that the variance is increasing on the x-axis. However, the line created by the zero densities may exaggerate this. These may not have such a large effect on the model itself. The Normal probability plot and the histogram of the residuals both suggest an improvement over the basic model without the landcover variables.

(Equation 4.2): $\delta = 1.89 - 0.478\ A - 0.234\ A^2 - 0.00348\ B + 0.00137\ C - 0.000004\ C^2 - 0.352\ D - 0.872\ D^2 - 1.63\ E - 3.10\ F - 0.721\ G$

Where:

δ = Log (Density +1) Trout Fry

A = Log_{10} Catchment Area - $\bar{x}$ ($\bar{x}$ = 1.183)

B = Site Altitude

C = Average Altitude U/S - $\bar{x}$ ($\bar{x}$ = 300.8)

D = Log_{10} Average Slope U/S - $\bar{x}$ ($\bar{x}$ = 0.284)

E = Proportion Deciduous forested land U/S

F = Proportion Urbanised land U/S

G = Proportion Arable land U/S

Table 4.8   Multiple Regression Minitab output for trout fry, Full Model

Predictor	Coef	SE Coef	T	P
Constant	1.89318	0.06651	28.47	0.000
A	-0.47827	0.02761	-17.32	0.000
A^2	-0.23400	0.02960	-7.91	0.000
B	-0.0034775	0.0003562	-9.76	0.000
C	0.0013667	0.0003407	4.01	0.000
C^2	-0.00000371	0.00000093	-4.00	0.000
D	-0.3519	0.1005	-3.50	0.000
D^2	-0.8724	0.1849	-4.72	0.000
E	-1.6308	0.2825	-5.77	0.000
F	-3.1008	0.3054	-10.15	0.000
G	-0.7206	0.1474	-4.89	0.000

S = 0.5755   R-Sq = 23.6%   R-Sq(adj) = 23.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	201.541	20.154	60.85	0.000
Residual Error	1966	651.194	0.331		
Lack of Fit	1922	646.340	0.336	3.05	0.000
Pure Error	44	4.854	0.110		
Total	1976	852.736			

Residuals Versus the Fitted Values
(response is Log (Density +1))
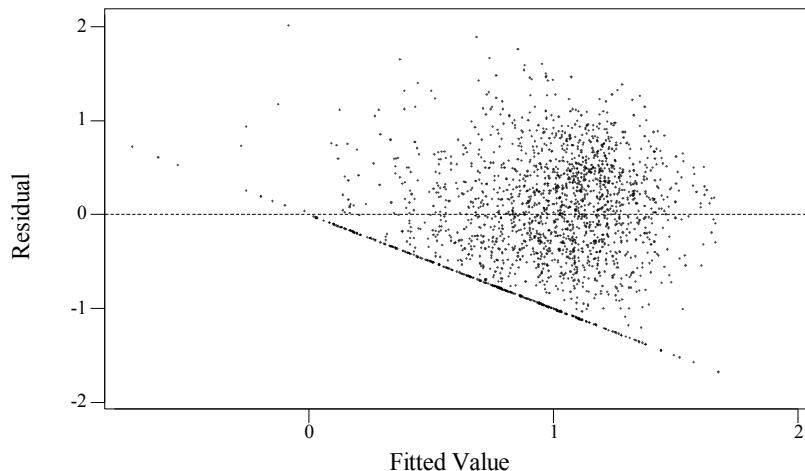
Normal Probability Plot of the Residuals
(response is Log (Density +1))
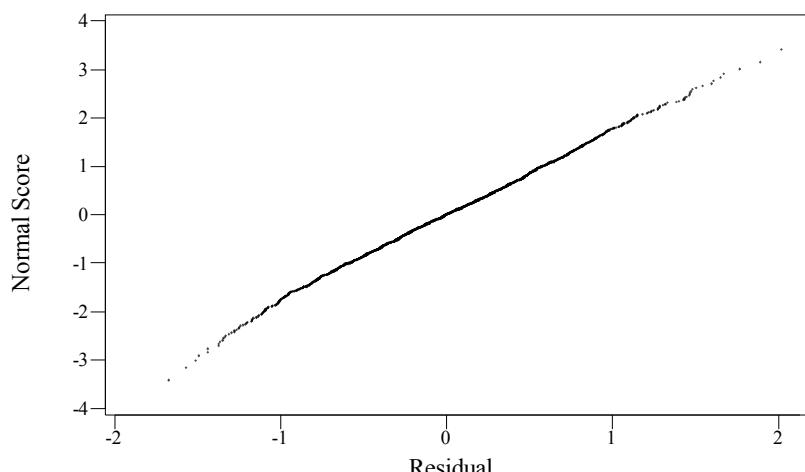
Histogram of the Residuals
(response is Log (Density +1))
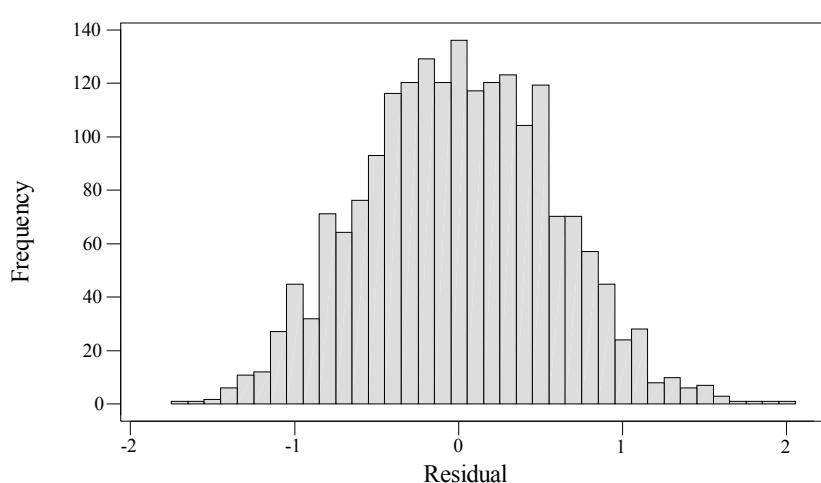
Figure 4.13   Multiple Regression Minitab residual plots for trout fry, Full Model

4.3.4 Salmon Fry Multivariable Relationships

The investigation into the relationship between the GIS derived variables and salmon fry densities was performed in a similar way to that of the trout fry. However, as access is an issue with salmon, zero densities were removed as a crude way of taking out those sites where access was not possible for salmon. Log density transformation was now possible as the problem of zero densities was removed.

Again the initial idea was to come up with a model based on the sort of variables that may be used in current relationship models such as HABSCORE. Table 4.5 displays the Minitab output of the best subsets regression performed. Primarily it suggested a number of variables for further analysis these were:
1. Site Altitude,
2. Average Altitude Upstream,
3. Average Slope Upstream,
4. Distance to Mouth and
5. Position $= \dfrac{\text{Distance to Source}}{(\text{Distance to Source} + \text{Distance to Mouth})}$

Table 4.9   Best Subsets Regression Minitab output for salmon fry

```
Response is Log Density

Vars    R-Sq    R-Sq(adj)        C-p        S     A B C D E F G

  3     11.4      11.2           9.3     0.63612     X     X X
  3     11.3      11.0          11.1     0.63669   X X     X
  3     10.6      10.3          18.7     0.63905     X   X X
  4     11.8      11.4           7.0     0.63506   X X X   X
  4     11.7      11.4           7.8     0.63533   X X     X X
  4     11.5      11.1          10.6     0.63620     X     X X
  5     12.1      11.6           5.7     0.63435   X X X   X X
  5     11.9      11.4           8.1     0.63511   X X X   X
  5     11.8      11.4           8.5     0.63524   X X X X X
  6     12.3      11.7           5.6     0.63401   X X X X X X
  6     12.2      11.6           6.7     0.63435   X X X   X X X
  6     12.2      11.6           6.7     0.63436   X X X   X X
```

```
A = Site Altitude
B = Average Altitude U/S
C = Average Slope U/S
D = Distance to Source
E = Distance to Mouth
F = Position
G = Catchment Area
```

Further analysis with multiple regressions of the selected site and catchment variables were undertaken to further refine the models and add or remove variables when appropriate. A number of transformations of the data were also examined such as log10 and X^2 terms.

Equation 4.3 displays the final model produced. This is termed the salmon fry Basic Model as no new attributes were included. Table 4.6 displays the Minitab output of this regression and Figure 4.14 displays the residual plot outputs from Minitab. These residual plot outputs show how well the model fits the assumptions of a regression model. The plots suggest a satisfactory model in terms of the assumptions of regression analysis.

(Equation 4.3): δ = 1.12 + 0.000609 A - 0.000012 A^2 + 0.00135 B -0.000008 B^2 + 0.334 C + 0.354 C^2 + 0.582 D

Where:

δ = Log Density Salmon Fry

A = Site Altitude - $\bar{x}$ ($\bar{x}$ = 157.5)

B = Average Altitude U/S - $\bar{x}$ ($\bar{x}$ = 300.8)

C = Log$_{10}$ Distance to Mouth - $\bar{x}$ ($\bar{x}$ = 4.73)

D = Position = Distance to Source / (Distance to Source + Distance to Mouth)

Table 4.10　　Multiple Regression Minitab output for salmon fry, Basic Model

Predictor	Coef	SE Coef	T	P	VIF
Constant	1.11932	0.05017	22.31	0.000	
A	0.0006090	0.0005632	1.08	0.280	6.0
A^2	-0.00001192	0.00000363	3.28	0.001	2.1
B	0.0013517	0.0003211	4.21	0.000	2.8
B^2	-0.00000770	0.00000187	-4.13	0.000	1.4
C	0.33408	0.07992	4.18	0.000	4.6
C^2	0.35410	0.06374	5.56	0.000	1.7
D	0.5821	0.2309	2.52	0.012	5.8

S = 0.6211　　　　R-Sq = 15.9%　　　R-Sq(adj) = 15.3%
PRESS = 392.847　　R-Sq(pred) = 14.51%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	73.008	10.430	27.04	0.000
Residual Error	1002	386.533	0.386		
Total	1009	459.540			

Residuals Versus the Fitted Values
(response is Log Density)

Normal Probability Plot of the Residuals
(response is Log Density)

Histogram of the Residuals
### (response is Log Density)
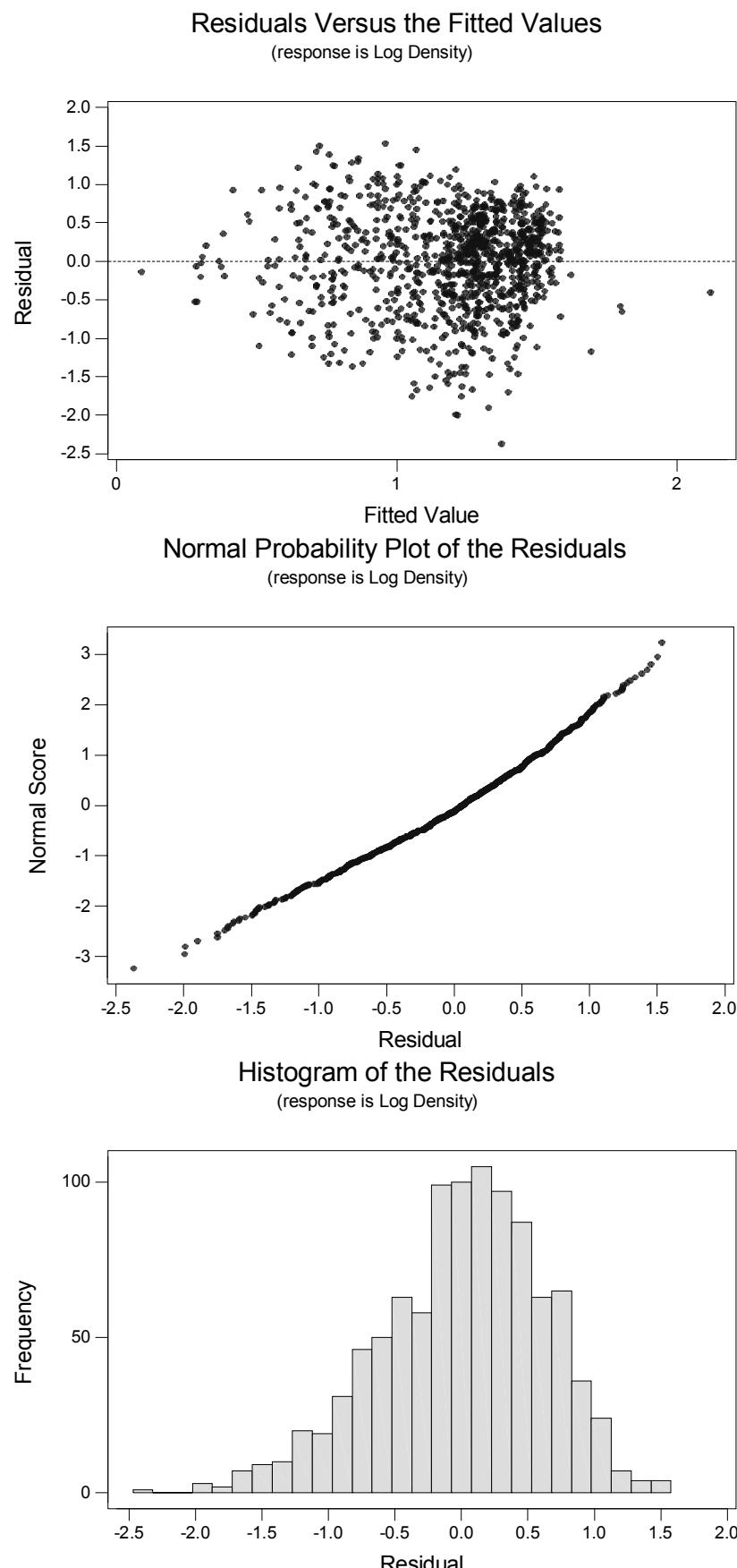
Figure 4.14   Multiple Regression Minitab residual plots for salmon fry, Basic Model

Figure 4.15 describes how each individual attribute acts within the model if all the other attributes are at their average value and remain constant. Figures 4.16 and 4.17 try to describe some of the interactions between the variables within the model. All of these figures display trends on fish population rather than absolute densities as they are only picking out the influence of each part of the model.
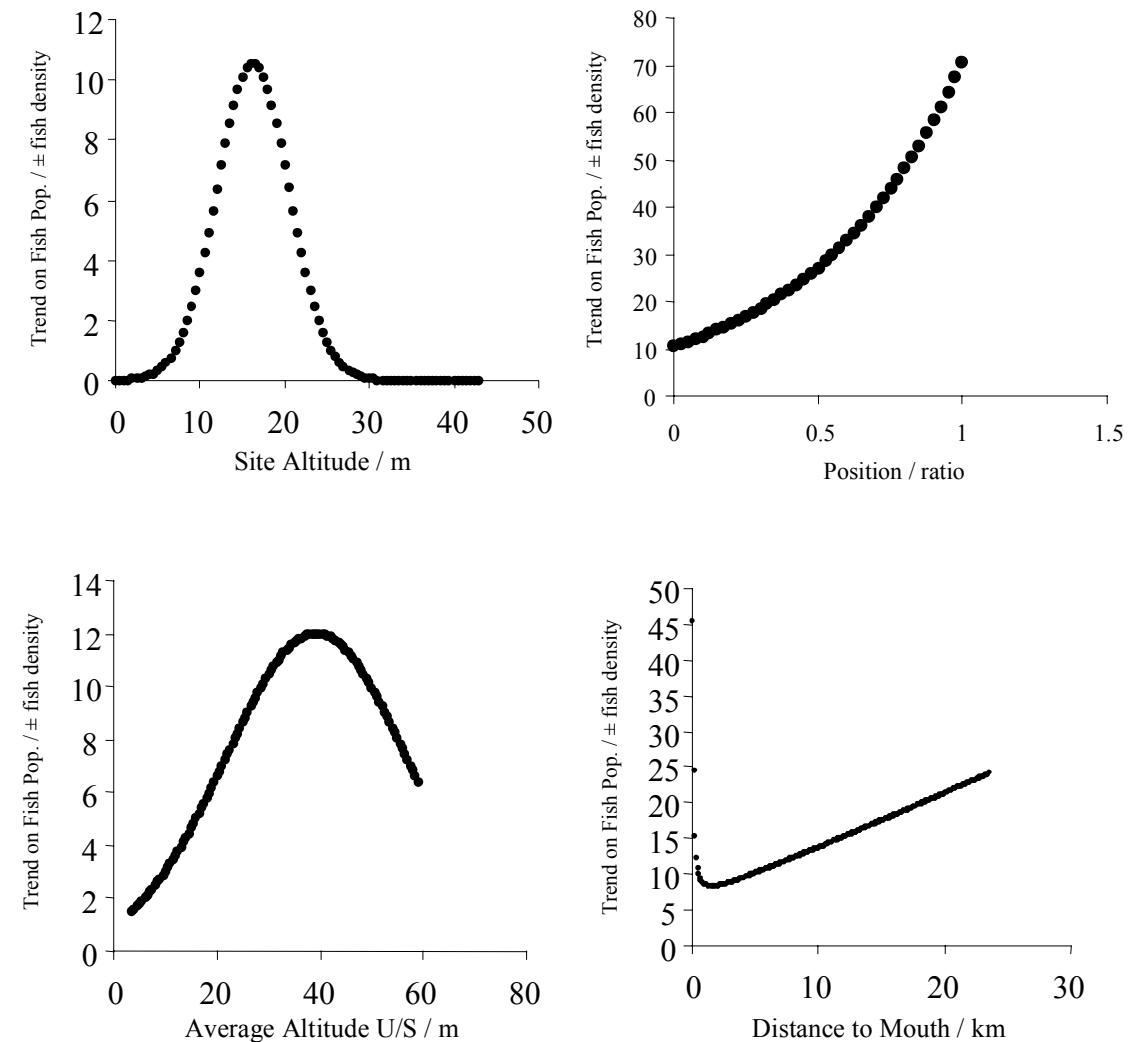
Figure 4.15   Relationship between selected catchment features from the model and the trend on salmon fry populations
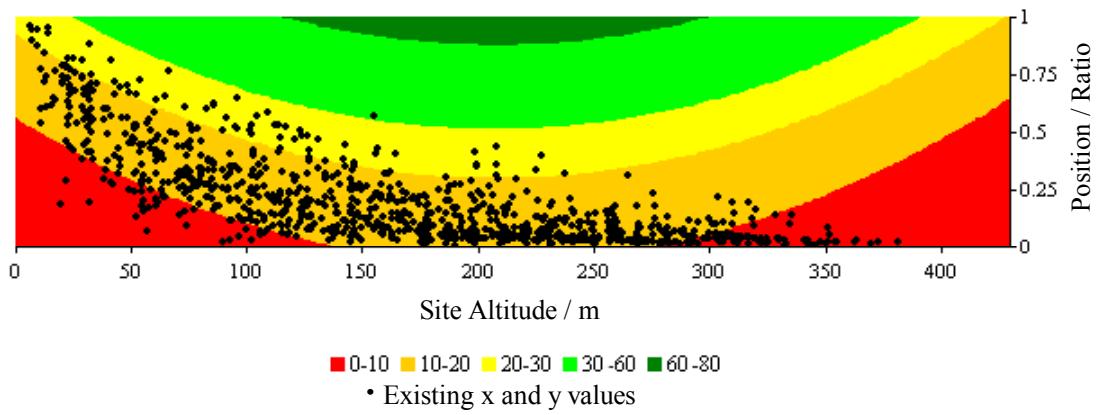
Figure 4.16   Tendency on salmon fry density from Site Altitude and Position within the model
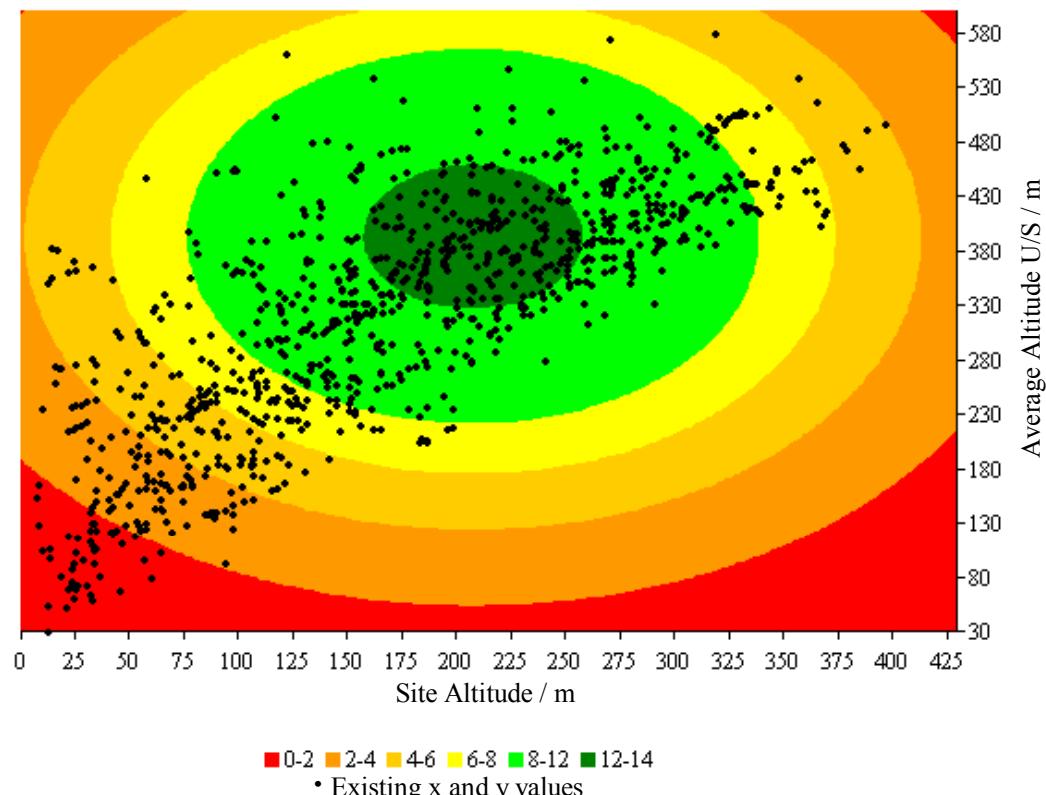
Figure 4.17   Tendency on salmon fry density from Site Altitude and Average Altitude Upstream within the model

Multivariable Relationships with Landcover Variables

As with trout fry a number of landcover upstream variables were tested in addition to the basic model to see if they helped in explaining the variability of the salmon fry density data. There were two landcover variables that were found to be of use in describing the variation. These were:
1. Proportion Deciduous Forested Land Upstream and
2. Proportion Urbanised Land Upstream

Equation 4.4 and Table 4.7 display the regression model resulting from this investigation. An extra 2.1% of the variation was explained by the addition of these landcover variables. Figure 4.18 displays the residual plot outputs from Minitab. The plots suggest a satisfactory model in terms of the assumptions of regressions. The Normal probability plot and the histogram of the residuals both suggest an improvement over the basic model without the landcover variables.

(Equation 4.4): δ = 1.02 + 0.00121 A - 0.000012 A^2 + 0.00128 B - 0.000007 B^2 + 0.356 C + 0.330 C^2 + 0.829 D - 3.92 E + 1.73 F

Where:
δ = Log Density Salmon Fry
A = Site Altitude - $\bar{x}$ ($\bar{x}$ = 157.5)
B = Average Altitude U/S - $\bar{x}$ ($\bar{x}$ = 300.8)
C = Log_{10} Distance to Mouth - $\bar{x}$ ($\bar{x}$ = 4.73)
D = Position = Distance to Source / (Distance to Source + Distance to Mouth)
E = Proportion Urbanised land U/S
F = Proportion Deciduous forested land U/S

Table 4.11   Multiple Regression Minitab output for salmon fry, Full Model

Predictor	Coef	SE Coef	T	P	VIF
Constant	1.01874	0.06106	16.68	0.000	
A	0.0012058	0.0005779	2.09	0.037	6.5
A^2	-0.00001184	0.00000361	-3.28	0.001	2.1
B	0.0012789	0.0003262	3.92	0.000	3.0
B^2	-0.00000715	0.00000185	-3.87	0.000	1.4
C	0.35588	0.07928	4.49	0.000	4.6
C^2	0.33032	0.06347	5.20	0.000	1.8
D	0.8292	0.2333	3.55	0.000	6.1
E	-3.915	1.035	-3.78	0.000	1.3
F	1.7341	0.4964	3.49	0.000	1.6

S = 0.6138   R-Sq = 18.0%   R-Sq(adj) = 17.3%
PRESS = 384.587   R-Sq(pred) = 16.31%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	82.7474	9.1942	24.40	0.000
Residual Error	1000	376.7928	0.3768		
Total	1009	459.5402			

Residuals Versus the Fitted Values
(response is Log Density)

Normal Probability Plot of the Residuals
(response is Log Density)

Histogram of the Residuals
### (response is Log Density)
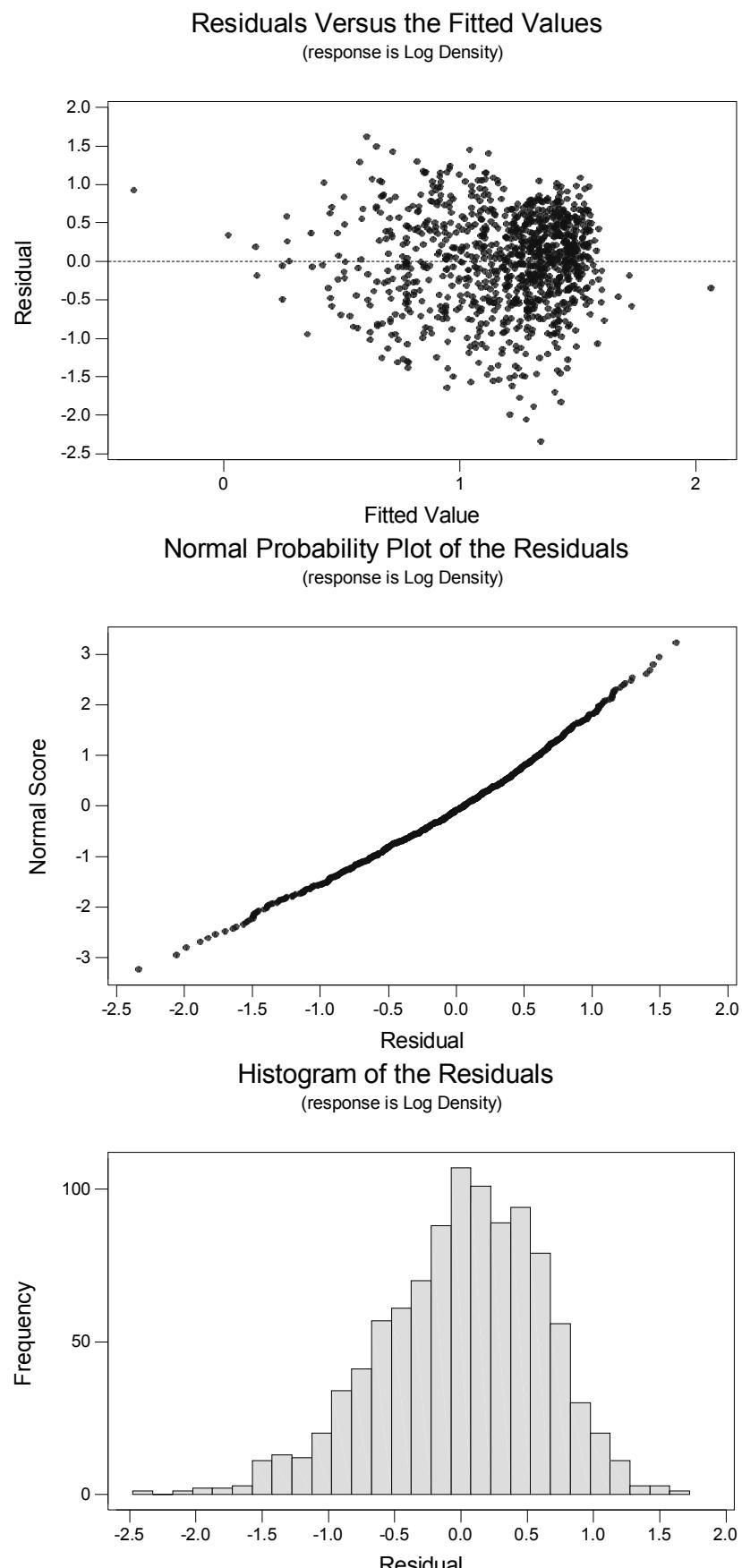
Figure 4.18   Multiple Regression Minitab residual plots for salmon fry, Full Model

4.3.5 Application of Models Within the GIS

The equations produced from the regression models described in previous sections can be applied within ArcView's Spatial Analyst to display predicted fish densities using the variables within the model.

For both salmon fry and trout fry the two models were executed within the Spatial Analyst raster calculator to produce a Basic and a Full model. By taking the basic model away from the full model a grid of the effect of landcover within the model was also produced.

Salmon Fry

To execute the basic model within ArcView Equation 4.3 was transformed into the MapAlgebra format as Equation 4.5. The full model was transformed from Equation 4.4 to produce Equation 4.6. The resultant grids from Equations 4.5 and 4.6 were then used to determine the effect of landcover on salmon fry within the model as Equation 4.7.

(Equation 4.5): Salmon Fry - Basic (SF_B)

```
1.12 + 0.000609 * ([dem_50m] - 157.5) - 0.000012 * (pow([dem_50m] -
157.5, 2)) + 0.00135 * ([avupalt100] - 300.8) - 0.000008 *
(pow([avupalt100] - 300.8, 2)) + 0.334 * (Log10([distmouth] - 4.73)) +
0.354 * (pow(Log10([distmouth] - 4.73)), 2) + 0.582 * [posi]
```

    Note: Spaces before and after each operator.
    `[dem_50m]` = Altitude Grid
    `[avupalt100]` = Average Altitude U/S Grid
    `[distmouth]` = Distance to Mouth Grid
    `[posi]` = Position Grid (Distance to Source / (Distance to Source + Distance to Mouth))

(Equation 4.6): Salmon Fry - Full (SF_F)

```
1.02 + 0.00121 * ([dem_50m] - 157.5) - 0.000012 * (pow([dem_50m] -
157.5, 2)) + 0.00128 * ([avupalt100] - 300.8) - 0.000007 *
(pow([avupalt100] - 300.8, 2)) + 0.356 * (Log10([distmouth] - 4.73)) +
0.330 * (pow(Log10([distmouth] - 4.73)), 2) + 0.829 * [posi] - 3.92 *
[p_urbanf100] + 1.73 * [p_decid100]
```

    Note: Spaces before and after each operator.
    `[dem_50m]` = Altitude Grid
    `[avupalt100]` = Average Altitude U/S Grid
    `[distmouth]` = Distance to Mouth Grid
    `[posi]` = Position Grid (Distance to Source / (Distance to Source + Distance to Mouth))
    `[p_decid100]` = Proportion Deciduous Forest U/S Grid
    `[p_urbanf100]` = Proportion Urbanised Land U/S Grid

(Equation 4.7): Effect of Landcover on Salmon Fry

```
[SF_F] - [SF_B]
```

    Note: Spaces before and after each operator.
    `[SF_F]` = Full Salmon Fry Model
   `[SF_B]` = Basic Salmon Fry Model

The models were executed within ArcView to produce three grids for salmon fry.
1.   Basic Model
2.   Full Model
3.   Landcover Effect Model

Figures 4.20 to 4.28 display three example areas of the full model and Landcover effect model grids. Figure 4.19 displays the legend for the example areas below. The three example areas are:
1.   The River Aln and Coquet,
2.   The Middle River Severn,
3.   For the River Dart.

The first Figure for each river examples displays the landcover of the types within the model for the same areas.
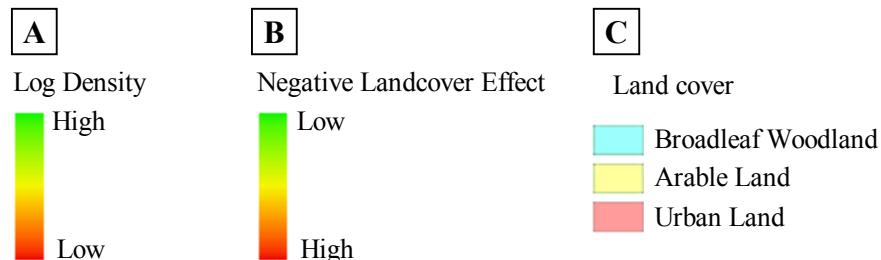
A   B   C

Log Density   Negative Landcover Effect   Land cover

High   Low   Broadleaf Woodland

      Arable Land

Low   High   Urban Land

A = Predicted Fish Density (Log Density)
B = Predicted Effect of Landcover (Model with Landcover – Model without Land cover)
C = Selected Land cover

Figure 4.19   Legends for following Figures
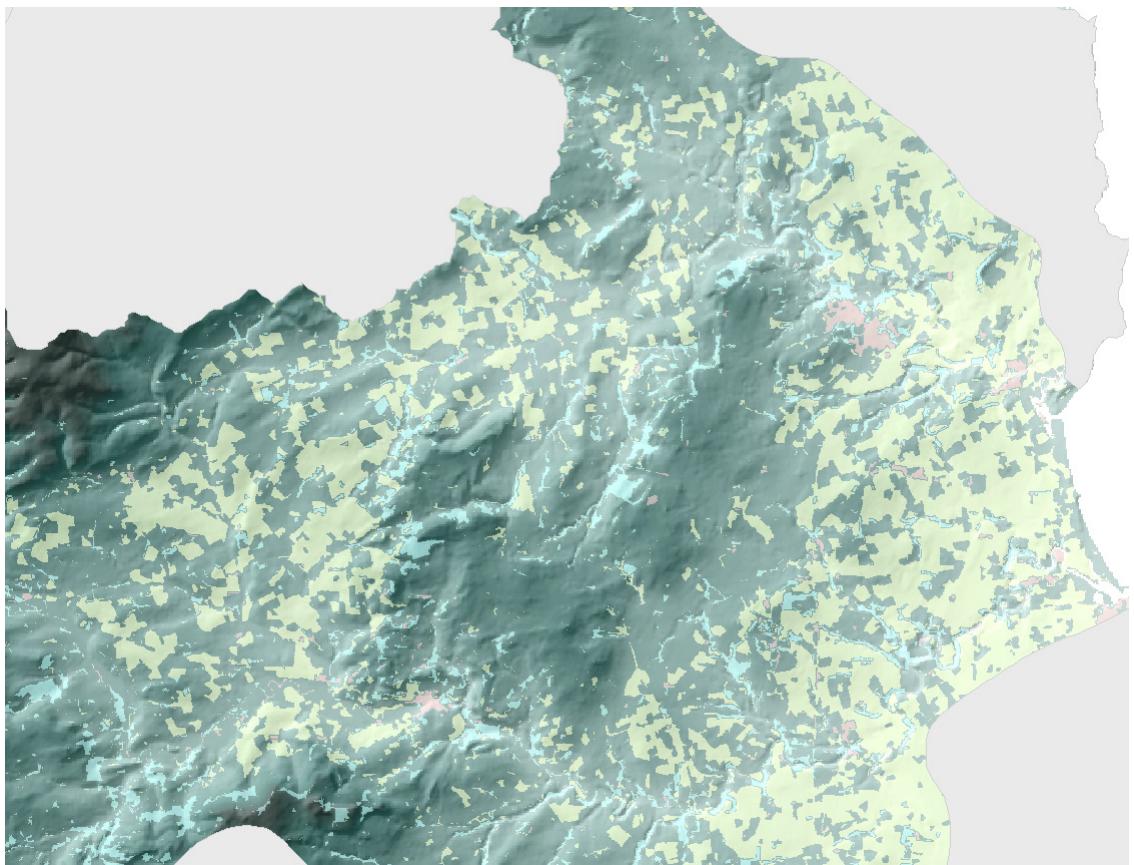
Figure 4.20   Selected Landcover for the River Aln and Coquet example
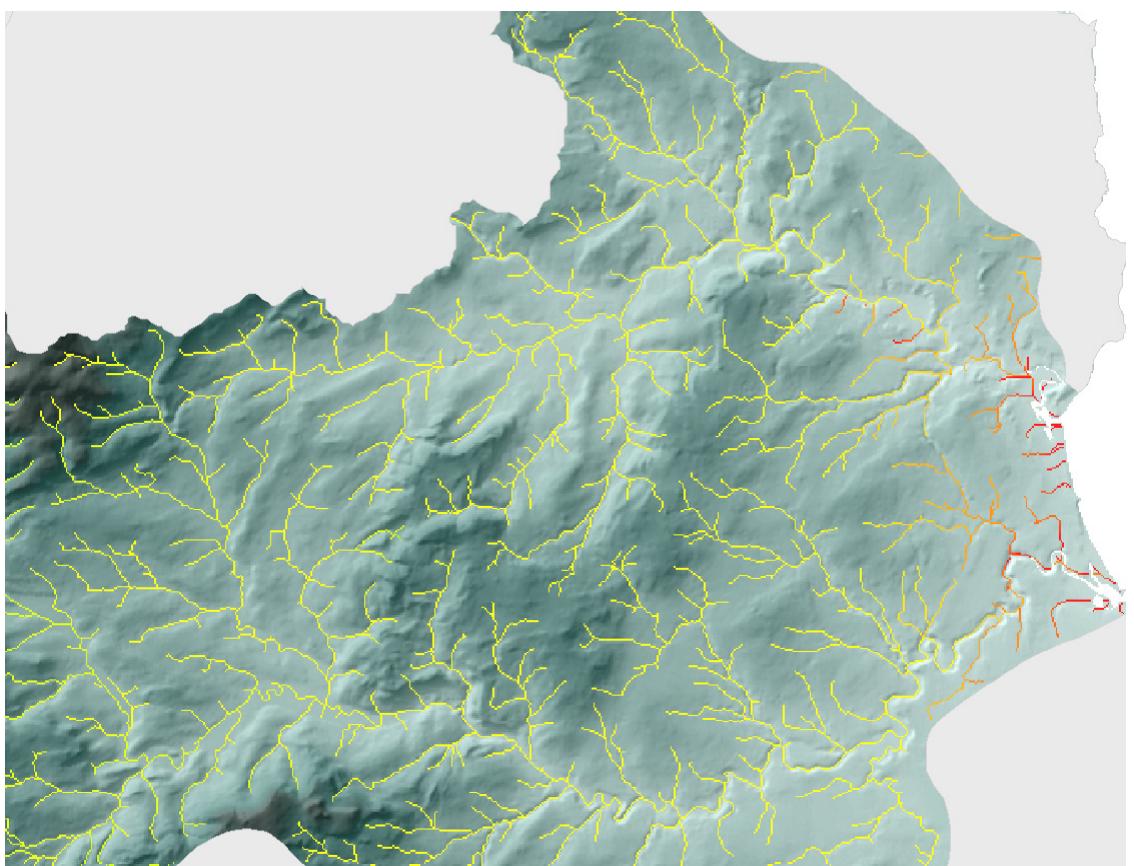
**Figure 4.21   Full Model for salmon fry the River Aln and Coquet example**
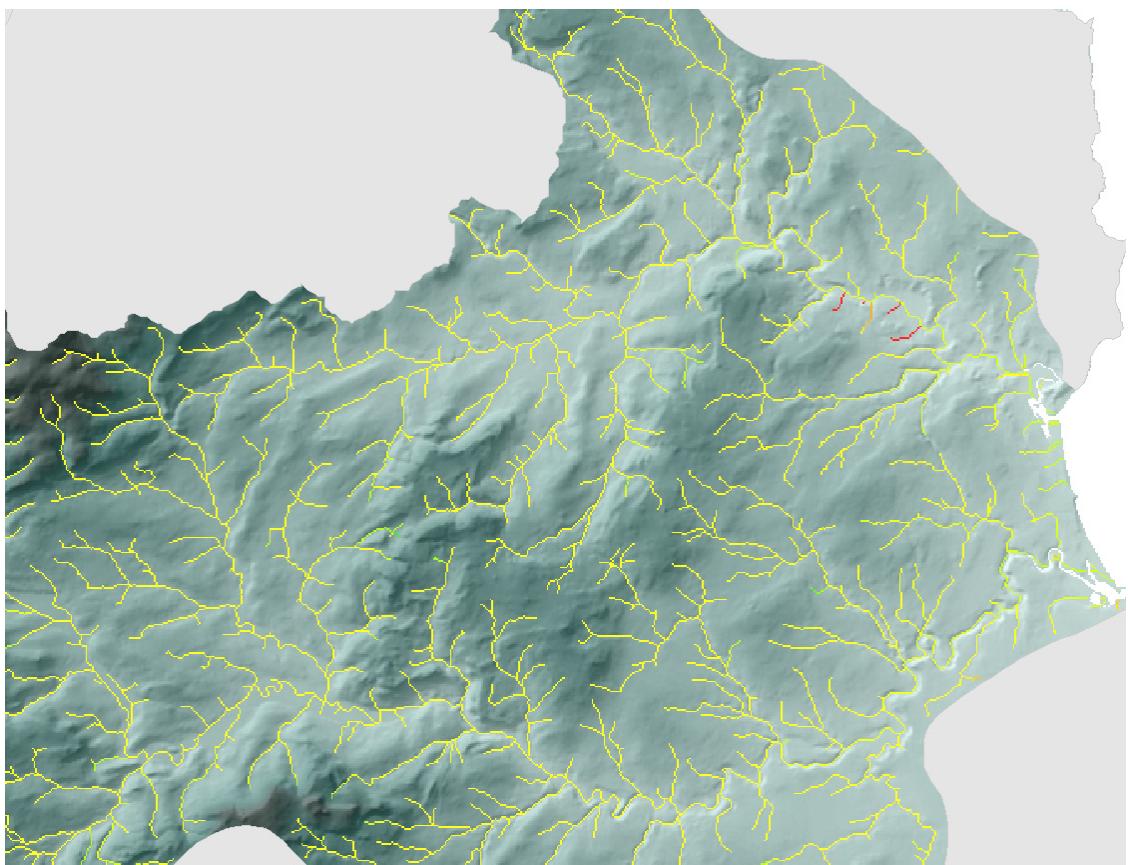
**Figure 4.22   Landcover effect for salmon fry the River Aln and Coquet example**
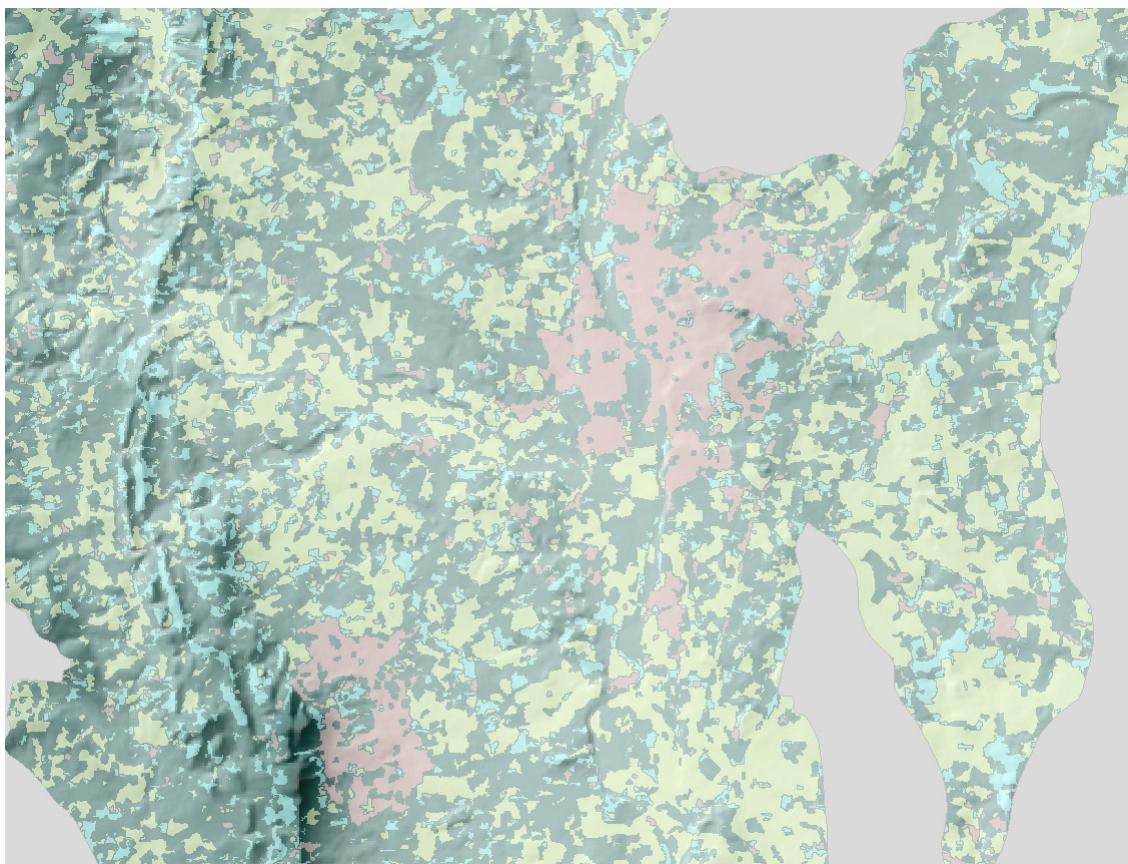
**Figure 4.23   Selected Landcover for the Middle River Severn example**
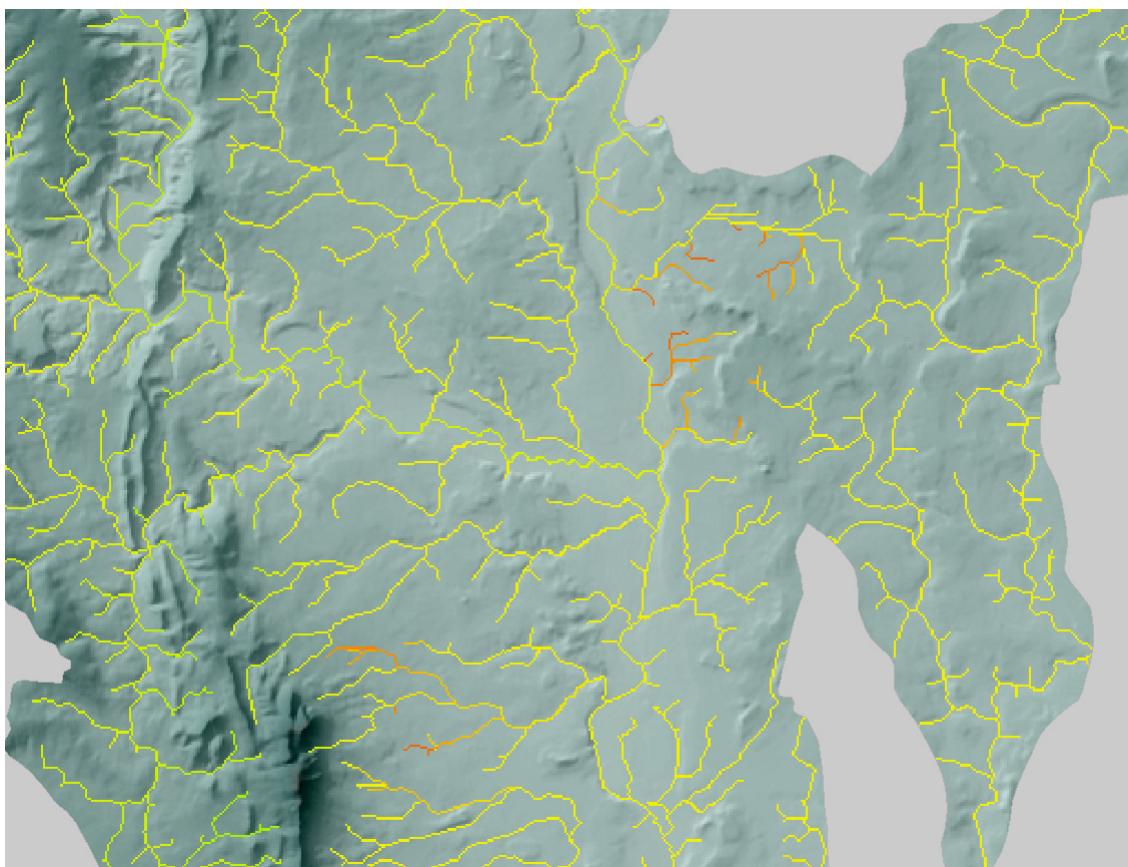
**Figure 4.24   Full Model for salmon fry the Middle River Severn example**
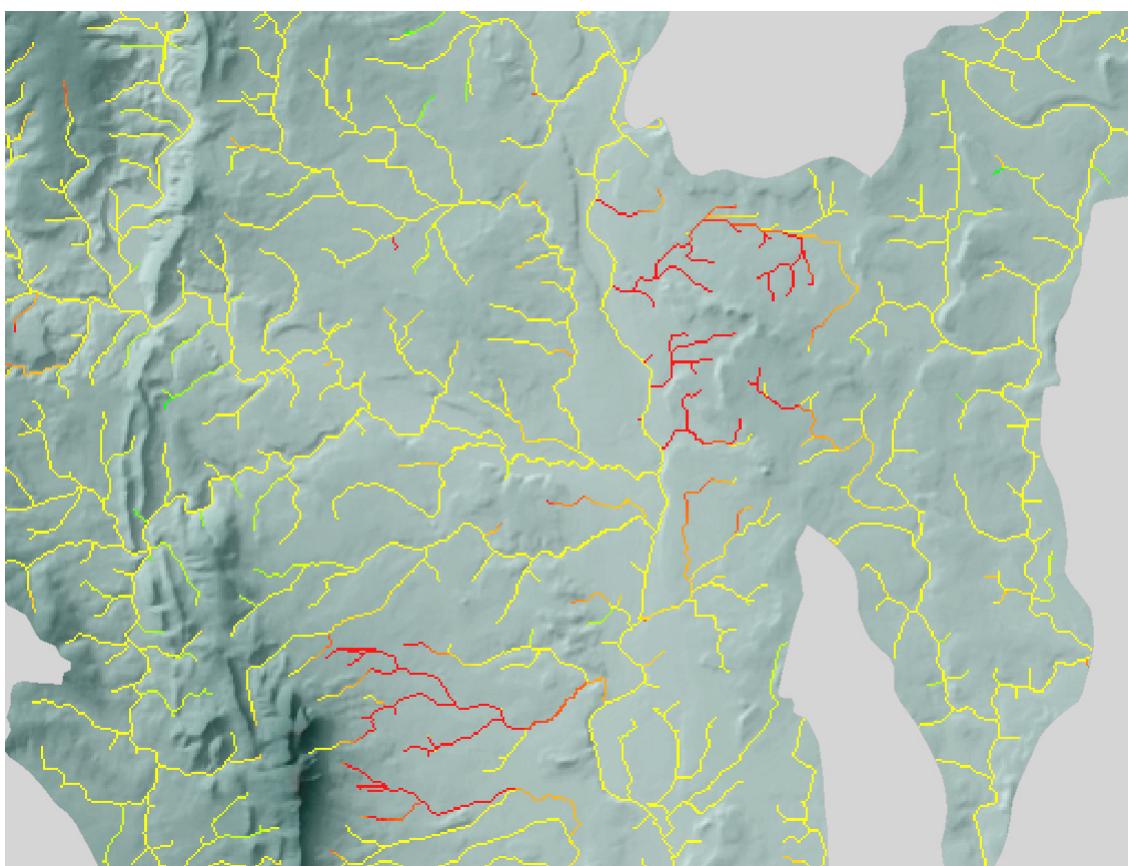
**Figure 4.25   Landcover effect for salmon fry the Middle River Severn example**
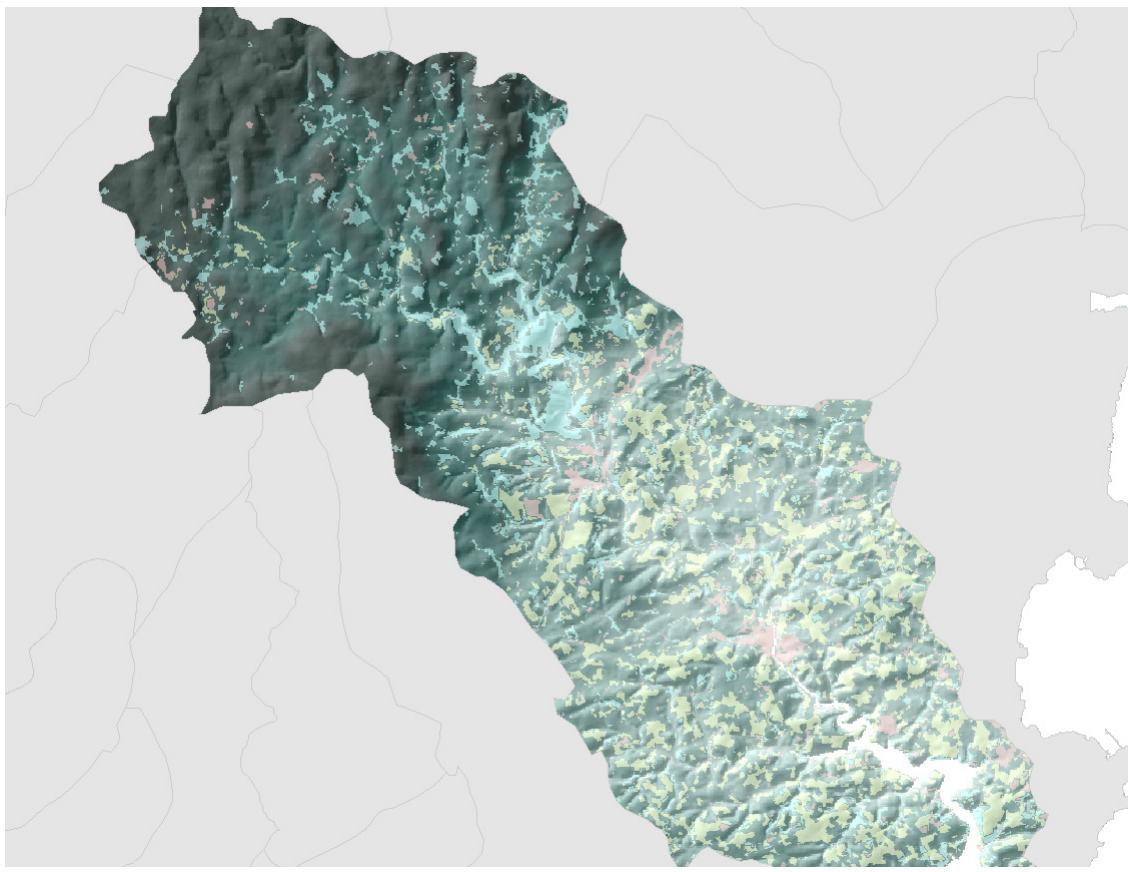
Figure 4.26   Selected Landcover for the River Dart example

Figure 4.27   Full Model for salmon fry the River Dart example

Figure 4.28   Landcover effect for salmon fry the River Dart example

The River Aln and Coquet examples show a catchment coloured as expected with a gradation from poor habitat at the bottom of the catchment (in terms of the model) to good habitat higher up the catchment. The Landcover effect model (Figure 4.22) suggests that there are only a few small tributaries with high negative landcover effects, these correspond to tributaries that source in predominantly urbanised land.

The Middle River Severn examples show a catchment coloured as the majority of the habitat (in terms of the model) being good. There are two sections that shows poor habitat quality (reduced predicted salmon fry densities). These sections are highlighted even more strongly by the landcover effect model (Figure 4.25).   Again these correspond to tributaries that source in predominantly urbanised land (towns).

The River Dart example is another typical catchment coloured with a gradation from poor habitat in the coastal areas (in terms of the model) to good habitat higher up the catchment. The Landcover effect model (Figure 4.28) suggests that there are only a few small tributaries with high negative landcover effects, these again correspond to tributaries that source in predominantly urbanised land.

Figure 4.29 depicts the proportions of the river network that fall into 6 categories of predicted log salmon fry density. The y-axis corresponds to the number of 50m2 cells that fall in each category. This equates to kilometres of river network if multiplied by 50 and divided by 1000 (as 50m cells). This chart shows that of the modelled areas in England and Wales the majority of cells fall in the central region of predicted values shifted slightly towards the higher end of the scale.

Figure 4.30 depicts the proportions of the river network that fall into 6 categories of predicted log salmon fry density relating from the landcover effect model. This chart shows that of the modelled areas in England and Wales the majority of cells fall in the central region of predicted values shifted slightly towards the higher end of the scale (medium to low landcover effect).
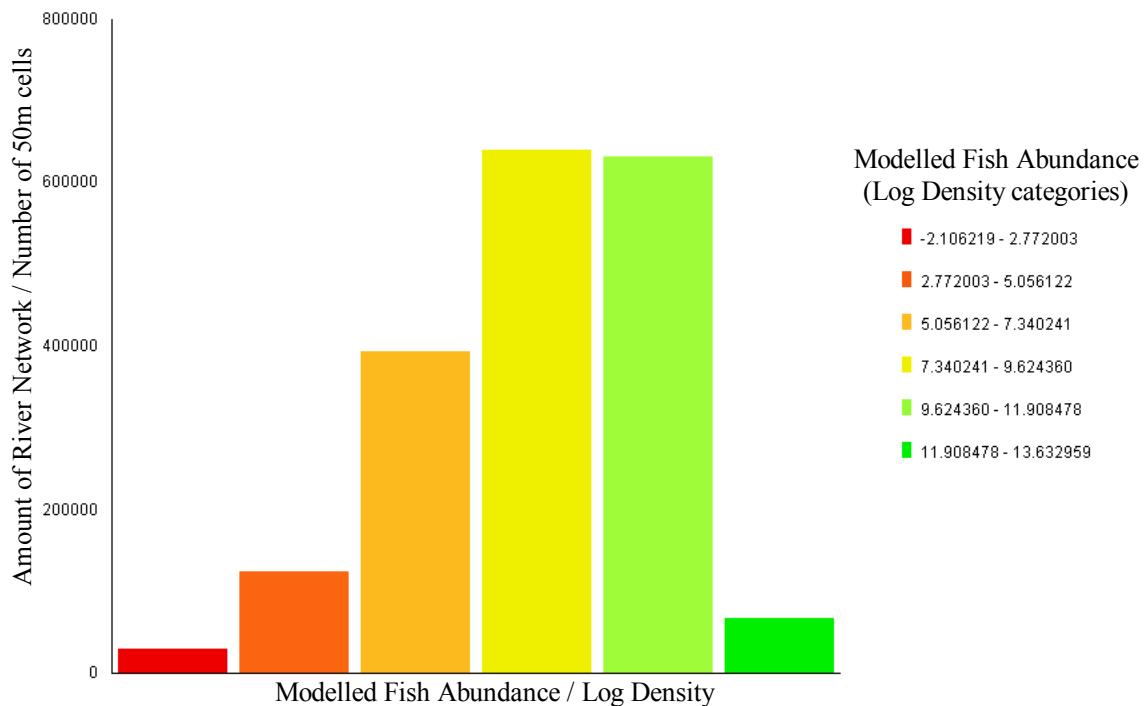

Figure 4.29   Chart depicting the proportion of 50m² cells in each category (Log Density salmon fry)
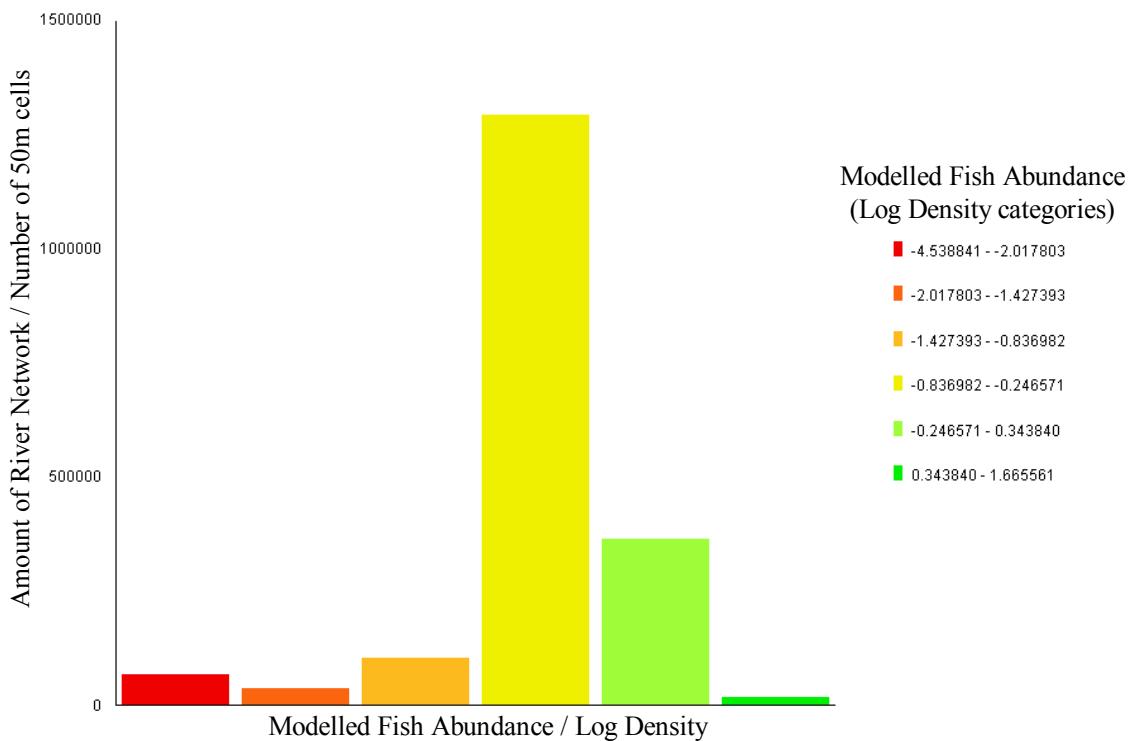
Figure 4.30 **Chart depicting the proportion of 50m² cells in each category of landcover effect (Log Density salmon fry) where red = high landcover effect**

Trout Fry

As with the salmon fry model the basic model within ArcView Equation 4.1 was transformed into the MapAlgebra format as Equation 4.8. The full model was transformed from Equation 4.2 to produce Equation 4.9. The resultant grids from Equations 4.8 and 4.9 were then used to determine the effect of landcover on the trout fry within the model as Equation 4.10.

(Equation 4.8): Trout Fry basic (TF_B)
```
1.65 - 0.486 * (Log10([catch] - 1.183) - 0.239 * (pow(Log10([catch] -
1.183)), 2)  - 0.00297 * [dem_50m] + 0.00236 * ([avupalt100] - 300.8)
- 0.000007 *  (pow([avupalt100] - 300.8)), 2) - 0.394 *
(Log10([avslopeus] – 0.284) - 0.696 * (pow(Log10([avslopeus] – 0.284),
2)
```

    Note: Spaces before and after each operator.
    `[dem_50m]` = Altitude Grid
    `[catch]` = Catchment Area Grid
    `[avupalt100]` = Average Altitude U/S Grid
    `[avslopeus]` = Average Slope U/S Grid

(Equation 4.9): Trout Fry full (TF_F)

```
1.89 - 0.478 * (Log10([catch] - 1.183) - 0.234 * (pow(Log10([catch] -
1.183)), 2) - 0.00348 * [dem_50m] + 0.00137 * ([avupalt100] - 300.8) -
0.000004 * (pow([avupalt100] - 300.8)), 2) - 0.352 *
(Log10([avslopeus] - 0.284) - 0.872 * pow(Log10([avslopeus] - 0.284),
2) - 1.63 * [p_decid100] - 3.1 * [p_urbanf100] - 0.721 * [p_arable100]
```

Note: Spaces before and after each operator.

`[dem_50m]` = Altitude Grid

`[catch]` = Catchment Area Grid

`[avupalt100]` = Average Altitude U/S Grid

`[avslopeus]` = Average Slope U/S Grid

`[p_decid100]` = Proportion Deciduous Forest U/S Grid

`[p_urbanf100]` = Proportion Urbanised Land U/S Grid

`[p_arable100]` = Proportion Arable Land U/S Grid

(Equation 4.10): Effect of landcover on Trout Fry

```
[TF_F] - [TF_B]
```

Note: Spaces before and after each operator.

`[TF_F]` = Full Trout Fry Model

`[TF_B]` = Basic Trout Fry Model

The models were executed within ArcView to produce three grids for trout fry.
1.   Basic Model
2.   Full Model
3.   Landcover Effect Model

Figures 4.32 to 4.37 display three example areas of the full model and Landcover effect model grids. These are the same rivers as used for the salmon fry examples. Figure 4.31 displays the legend for the example areas below. The three example areas are:
1.   The River Aln and Coquet,
2.   The Middle River Severn,
3.   For the River Dart.

The first Figure for each river examples of the salmon fry models (Figures 4.20 to 4.28) displays the landcover of the types within the model for the same areas.
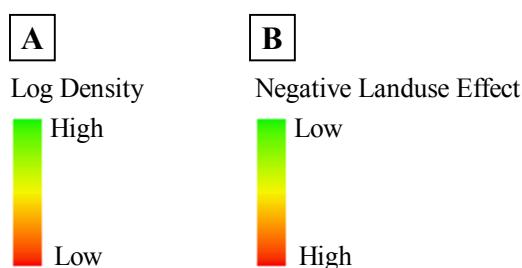
A = Predicted Fish Density (Log Density)
B = Predicted Effect of Landcover (Model with Landcover – Model without Land cover)

**Figure 4.31   Legends for following Figures**
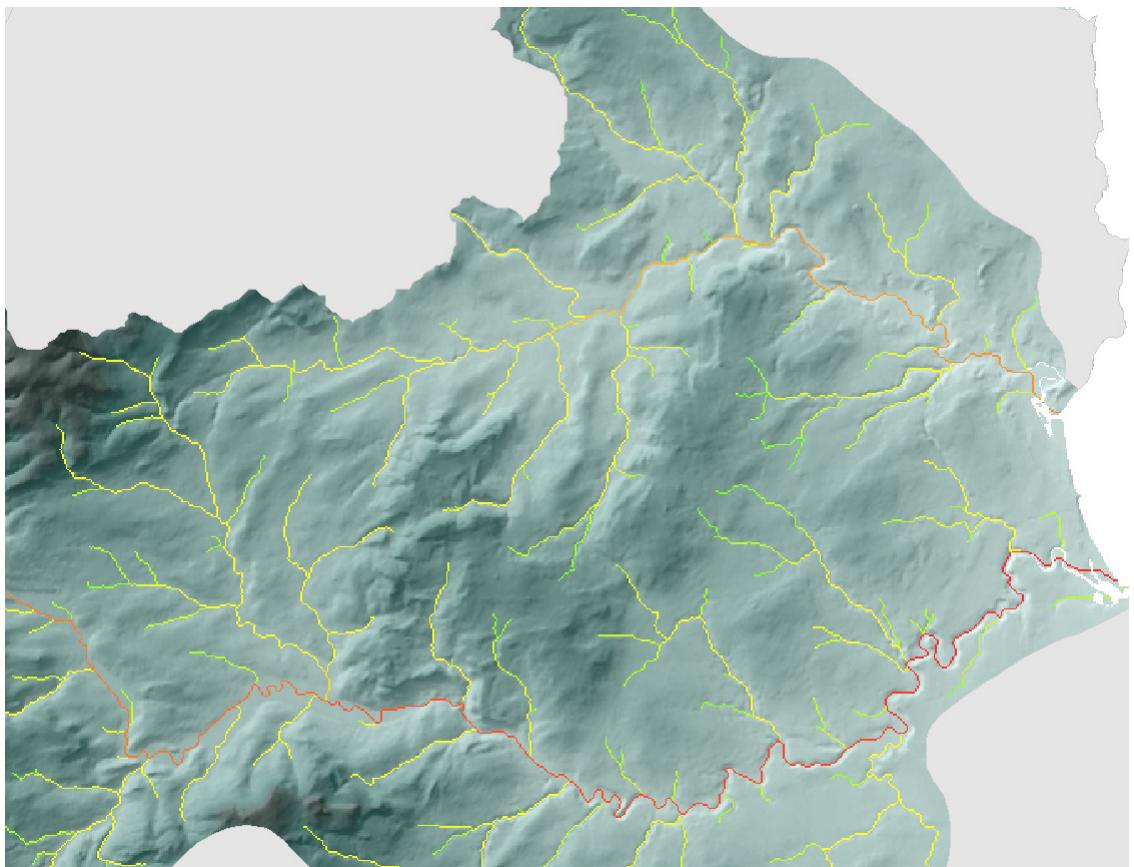
Figure 4.32   Full Model for trout fry the River Aln and Coquet example
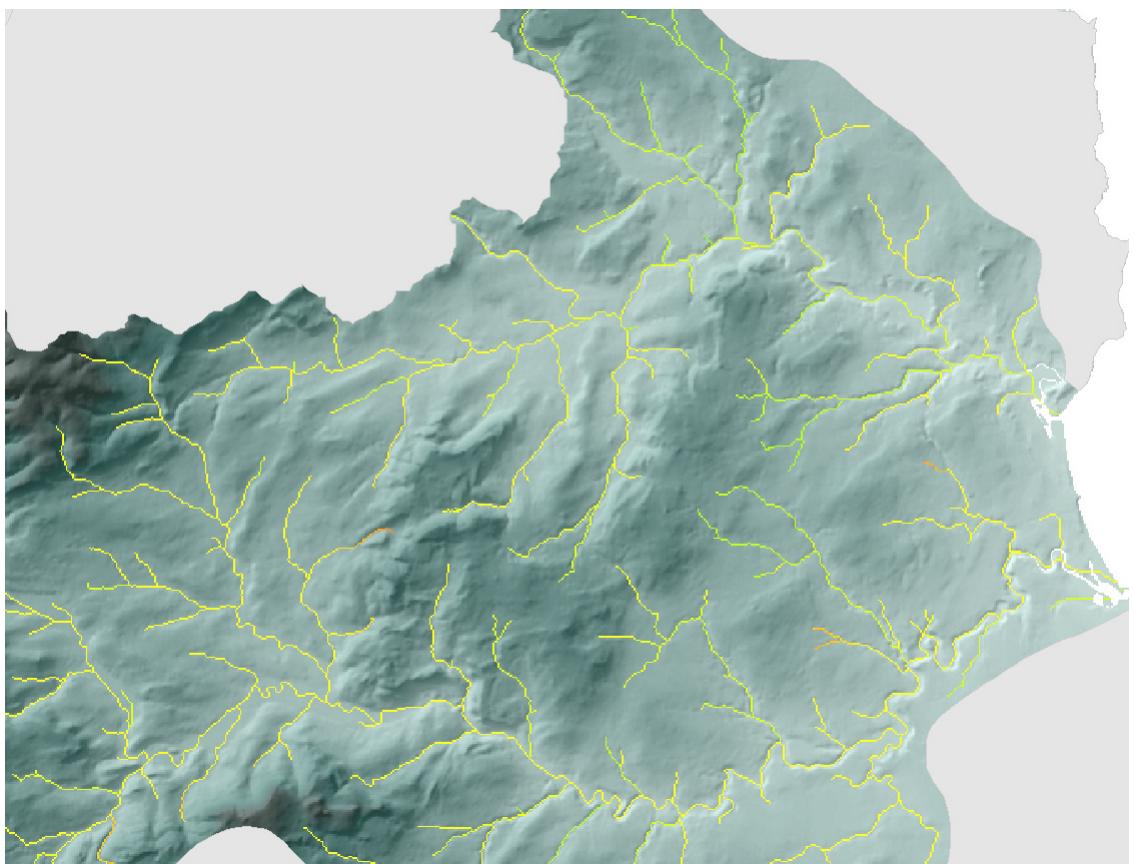
**Figure 4.33   Landcover effect for trout fry the River Aln and Coquet example**
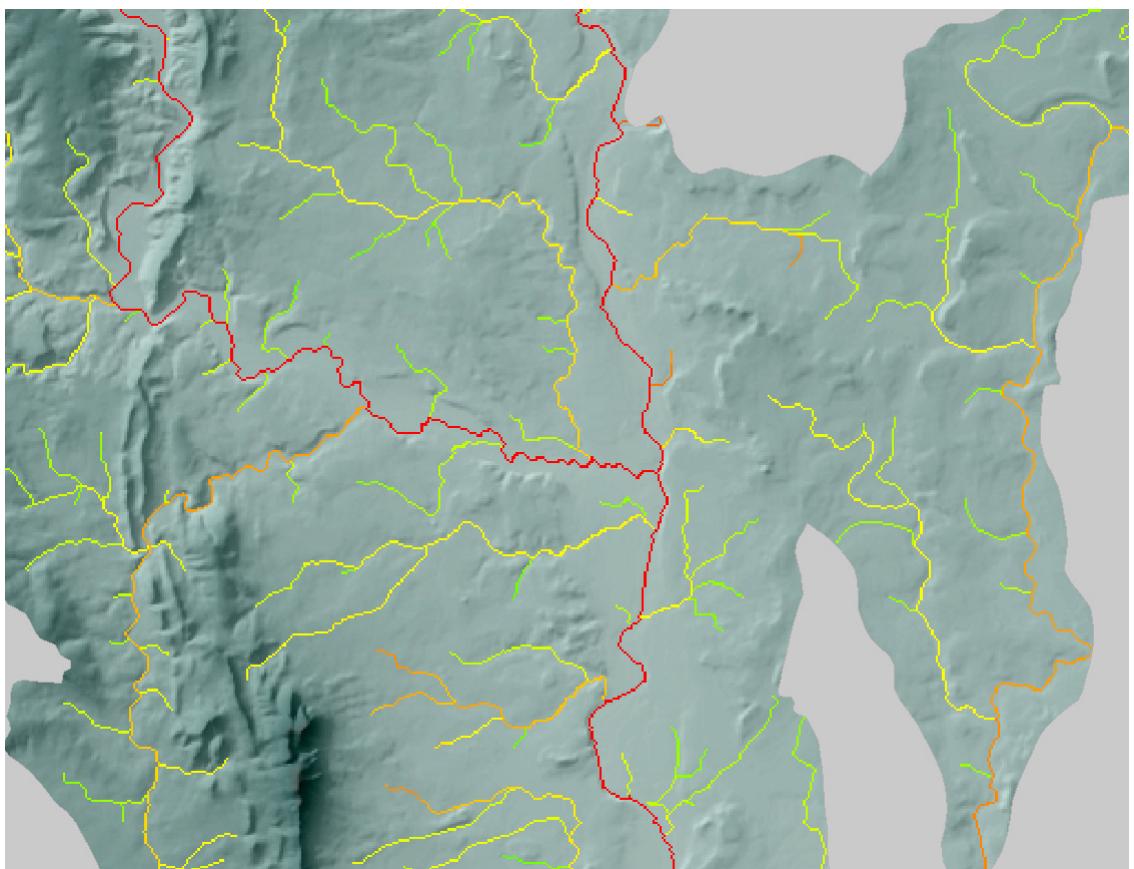
Figure 4.34   Full Model for trout fry the Middle River Severn example
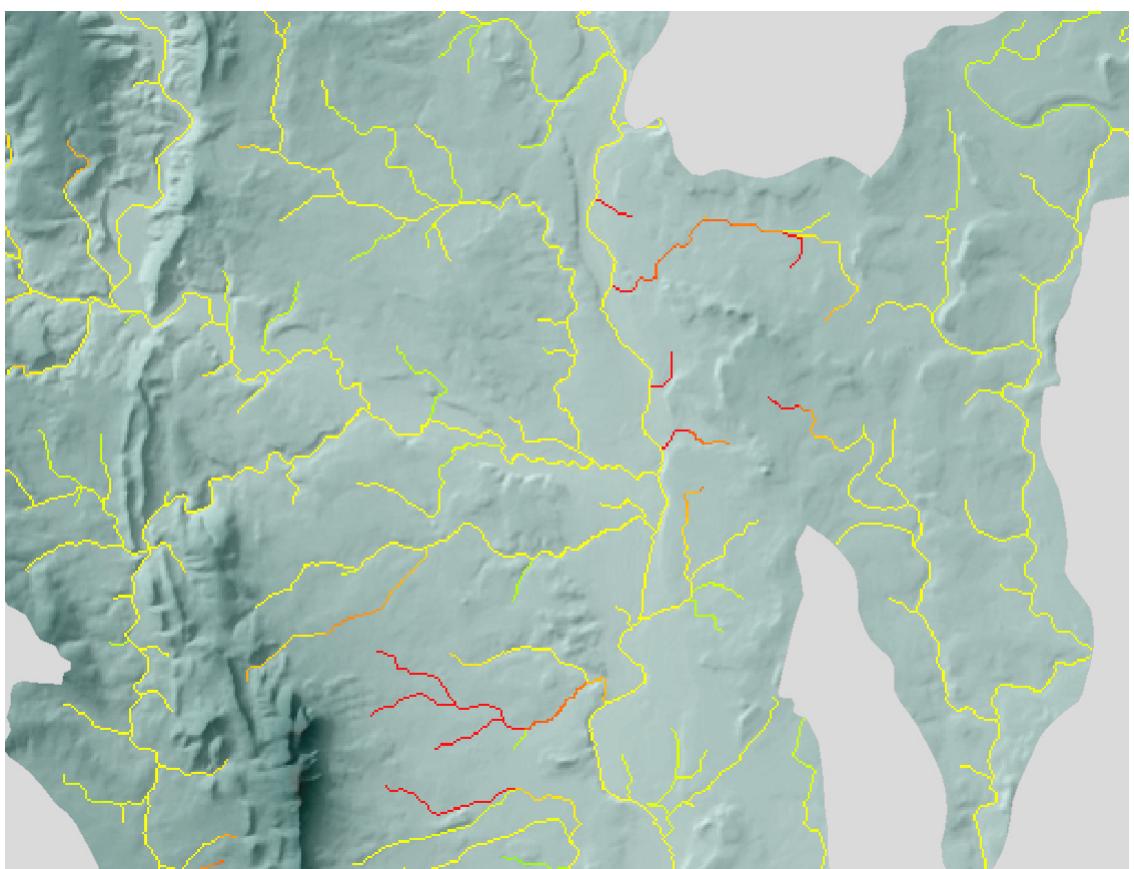

Figure 4.35   Landcover effect for trout fry the Middle River Severn example

Figure 4.36   Full Model for trout fry the River Dart example

Figure 4.37   Landcover effect for trout fry the River Dart example

The River Aln and Coquet examples show a catchment coloured as expected with a gradation from poor habitat (in terms of the model) to good habitat higher up the catchment. The main river is seen to be less suitable for trout fry than the tributaries. However, to the extremes main rivers may lie outside of the models. This effect was not seen in the salmon fry model. The Landcover effect model (Figure 4.33) suggests that there are only a few small tributaries with medium negative landcover effects, these correspond to tributaries that source in predominantly urbanised land.

The Middle River Severn examples show a catchment coloured as the majority of the habitat (in terms of the model) being good. Again the main river is seen to be less suitable for trout fry than the tributaries. There are two sections that show poor habitat quality (reduced predicted trout fry densities). The landcover effect model (Figure 4.35) highlights the sections even more strongly. Again these correspond to tributaries that source in predominantly urbanised land (towns).

The River Dart examples is another typical catchment coloured with a gradation from poor habitat in the lower catchment areas (in terms of the model) to good habitat higher up the catchment. As with the other two examples the main river is seen to be less suitable for trout fry than the tributaries. The Landcover effect model (Figure 4.37) suggests that there are only a few small tributaries with high negative landcover effects, these again correspond to tributaries that source in predominantly urbanised land.

Figure 4.38 depicts the proportions of the river network that fall into 6 categories of predicted log trout fry (density +1). The y-axis corresponds to the number of $50m^2$ cells that fall in each category. This equates to kilometres of river network if multiplied by 50 and divided by 1000. This chart shows that of the modelled areas in England and Wales the majority of cells fall in the central region of predicted values as with the salmon fry example shifted slightly towards the higher end of the scale.

Figure 4.39 depicts the proportions of the river network that fall into 6 categories of predicted log trout fry density relating from the landcover effect model. This chart shows that of the modelled areas in England and Wales the majority of cells fall in the central region of predicted values (medium landcover effect).
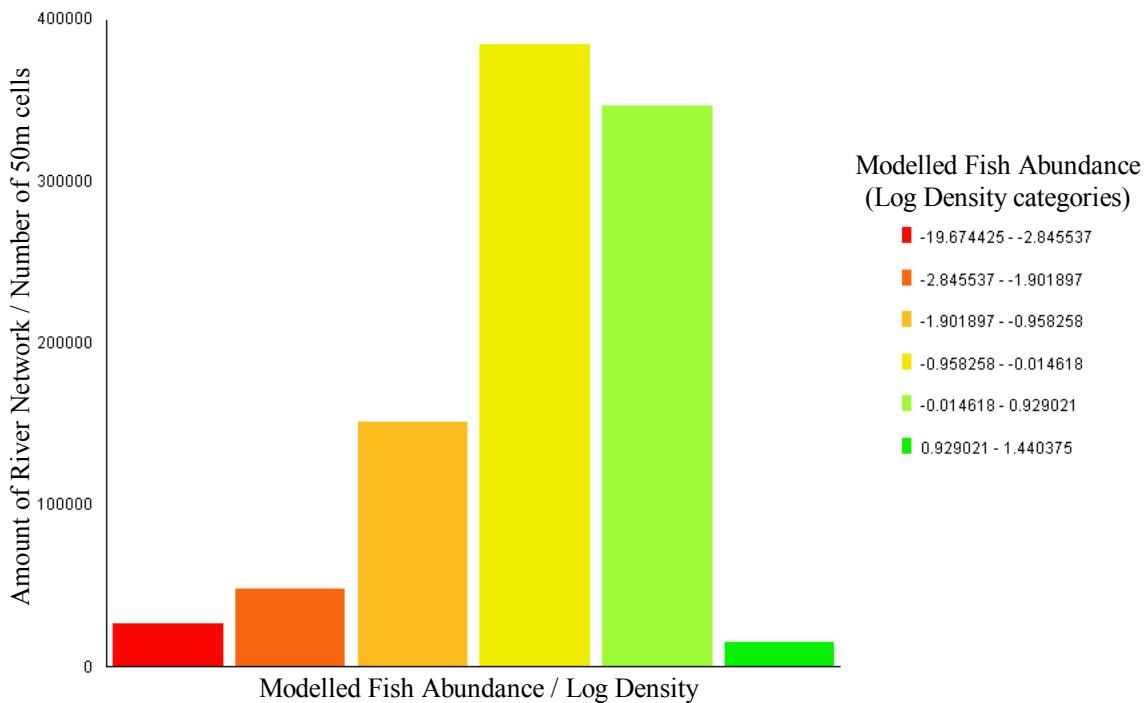
Figure 4.38 Chart depicting the proportion of 50m² cells in each category (Log Density trout fry)
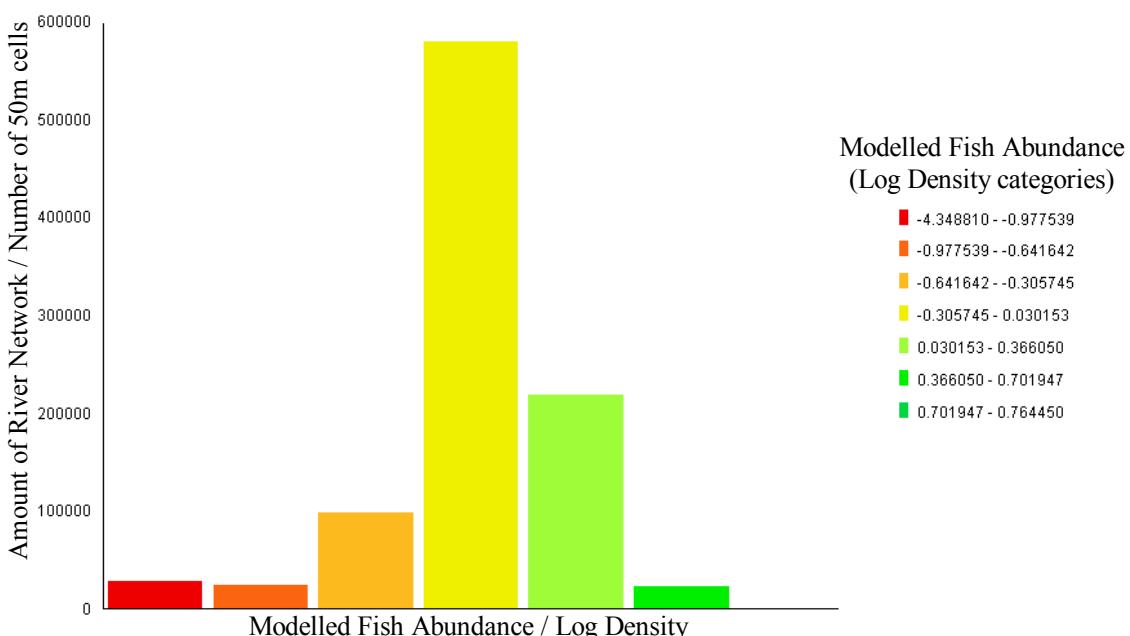

Figure 4.39 Chart depicting the proportion of 50m² cells in each category of landcover effect (Log Density trout fry) where red = high landcover effect

4.4   Discussion

The project can be split into a number of distinct but interrelated areas that will be discussed separately:
1. GIS methods and data,
2. Juvenile salmonid data,
3. Model development,
4. Model outputs – predicted densities,
5. Landcover effects.

GIS methods and data

The analysis has demonstrated that GIS derived variables are certainly of use in helping to understand juvenile salmonid populations, although the predictive value of the particular example models are poor. However, in more detailed studies designed specifically to produce predictive models or in studies of specific catchments the predictive value of the models could be greatly improved. In addition a number of the other new attributes that will be available should introduce a greater scope to improve the models and introduce the ability to predict influencing factors and risk. One particular part of the GIS methodology that is lacking is the production of site gradient that is comparable to that of the manual methods. This is a particular weakness, as a number of methods will certainly require this attribute that has been shown to be of use in describing salmonid populations. HABSCORE (Barnard, Wyatt and Milner, 1995, Wyatt *et al.*, 1995b), the Fisheries Classification Scheme (Hay *et al.*, 1995, Mainstone *et al.*, 1994, Wyatt and Lacey, 1998), the River Fisheries Habitat Inventory project (Wyatt, 2001) and the Salmon Lifecycle project (Wyatt and Barnard, 1997a) may all use site gradient within their methodologies. Further work will be required to achieve the automated production of this attribute within the GIS.

Juvenile salmonid data

The fish survey data available for the investigation are predominantly from 1985 to 1998 but cover the period from 1974 to 2001. This data is of a limited coverage in terms of available sites and in terms of the years available. More recent data exists but was not available, as national querying of the National Fish Population Database (NFPD) is currently not feasible. In this study only fully quantitative fish survey data was used, however, there is a considerable quantity of semi-quantitative fish survey data available. Future work should aim to use a greater quantity of available data and not rely on just the fully quantitative survey data. Such work is proposed within the River Fisheries Habitat Inventory project (Wyatt, 2001) and will require the use of GIS derived variables as described within this study.

Model development

The methods employed were not designed to produce a model for predicting fish densities using both field and map based attributes such as in HABSCORE (Barnard, Wyatt and Milner, 1995, Wyatt *et al.*, 1995b). They were designed more as a relatively crude method to investigate and demonstrate the usefulness of the new GIS derived attributes in describing the variation of fish populations as had been done at an earlier stage by McMellin (2002), and not dissimilar to the principles described by Burrough (1986) and ESRI (2000).

The models were developed using standard statistical techniques as described by Minitab (2000). The models were developed as (nearly) national models and as such are

at a relatively large spatial scale. These models may not work well on individual catchments, where smaller scale variations might be more important and the effects of the larger scale variations masked. Also at the smaller scale, instream habitat features may be much more important.

Model outputs – predicted densities

Salmon and trout fry show quite different relationships with the variables. This has also been shown in other work such as in HABSCORE (Barnard, Wyatt and Milner, 1995, Wyatt *et al.*, 1995b) different, variables have been chosen to model salmon and trout. The trout model resulted in a slightly higher R-Squared value (23.6 % for trout fry and 18% for salmon fry). This is what might be expected, as trout populations tend to be more stable at the site level and might be considered more predictable. Salmon populations may be subject to greater temporal variation and the marine component of the lifecycle may add more variability which would make prediction of the spatial component of the total variation more difficult (Wyatt R, personal communication). The R-Squared values are still low and do not explain a high amount of the total variation as compared to models developed for HABSCORE (Wyatt *et al.*, 1995b). This is partly due to the HABSCORE models using both map-based and site-based variables. However, an additional reason is that the HABSCORE models were developed using pristine sites only, where all available site data were used in this investigation, increasing the likely variation.

The examples show that the main rivers are less suitable for trout fry than the tributaries. This effect was not seen in the salmon fry models. This result should be expected, as trout generally will not spawn in the main river whereas salmon fry and parr occupy shallow, fast flowing water whether tributaries or main river (Symons and Heland 1978, Baglinière and Champigneulle 1986, Baglinière, and Maisse, 1991).

The proportions of the river network that fall into 6 categories of predicted log trout fry or salmon fry density suggest that the majority of cells fall in the central region of predicted values shifted slightly towards the higher end of the scale. This would equate to the area modelled having generally medium to high-predicted values. It should be noted that the majority of the areas modelled are salmonid areas and so large areas of the country where no salmonids would be expected play no part in the modelling. In addition, the GIS modelling does not take into account the deeper areas of the river network and the riffle / pool sequences that characterise salmonid rivers. This prevents the model predictions being used to quantify population numbers on the whole river network.

Landcover effects

From the examples above the landcover effect that appears to be most influential on salmon fry and trout fry is that of urbanised land. For example urbanised land generally in the form of towns will result in very high values upstream in some small tributaries (in terms of proportion urbanised land upstream). The other landcover effects may be subtler and their distribution may be less clumped. This may result in the effect being not so dramatic or visually obvious as with the urban area examples. The effects of landcover will relate more to the actual landuse. This may not be defined sufficiently by the landcover classes to be picked by the modelling procedure. Impact agents such as runoff, siltation, diffuse pollution and so on may be more accurately picked up by other datasets such as agricultural landuse and impact modelling such as in McHugh *et*

al. (2003), McMellin (2002) and McMellin *et al.* (2002). Further work to acquire these additional attributes is required, as the datasets become available in the near future.

The proportions of the river network that fall into 6 categories of predicted log trout or salmon fry densities relating from the landcover effect model suggest the majority of the modelled area is mildly effected by land cover. This is to be expected as areas that have very high juvenile salmonid densities would be expected to be in areas which are near to pristine and not affected by adverse landcover effects to the degree that may occur in other areas. This methodology may provide the means to assist in the prioritisation of river reaches requiring further analysis and possible mitigation / remediation work.

4.5   Conclusions

The purpose of the project was twofold:
1. To look at both new GIS methods and to describe where appropriate improvements to existing methods to enable the calculation of attributes.
2. To examine the suitability and usefulness of the new attributes and the existing attributes for telling us something about juvenile salmonid populations.

The introduction of the improved hydrological analysis methods including the AGREE methodology have greatly enhanced the ability of the GIS to provide attributes that help us to understand juvenile salmonid populations. Methodologies have been refined so those variables can be obtained in a systematic way, allowing for the application of any regression models to the dataset as a whole. There is a much greater availability of attributes that can be calculated and a number of these have been shown to be useful. Landcover variables have been shown to be of some use in describing juvenile salmonid populations.

There are a number of significant advantages of the GIS methods over existing manual methods.
1. Mapping and the ability to view distributions spatially
2. Speed of attribute derivation is greatly increased when deriving for a large number of sites.
3. The methodology can be standardised and removes some of the scope for error.
4. The methods allow for the calculation of attributes for every "cell" on the river network.
5. The methods allow for the application of any regression models within the GIS.

These GIS derived attributes may be suitable for any Environment Agency methods that use or involve parameters that are taken from maps. There are three significant limitations that need to be overcome:
1. Feasibility of deriving stream widths. There are future data sets that may contain width attributes at set intervals or averages for a reach. Work by Wyatt (in press) allows for the prediction of the widths at points on a river by interpolating from measurements upstream and downstream on the same catchment. The method would need to be refined to enable its use at scales larger than individual catchments.
2. Calculating Site Gradient in a method comparable to the manually derived method.
3. Modelling fish in deeper waters. A pool/riffle model may be required in the future.

Once these two questions are overcome the use of GIS within fisheries will become more powerful as a management and investigation tool.

5. RECOMMENDATIONS

5.1 Methodology and GIS Data

The current methodology provides the ability to calculate a wide selection of variables for fish survey sites in England and Wales. The following recommendations would allow for further improvements and a consistent approach to the calculation of required data.

- Productions of attribute data sets for all of England and Wales following the methodology. Selected data sets could be available via the national I-drive to provide consistent data for Regional and Area Fisheries Staff.

- Use of GIS methodology in preparing data sets, delivery of the ArcHydro tools to users that require this capability.

- Methods for production of attributes to be utilised within existing Environment Agency methodologies requiring manual map derived attributes.

- Acquisition and investigation of new data sets that may provide useful information in understanding fish populations.

- Study to improve the detail of derived river networks using SPR grids as described in Section 2.5.1

- Study to achieve a method of deriving Site Gradient in a way that is consistent with the manual map derived method. This must be undertaken in a way that would provide data for the whole river network to allow for analysis techniques to utilise these data.

5.2 Fish Data

The following recommendations aim to improve the quality and quantity of national data sets.

- There are a number of areas where fully quantitative fish survey data is missing, particularly the River Severn, the River Tyne, North West Region and the Southern Chalk streams. These holes in the data could be filled.

- Semi-quantitative datasets should also be included to provide data with a greater spatial distribution and to provide national data for improved modelling techniques.

- The National Fisheries Population Database (NFPD) provides the opportunity to investigate a wide dispersion of data covering all of the national monitoring not just for salmonids but also for coarse fish. The ability to export data into a format that would allow for national studies would be essential in achieving the potential of these data stores.

5.3 Proposed Research

There are a number of GIS based techniques not covered by this project that may provide useful options in the investigation and understanding of fish populations. For example the following could be investigated further.

- The use of Geostatistics to help in prediction, these spatial statistics methods such as interpolation of data, linked to prediction could improve modelling capabilities, particularly in relation to coarse fish.

- The use of other parts of the ArcHydro data model to facilitate catchment management techniques enables better use of data. For example the method termed "dynamic segmentation" for reach based work could provide some use in reporting of the monitoring programme, water framework directive etc. Also the ArcHydro data model would provide the ability to calculate link number and down stream link number. This may give a better indication where on a catchment a site lies and whether the stream goes into another small stream or main river. Link numbers also provide data that can be interpreted by statistical packages looking at the connectivity or adjacency of sites to each other.

- GIS based attributes could be investigated in a similar way in relation to coarse fish data, to provide a broad scale prediction tool. This could provide a useful addition to the suite of management tools available to Area Fisheries Staff.

- Wyatt and Lacey (1999) suggest that the Fisheries Classification Scheme (FCS) can be used with semi-quantitative data by using map based data to enable the semi-quantitative data to be corrected for fishing conditions (that is to say habitat, for example with sea-trout, these could be gradient and distance to source). This calibration of population estimates increases the usefulness of the semi-quantitative surveys. Using GIS would reduce the time required, increase the accuracy and effectiveness. However, the issue of site gradient still requires some work.

REFERENCES

Argent D G, Bishop J A, Stauffer Jr J R, Carline R F, and Myers W L, 2003 *Predicting freshwater fish distributions using landscape-level variables*. Fisheries Research 60:1 17-32

Baglinière J-L, and Chamigneulle A. 1986 *Population estimates of juvenile Atlantic salmon (Salmo salar) as indices of smolt production in the Scorff River, Brittany France*. Journal of Fish Biology, 29(4): 467-482.

Baglinière J-L, and Maisse G. 1991 *Biology and ecology of the Brown and Sea Trout.* pp286 Chichester: Praxis Publishing

Barnard S, Wyatt R J, Milner N, 1995 *The development of habitat models for stream salmonids, and their application to fisheries management*. Bulletin Fr. Peche Piscic 337/338/339: 375-385

Boorman, D B, Hollis J M, and Lilly A, 1995 *Hydrology of soil types: a hydrologicaly-based classification of the soils of the United Kingdom.* Institute of Hydrology Report No. 126.

Broadbridge N, 2000 *The impact of conifer harvesting on upland fish populations: Determining fish distribution as part of an ameliorative strategy*. R&D Technical Report P2-268/TR & PR Bristol: Environment Agency

Burrough, P A, 1986 *Principles of Geographical Information Systems for Land Resources Assessment*. New York: Oxford University Press, p 50.

Crundwell C R, 2002 *Transporting Salmon Spawning Targets.* Environment Agency R&D report: W2-034/PR

Elser A A, 1968 *Fish populations of a trout stream in relation to major habitat zones and channel alterations*. Transactions of the American Fisheries Society. 97: 389-397.

Environment Agency. 1996 *River Habitat Survey. Field Survey Guidance Manual*. Environment Agency

ESRI, 2000 *Spatial Analyst Functional Reference.* Redlands: Environmental Systems Research Institute (ESRI)

ESRI, 2000b *Spatial Analyst Online Help*. Redlands: Environmental Systems Research Institute (ESRI)

ESRI, 2002 Online Help for Arc Hydro Tools. Version 1.0 Release 1, June 2002 Redlands: Environmental Systems Research Institute (ESRI)

Fairfield J, and Leymarie P, 1991 *Drainage networks from grid elevation models*. Water Resources Research. 30(6): 1681-1692

Fuller R M, Smith G M, Sanderson J M, Hill R A, Thomson A G, Cox R, Brown N J, Clarke R T, Rothery P, and Gerard F F. 2002 *Countryside survey 2000 module 7: landcover map 2000 - a guide to the classification system*. Centre for Ecology and Hydrology (Natural Environment Research Council)

Garbrecht J, and Martz L W, 1994 *Grid size dependency of parameters extracted from digital elevation models*. Computers and Geosciences 17(3): 413-422

Garbrecht J, and Martz L W, 2000 *Digital elevation model issues in water resources modelling*. In: Hydrolic and hydraulic modelling support with GIS (ed Maidment D, and Djokic D). 232 pp Redlands: ESRI Press.

Garbrecht J and Starks P, 1995 *Note on the use of GSGS level 1 7.5-minute DEM coverages for landscape drainage analysis*. Photogrammetric Engineering and Remote Sensing 61(5):519-522

Haines-Young R H, Barr C J, Black H I J, Briggs D J, Bunce R G H, Clarke R T, Cooper A, Dawson F H, Firbank L G, Fuller R M, Furse M T, Gillespie M K, Hill R, Hornung M, Howard D C, McCann T, Morecroft M D, Petit S, Sier A R J, Smart S M, Smith G M, Stott A P, Stuart R C and Watkins J W, 2000 *Accounting for nature: assessing habitats in the UK countryside*. DETR: London

Hay S V, Barnard S, and Mainstone C P, 1995 *The National Fisheries Classification Scheme*. National Rivers Authority R&D Note: 441

Hellweger F, 1997 *AGREE - DEM Surface Reconditioning System*. University of Texas. Available online at (Page accessed 11/12/2002):
http://www.ce.utexas.edu/prof/maidment/GISHYDRO/ferdi/research/agree/agree.html

Hutchinson M F, 1989 *A new procedure for gridding elevation and stream line data with automatic removal of spurious pits*. Journal of Hydrology 106: 211-232

Jenson S K and Domingue J O, 1988 *Extracting Topographic Structure*. In Digital Elevation Data for Geographic Information System Analysis, Photogrammetric Engineering and Remote Sensing. Vol. 54, No. 11, November 1988, pp. 1593–1600.

Johnson I W, Elliott C R N, Gustard A, Armitage P D, Ladle M, Dawson F H and Beaumont W R C. 1993 *Ecologically acceptable flows. Assessment of instream flow incremental Methodology*. Report to National Rivers Authority. Project Record 282/1/Wx.

Johnson L B, and Gage S H, 1997 *Landscape approaches to the analysis of aquatic ecosystems*. Freshwater Biology. 37: 113-132

Maidment D, 1999 *Spatial hydrology using ArcView 3.x*. ESRI Virtual campus: Environmental Systems Research Institute. Available online at
(Page accessed 15/01/2003):
http://campus.esri.com/acb2000/showdetl.cfm?&DID=6&User_ID=895963&st=9876&st2=52294115
&st3=-80109075&Product_ID=604&CATID=83&CFID=4935734&CFTOKEN=48775019

Maidment D, 2002 *ArcHydro GIS for water resources*. Redlands: ESRI Press.

Mainstone C P, Barnard S, and Wyatt R J, *1994 Development of a Fisheries Classification Scheme.* National Rivers Authority R&D Output: PR 244/7/NY

Mark D M, 1988 *Network Models in Geomorphology*, In Modelling in Geomorphological Systems. New York: J Wiley & Sons

McCoy J and Johnson K, 2001 Using ArcGIS Spatial Analyst. Redlands: Environmental Systems Research Institute (ESRI) pp232

McHugh M, Wood G, Walling D, Morgan R, Zhang Y, Anthony S, and Hutchins M 2002 *Prediction of sediment delivery to watercourses from land (Phase II)*. Environment Agency R&D Technical Report No P2-209

McMellin G, 2002 *Land use and fisheries: Results of sediment trapping in artificial salmon redds, winter 1999/2000*. R&D Technical Report W2-046/TR2 Bristol: Environment Agency

McMellin G, Walling D, and Nicholls D, 2002 *Land use and fisheries: Report of the pilot study* R&D Technical Report W2-046/TR1 Bristol: Environment Agency

Milner N J, Hemsworth R J, and Jones B E. 1985 *Habitat evaluation as a fisheries tool*. Journal of Fish Biology. 27(Sppl.A), 85-108.

Minitab 2000 *Minitab online help and statistics guide*. Minitab inc.

Mizgaleuicz P J, and Maidment D, 1996 *Modelling agrichemical transport in midwest rivers using geographical information system*s. Center for Research in Water Resources Online Report 96-6, University of Texas: Austin, Texus.

Moore I D, Grayson R B, and Ladson A R, 1991 *Digital terrain modelling: A review of hydrological, geomorphological and biological applications*. Hydrological Processes 5(1):13-141

O'Callaghan J F, and Mark D M, 1984 *The extraction of drainage networks from digital elevation data*. Computer Vision, Graphics and Image Processing 28:323-344

Ormsby T, Napoleon E, Burke R, Groessl C and Feaster L. 2001 *Getting to know ArcGIS desktop*. Redlands: ESRI Press pp538.

Ordnance Survey, 1997 *Land-Form PANORAMA User Guide.* v2.0 - 06/1997 Southampton: Ordnance Survey

Saunders W, 2000 *Preparation of DEMs for Use in Environmental Modelling Analysis*. In: Hydrolic and hydraulic modelling support with GIS (ed Maidment D, and Djokic D). 232 pp Redlands: ESRI Press.

Symons P E K, and Heland M. 1978 *Stream habitats and behavioural interactions of underyearling and yearling Atlantic salmon*. Journal of the Fisheries Research Board of Canada, 35: 175-183.

Tribe A, 1992 *Automated recognition of valley heads from digital elevation models*. Earth Surface Processes and Landforms 16(1):33-49

Williamson A, Brandsma R, Pettit L and McMellin G, 2001 *The Agency's version of the NSRI spatial soil database*. Environment Agency Technical Reference (version 3.0)

Wright J F, Furse M T, Clarke R T, Moss D, Gunn R J M, Blackburn J H, Symes K L, Winder J M, Grieve N J, and Bass J A B. 1994 *Testing and further development of RIVPACS Phase 1: Main Report*. National River Authority R&D Note 453.

Wyatt R J, Barnard S, Lacey R F, 1995a *Salmonid modelling literature review and subsequent development of HABSCORE models (338/20/W)*. National Rivers Authority R&D Output: Project Record 338/20/W.

Wyatt R J, Barnard S, Lacey R F, 1995b *Use of HABSCORE V Software and application to impart assessment*. National Rivers Authority R&D Note 400.

Wyatt R J, Barnard S, 1997a *Development of a Lifecycle Simulation Model for Atlantic Salmon*. Environment Agency R&D report: TR W91

Wyatt R J, Barnard S, 1997b *The Transportation of the Maximum gain Salmon Spawning Target for the River Bush(N.I) to England & Wales*. Environment Agency R&D report: TR W65

Wyatt R J, Barnard S, 1997c *River Fisheries Habitat Inventory (Phase 1 Scoping Study)*. Environment Agency R&D technical report: TR W95

Wyatt R J, and Lacey R F, 1998 *Semi-Quantitative Methods for Fisheries Classification*. Environment Agency R&D report: TR W167

Wyatt R J, 2001 *River fisheries habitat inventory (RFHI) – Progress report*. Environment Agency R&D Progress Report Project Ref: W2-040.

Wyatt R J, In press. *Mapping the abundance of riverine fish populations: integrating hierarchical Bayesian models with a Geographic Information System (GIS)*. Canadian Journal of Fisheries and Aquatic Sciences.

ACKNOWLEDGEMENTS

APPENDIX 1   MAP ALGEBRA AND RASTER CALCULATOR

The Spatial Analyst extension allows you to solve a wide range of spatial problems by providing a powerful set of tools to perform spatial analysis (ESRI, 2000). The Spatial Analyst Functional Reference (an online help provided with the Spatial Analyst extension) provides access to information about individual commands, functions, operators, and methods available within Spatial Analyst (ESRI, 2000).

Raster Calculator

Basic or commonly used commands can be selected from the drop down menu on the Spatial Analyst toolbar (Figure A1.1). However, there are many other commands that you may wish to use and these are run by using the Raster Calculator. The Raster Calculator can be called from the drop down menu on the Spatial Analyst toolbar (Figure A1.1), which opens a dialogue box allowing you to create equations and commands to perform within ArcMap (Figure A1.2).

Within the Raster Calculator available grids on which analysis can be performed are listed (Figure A1.2 [Box A]). Equations and commands are entered as text into the main area as Map Algebra (Figure A1.2 [Box B]).
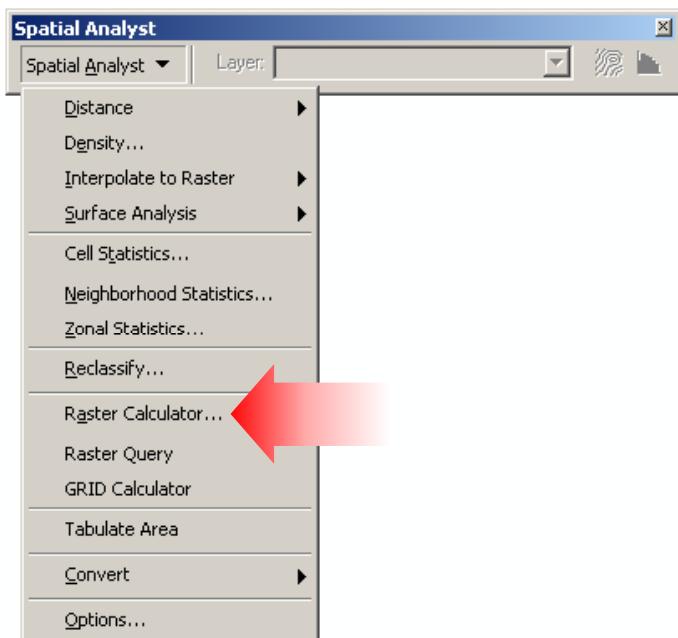

**Figure A1.1   Spatial Analyst Toolbar**
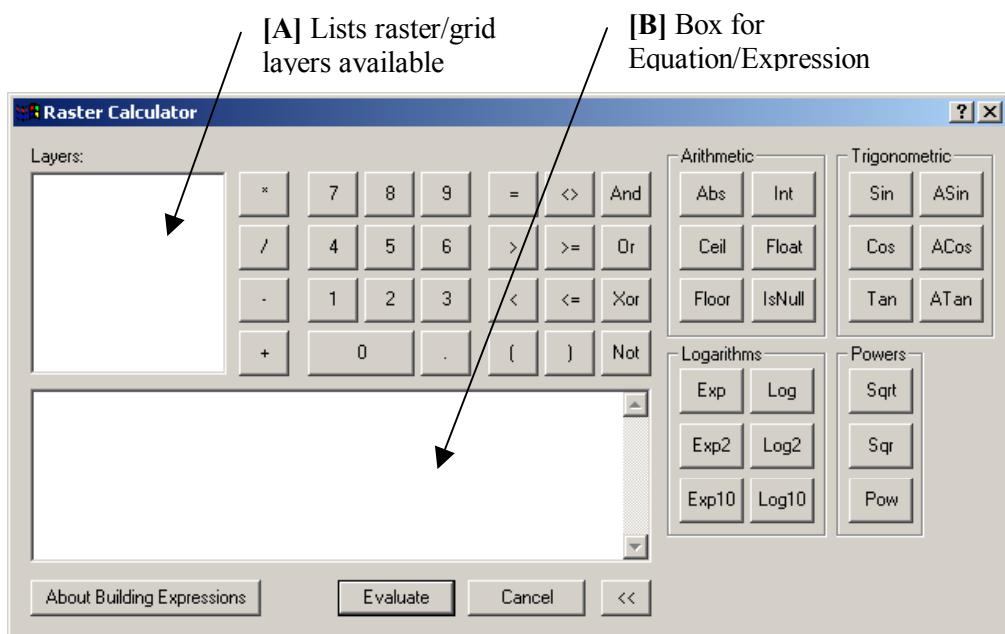
Figure A1.2   The Raster Calculator

Map Algebra

Map Algebra is the analysis language of Spatial Analyst. It is a simple language similar to any algebra where output data will result from some manipulation of input data as defined in the Map Algebra expression. The simplest way for Map Algebra expressions to be realised is for them to be entered within the Raster Calculator.

Details on the syntax of Map Algebra can be found in the Spatial Analyst manual (McCoy and Johnson, 2001), the Spatial Analyst Functional Reference (ESRI, 2000) and the Spatial Analyst online help (ESRI, 2000b).

The equations or expressions used within this report are, as they would be entered in the Raster Calculator. Grids are contained within square brackets and functions can be nested within other brackets.

`[GridName]` = Raster Grid name within square brackets

APPENDIX 2   TOOLS WITHIN ARCGIS

This appendix describes a number of tools used within the ArcGIS software within the project. This appendix also contains the Visual Basic script used for these tools. The following tools are described:

1. Batch Snapping of Selected Points
2. Adding X & Y Co-ordinates to attributes table
3. Extract Values of a Raster Layer to a Point Shapefile

If you are not familiar with the ArcGIS / ArcView 8.x software then the following book is recommended: Ormsby T, Napoleon E, Burke R, Groessl C and Feaster L. 2001 *Getting to know ArcGIS desktop*. Redlands: ESRI Press pp538.

If you are not familiar with ArcObjects and writing Macros for ArcView then the following book is recommended: Zeiler M. 2001 *Exploring ArcObjects Volume 1 & 2*. Redlands: ESRI Press.

1. Batch Snapping of Selected Points

This code allows you to snap selected based on the current snapping environment settings.   Within this project the sites must be snapped to the derived river network to allow for attributes to be correctly derived (See Figure A2.1).

The code used is an ArcObjects Visual Basic script that is implemented as a macro in ArcMap. The code can be found on the ESRI ArcObjects web site at: (27/12/2000) http://arcobjectsonline.esri.com/arcobjectsonline/samples/editing/batchsnap.htm
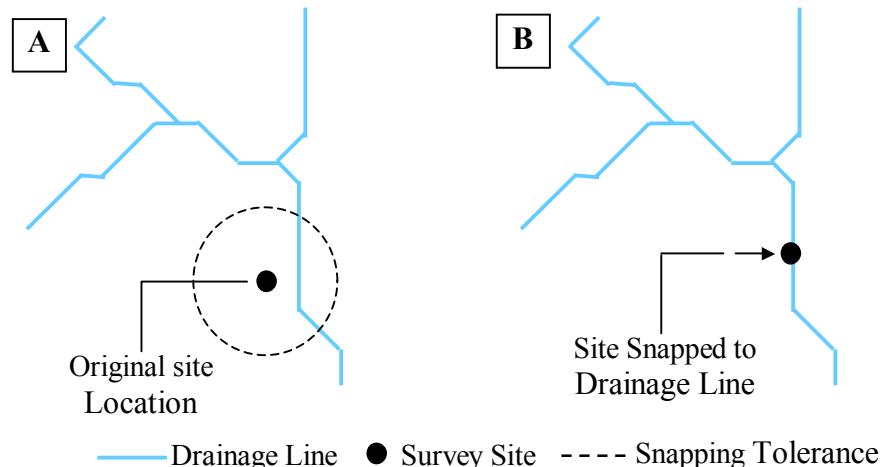

Figure A2.1   **Snapping of Survey Site to Derived River Network**

ArcObjects Code Batch Snapping:

```vb
Public Sub BatchSnapping()
  Dim pEditor As IEditor, pSnapEnv As ISnapEnvironment
  Dim pMxDoc As IMxDocument
  Dim pFeature As IFeature, pEnumFeat As IEnumFeature
  Dim pPoint1 As IPoint, iCount As Integer, iTotCount As Integer
  Dim pUID As New UID, pNetFeat As INetworkFeature

  'Set the editor and the snap environment variables
  pUID = "esricore.editor"
  Set pEditor = Application.FindExtensionByCLSID(pUID)
  Set pSnapEnv = pEditor

  'Make sure we have selected features
  If pEditor.SelectionCount = 0 Then
    MsgBox "You don't have any features selected!!"
    GoTo LeaveSub
  End If
    'Loop through the selected features snapping only the points
  Set pEnumFeat = pEditor.EditSelection
  iCount = 0
  iTotCount = 0
  Set pFeature = pEnumFeat.Next
  pEditor.StartOperation
  While Not pFeature Is Nothing

    'Check to make sure selected feature is a point feature
    iTotCount = iTotCount + 1
    If pFeature.Shape.GeometryType = esriGeometryPoint Then
      Set pPoint1 = pFeature.ShapeCopy
      'Check to see if the location of the point changed (indicating it snapped),
  'and store the new feature location if it did.
      If pSnapEnv.SnapPoint(pPoint1) Then
        Set pFeature.Shape = pPoint1
        iCount = iCount + 1
        pFeature.Store

        'Connect if a simple junction
        If pFeature.FeatureType = esriFTSimpleJunction Then
          Set pNetFeat = pFeature
          pNetFeat.Connect
        End If
      End If
    End If

    'Get the next feature
    Set pFeature = pEnumFeat.Next
  Wend
  pEditor.StopOperation "Bulk move"
  MsgBox CStr(iCount) + " of " + CStr(iTotCount) + " moved"

  'Refresh display if something moved
  If iCount > 0 Then
    Set pMxDoc = ThisDocument
    pMxDoc.ActiveView.Refresh
  End If
```

```
        GoTo LeaveSub

    Exit Sub

LeaveSub:
    'Clear out the object variables
    Set pMxDoc = Nothing
    Set pEditor = Nothing
    Set pSnapEnv = Nothing
    Set pPoint1 = Nothing
    Set pFeature = Nothing
    Set pEnumFeat = Nothing
    Set pNetFeat = Nothing

    End Sub
```

2. Adding X & Y Co-ordinates to attributes table

Moving shapes or points within layers will not update the Eastings and Northings that were used to create the layer it may be useful to add the new X and Y co-ordinates to the attribute table. Using a tool called "Add XY" this can be done. This tool is activated in ArcCatalog by registering the AddXY.dll file in the Customise dialog. The file can be found at:

C:\arcgis\arcexe82\ArcObjects Developer Kit\Samples\Geodatabase\Database Tools\Write XY Values\AddXY.dll

You then add the AddXY Sample command (found under Developer Samples) to a toolbar as in Figure A2.2.
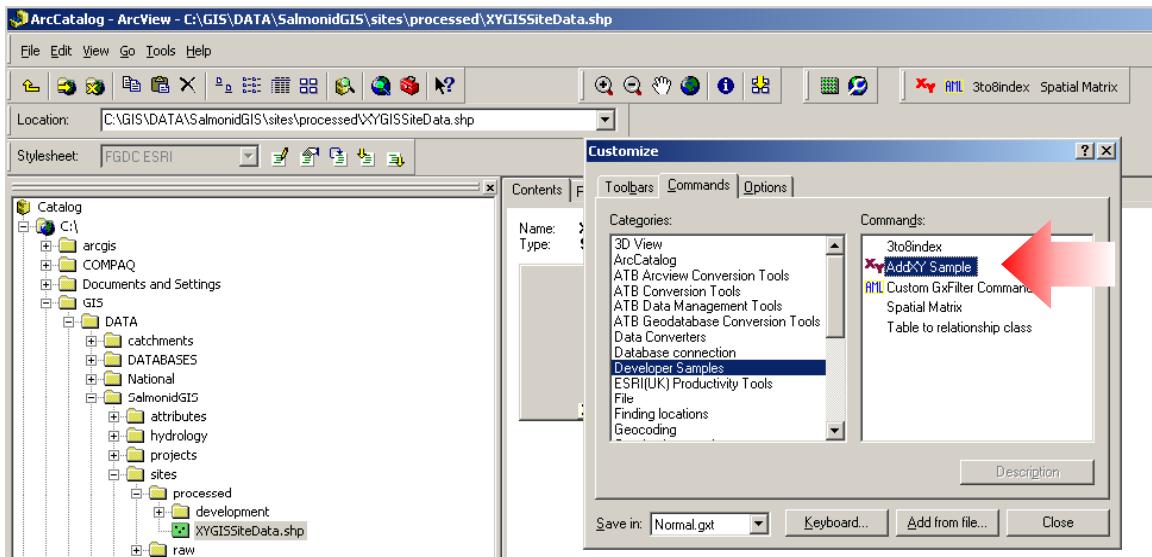

Figure A2.2   Adding the AddXY sample command

3. Extract Values of a Raster Layer to a Point Shapefile

This sample code is used to extract the values of a raster grid at given points defined by a point feature class and add those values to the feature class attribute table.

The code is in two parts, part one was produced with help from Chris Watson at ESRI UK. Part two is available on the ArcObjects website at: (5/14/2001)
http://arcobjectsonline.esri.com/ArcObjectsOnline/default.asp?URL=/ArcObjectsOnline/Samples/Raster/ExtractValueToPointFeatureClass.htm

To use the code the point file to be used should be uppermost in the table of contents and the raster value should be the next item in the table of contents. The field created will be called "rasterval".

Code Part 1:

```
Private Sub UIButtonControl1_Click()
Dim pDoc As IMxDocument
Set pDoc = ThisDocument

'get the raster which needs to be the first layer in the Table of Contents.
Dim pRasLayer As IRasterLayer
Set pRasLayer = pDoc.FocusMap.Layer(1)
Dim pInRas As IRaster
Set pInRas = pRasLayer.Raster
'get the feature layer
Dim thelayer As IFeatureLayer
Set thelayer = pDoc.FocusMap.Layer(0)
Dim theFC As IFeatureClass
Set theFC = thelayer.FeatureClass
ExtractValueTOPointFeatureClass pInRas, theFC, "rasterval"

End Sub
```

Code Part 2:

```
Sub ExtractValueTOPointFeatureClass(pInRaster As IRaster, pInFeatureClass As IFeatureClass, sFieldName As String)
    ' pInRaster: input raster
    ' pInFeatureClass: input point feature class
    ' sFieldName: name of the field that stores the values

    On Error GoTo ERH

    ' Define field name
    Dim pFld As IFieldEdit
    Set pFld = New Field
    pFld.Name = sFieldName

    ' Define field type
    Dim pProp As IRasterProps
    Set pProp = pInRaster
    If pProp.PixelType = PT_CHAR Or pProp.PixelType = PT_UCHAR Then
        pFld.Type = esriFieldTypeString
        pFld.Length = 20
        pFld.Required = 0
    ElseIf pProp.PixelType = PT_FLOAT Or pProp.PixelType = PT_DOUBLE Or pProp.PixelType Then
        pFld.Type = esriFieldTypeDouble
        pFld.Length = 24
        pFld.Required = 8
    Else  ' for integer case
        pFld.Type = esriFieldTypeInteger
        pFld.Length = 24
        pFld.Required = 0
```

```vb
            End If

             ' Add field
             pInFeatureClass.AddField pFld

             ' Get field index
             Dim FieldIndex As Integer
            FieldIndex = pInFeatureClass.FindField(sFieldName)
            If FieldIndex < 0 Then Exit Sub

             ' Create a raster layer and QI for IIdentify interface
             Dim pRLayer As IRasterLayer
            Set pRLayer = New RasterLayer
            pRLayer.CreateFromRaster pInRaster
            Dim pIdentify As IIdentify
            Set pIdentify = pRLayer

            Dim pIDArray As IArray
            Dim pRIDObj As IRasterIdentifyObj
            Dim I As Long
            Dim pPoint As IPoint
            Dim pFeature As IFeature
            Dim pNewPoint As IPoint
            Set pNewPoint = New Point

             'Loop through each point in the feature class and obtain value of the
            'raster on that point
             Dim NumOfRow As Integer
            NumOfRow = pInFeatureClass.FeatureCount(Nothing)
            For I = 0 To NumOfRow - 1
                'Get point
                Set pFeature = pInFeatureClass.GetFeature(I)
                Set pPoint = pFeature.Shape
                pNewPoint.X = pPoint.X
                pNewPoint.Y = pPoint.Y
                'Get RasterIdentifyObject on that point
                Set pIDArray = pIdentify.Identify(pNewPoint)
                If Not pIDArray Is Nothing Then
                    Set pRIDObj = pIDArray.Element(0)
                     'Get the value of the RasterIdentifyObject and add it to the field
                     If pProp.PixelType = PT_CHAR Or pProp.PixelType = PT_UCHAR Then
                    pFeature.Value(FieldIndex) = pRIDObj.Name
                    ElseIf pProp.PixelType = PT_FLOAT Or pProp.PixelType = PT_DOUBLE Or
    pProp.PixelType Then
                        If pRIDObj.Name < >  "NoData" Then
                        pFeature.Value(FieldIndex) = CDbl(pRIDObj.Name)
                        End If
                    Else    ' for integer case
                        If pRIDObj.Name < > "NoData" Then
                        pFeature.Value(FieldIndex) = CLng(pRIDObj.Name)
                        End If
                    End If
                    pFeature.Store
                End If
            Next I
            Exit Sub
        ERH:
            MsgBox Err.Description
End Sub
```