Inverse modelling of diffuse pollution sources in the River Eden catchment

BUCKLEY, CHRISTOPHER,PATRICK

DURHAM UNIVERSITY

INVERSE MODELLING OF DIFFUSE POLLUTION SOURCES IN THE RIVER EDEN

by

Christopher P. Buckley

A thesis submitted for the
degree of Master of Science

in the

Department of Geography

May 2010

"The activist is not the man who says the river is dirty. The activist is the man who cleans up the river. "

Ross Perot

Abstract

The UK has an obligation for its waters to meet the minimum standards as set out in the Water Framework Directive legislation by 2015. In recent years tighter controls on pollutants from point sources has led to diffuse sources (i.e. agricultural) having a greater contribution to degradation in water quality.

The Catchment Sensitive Farming program has been set up to attempt to advise and support landowners and farmers with various land mangement techniques which can be applied to rural areas to mitigate against some of the contributions that agricultural activities have to poor water quality.

In order for any such measures to be either cost-effective or successful at improving water quality they must be applied in suitable areas of a catchment. This research takes the River Eden catchment in Cumbria as a case study and uses mathematical modelling of measured low resolution field nutrient data together with high-resolution quasi-continuous discharge data to drive a reduced complexity diffuse pollution modelling framework (SCIMAP) to identify the areas most likely to be causing water quality problems.

Results of inverse modelling showed arable land was a particular risky land use within the Eden catchment. Several areas (mainly surrounding the River Eden in the lower reaches) within the catchment were identified as being the most likely to be causing water quality problems. As a form of control the SCIMAP model was run with logical risk values assigned to diferent landuses as well as those derived from inverse modelling of nutrient data. The model outputs driven by the statistically improved data were very similar to those which were driven by *a priori* judgment.

Several conclusions were drawn; (1) the SCIMAP model run driven by a very simple dataset based on nationally available data produced similar results to an identical model run driven by a large nutrient and discharge dataset, suggesting that the process of identifying risky areas to further examine within a catchment can be completed relatively easily (in terms of data availability), and (2) even low infrequent nutrient data can capture enough information when combined with continuous discharge data to be used in the SCIMAP model.

Acknowledgements

I would like to thank my supervisors Professor Stuart Lane and Dr Sim Reaney of the Institute of Hazard & Risk Research at Durham University for their support, guidance, help and encouragement in the past year and a bit.

Stuart was especially patient when it came to explaining various statistical techniques and modelling concepts to me, and I am particularly grateful for his willingness to repeat the same thing multiple times if necessary! Sim has been a great source of advice on all things modelling and thesis writing!

My fellow students who I shared an office with; Emma, Dave, Ian and Jonah were a great bunch to spend time with and I am especially grateful to Dave and Emma for their friendship, help and support throughout the last year. Dave was amazingly helpful towards the end of writing up, by allowing me use of his computer and office space...thanks mate!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter outlines the background to the research contained in the thesis, gives a brief summary of the methodology used, and also sets out the structure that the thesis follows.

1.1 Background

There is strong evidence that agricultural activities are responsible for a large contribution to degradation in water quality as a result of diffuse pollution in the UK (e.g. Harris *et al.* 1992, Withers *et al.* 2002, Harris *et al.* 2005).

The large impact of agriculture is due to several reasons: (1) the many activities and land uses associated with agriculture including ploughing, nutrient rich fertiliser application and livestock impacts on watercourses which are likely to contribute to water quality issues through fine sediment addition, animal waste and leaching of fertilisers into channels; (2) agricultural land makes up a large area (76%) of the total land area of England and Wales (DEFRA 2002); (3) agriculture generally has a low impact per unit area in terms of contribution to diffuse pollution which makes it difficult to manage because there are potentially many small inputs across a wide area.

The UK has an obligation to meet the most strignent set of water quality standards ever with the implementation of the Water Framework Directive (Directive 2000/60/EEC, OJ L 327 of 22.12.2000) which came into force in 2003. Member states must meet good chemical and ecological standards in lakes and waterbodies by 2015. Research has shown that diffuse pollution is a major barrier to meeting these standards (DEFRA 2002). Therefore in order to effectively apply mitigation measures and best practice land management techniques it is necessary to prioritise which areas of catchments are to be targeted. Research has shown that blanket targetting of entire catchments with mitigation measures is neither cost-effective (Schleich *et al.* 1996) or successful in stopping pollutants entering the channel (Jokela *et al.* 2002).

A keystone of diffuse pollution understanding and this research is the concept of "Critical Source Areas" (CSAs). CSAs are any piece of land (on any scale) where the local hydrological characteristics (i.e. flow from the land is directly connected to rivers or streams) combine with a significant source of nutrient inputs (Heathwaite *et al.* 2005). This concept is key for two reasons; (1) it enables more effective targeting of mitigation methods within catchments; and (2) justifies the use of the SCIMAP modelling framework for identification of CSAs within the River Eden catchment.

To date various models have been developed to attempt to quantify and understand how diffuse pollutants move and interact within a river catchment. One of the major problems with these models has been a lack of suitable validation data for model assessment. Lane (2008) has suggested that it may be time to sidestep this data availability issue and use inverse modelling techniques of the data which is available to redfine the modelling of processes involved in diffuse pollution.

1.2 Research Aims and Objectives

The aim of this research is to develop an inverse modelling methodology to identify land use hotspots responsible for the observed phosphorus concentrations in the River Eden catchment in the North-West of England. In order to accomplish this the following research objectives have been defined:

(1) Develop a technique to improve the quality of the available water quality data in the River Eden catchment

The United Kingdom has an extensive network of measuring stations operated by the Environment Agency. This monitoring scheme is known as the General Quality Assessment (GQA) scheme and monitors the chemical, biological and nutrient status of 40,000 km of rivers and streams (Environment Agency 2008b). The phosphorus data available from the GQA scheme consists of a monthly reading of phosphorus concentration (mgl^{-1}) from a particular point in the channel at that snapshot in time from 1990 to the present day. The GQA monitoring network is more suited to assessing the impacts of point source pollutants rather than diffuse sources, however to install an alternative monitoring system within the timeframe of an MSc project would be impractical and prohibitively expensive. A technique to improve the usefulness of the phosphate concentration data must be developed. To achieve this objective a Monte-Carlo simulation technique based on the Tarras-Wahlberhg & Lane (2003) Monte-Carlo based methodology will be developed and applied to field data collected by the Environment Agency as part of the GQA scheme.

(2) Develop and apply inverse modelling techniques to attempt to deal with problems of data availability

The relationships between landcover and the availability of phosphorus are poorly constrained and uncertain (Lane 2008). The uncertainty comes from a combination of; (1) the precision and resolution of available datasets and (2) a poor understanding of the process cascade in diffuse pollution risk. Inverse modelling is a technique which uses the available data to understand the process cascade and determine which processes matter in the diffuse pollution production / transport / delivery regime (Lane 2008). The inverse modelling technique will be based on the method used in a study into salmonid fry populations in the River Eden catchment (Reaney *et al.* in review).

(3) Assess the usefulnes of inverse modelling techniques for diffuse pollution risk identification

Work has already been undertaken in the River Eden catchment to identify diffuse pollution land use hotspots using expert *a priori* judgments for the likely risk of different land uses leading to diffuse pollution generation and transport. After applying the inverse modelling techniques to the River Eden catchment this project will assess the influence of inverse modelling (as opposed to *a priori* judgments) on predicted phosphorus risk within the catchment. This will be undertaken using the SCIMAP modelling framework developed at Durham and Lancaster Universities.

1.3 Thesis Structure

The first section of the thesis looks at general issues surrounding diffuse pollution including its sources, impacts and the costs associated with treating it (Chapter 2). Chapter 3 reviews the development of several previous attempts at diffuse pollution modelling, which leads to an introduction to the approach and mathematics behind the SCIMAP risk modelling framework (chapter 4). The study catchment (River Eden) is looked at in more detail in Chapter 5.

The next chapter (Chapter 6) looks at whether or not a Monte-Carlo based statistical simulation model can improve the quality of the existing Environment Agency dataset of nutrient concentrations by estimating flow weighted concentration measurements. Chapter 7 explains the inverse modelling which was undertaken on the phosphorus concentration dataset and also an application of the SCIMAP risk mapping tool using both; (1) *a priori* risk weightings for landuse, and (2) the landuse weightings from inverse modelling and compares the results.

Chapter 8 contains a summary discussion of the research approach and findings from the MSc project and finally chapter 9 briefly looks at further reccomendations for work in the River Eden catchment.

Chapter 2

Diffuse Pollution

This chapter looks at the sources, impacts and leglisation surrounding diffuse pollution in the UK and also considers possible techniques which can be used to mitigate against water quality degradation caused by diffuse pollutants.

2.1 What is diffuse pollution?

Diffuse pollution is defined as pollution which can not be attributed to a particualar point in the landscape, and hence is often referred to as non-point source. Unlike traditional point sources, such as sewage treatment plant ouputs, sources of diffuse pollution are both hard to identify and thus not directly controlled or monitored (Carpenter *et al.* 2008). Lane *et al.* (2008) recently suggested that diffuse pollution may not be as diffuse as once thought and suggests that diffuse pollution is in fact made up of individual point sources (i.e. fields with particular properties). There are a variety of sources of diffuse pollution (see Figure 2.1) but recently the attempts to monitor and control them have been focused on agricultural activities. There are good reasons for this; (1) 76% of land use in England and Wales is agricultural; (2) common agricultural practices and activities such as ploughing of land leading to increased erosion susceptibility and the application of fertilisers can all contribute to increased diffuse pollution risk; (3) agriculture uses significant amounts of fertilisers which contain large amoutns of nitrates and phosphorus.

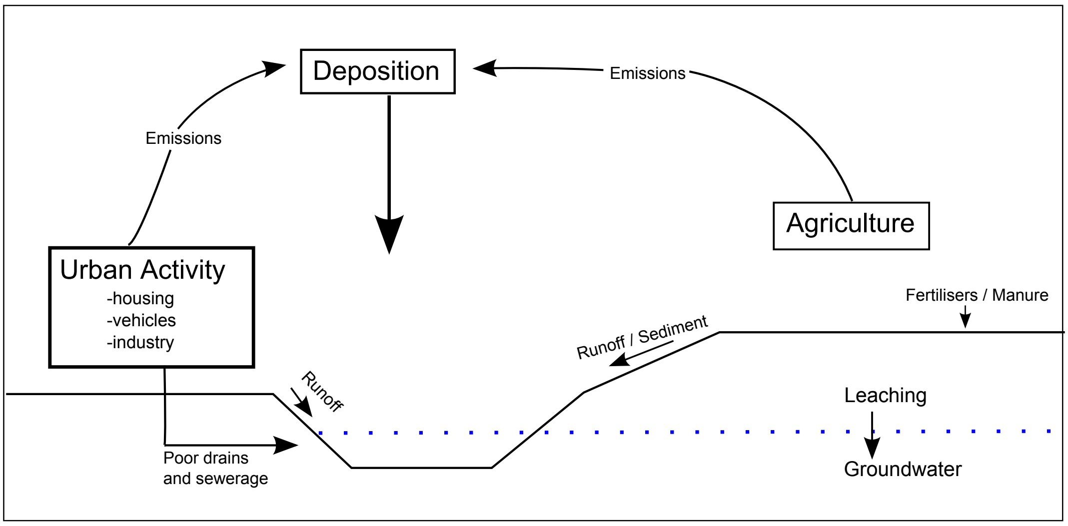

FIGURE 2.1: Conceptual model of common diffuse pollution sources

2.2 Water quality and leglislation

Both the UK and the EU are striving to improve water quality through legislation; nutrient over-enrichment is at the centre of the Water Framework (WFD) (Directive 2000/60/EEC, OJ L 327 of 22.12.2000) and Habitats and Urban Waste Water Treatment (HUWWT) (Directive 91/271/EEC, OJ L 135, 30.5.1991) directives which aim to both reduce the costs of water treatment and reduce eutrophication and its associated problems. The forthcoming implementation deadlines of these directives coupled with phosphorus being the limiting nutrient in freshwater systems (Smith *et al.* 1999), means that it is necessary to regulate the amount of phosphorus entering freshwater systems (rivers and lakes). In order to do this there is a refreshed need for quantification of changing phosphorus loads and subsequent changes in water quality to assess whether any mitigation measures (such as those introduced as part of the DEFRA (Department for Environment, Food and Rural Affairs) Catchment Sensitive Farming (CSF) program are effective and cost-effective. There are also other leglislative committments which the waters of the UK must meet; these include (1) The Nitrate Directive (91/676/EEC, (OJ L 375, 31.12.1991) which aims to reduce agricultural nitrate contributions to pollution and (2) The Habitats and Birds Directive (92/43/EEC, OJ L 206, 22.7.1992).

2.3 Phosphorus: the essential nutrient

Phosphorus (P) is the major limiting nutrient of primary productivity in many rivers and lakes in the UK, and therefore any increase in the loading of phosphorus to freshwater systems can significantly alter the nutrient balance and affect the composition and diversity of plants and creatures inhabiting the waters (Smith *et al.* 1999). It is an essential element in many of the compound synthesising processes upon which plant life depends (Brady & Weil 1999). Phosphorus is a foundation component of ATP (adenosine tri-phosphate) which drives most energy requiring bio-chemical processes as well as being an essential component of both RNA and DNA (key parts of genetic inheritence and protein synthesis) (Cartz *et al.* 1979). As a result, phosphate rich fertilisers are often applied to fields to ensure the crops have sufficient phosphorus available to them. However recent years have seen an over-application of these manures and fertilisers resulting in high levels of phosphorus accumulation in the soils in many parts of Europe, Asia and North America (Brady & Weil 1999).

Phosphorus inputs to freshwater systems can be classified as coming from either internal (cycling) or external (inputs) sources. Organic matter decay or bed sediments can contribute to in-stream phosphorus internally. External sources can be divided into point or non-point (diffuse) sources. Point source inputs come mainly from waste-water treatment works or industrial inputs. In recent years these sources have had much tighter controls on them, and as, such contributions to water pollution from waste-water treatment works are not as significant as they once were. The major source of diffuse phosphorus inputs in the UK is a result of the application of P rich fertilisers to crops and from livestock manure, coupled with the movement of sediment from landscape to river (providing a transport mechanism for nutrients).

Although rivers and streams are the principle routes for phosphorus transfer from landscape to lakes or the sea, they function as ecosystems rather than simple conduits (Melack 1995) which can store, release and transform nutrients. In research studies phosphorus is described in different forms depending on how the analysis takes place; for example the distinction between dissolved and particulate phosphorus is made based on the filter size used to separate the fractions. Within the two classes of total dissolved (TDP) and particulate (TPP) phosphorus there are further sub-divisions of organic and inorganic. Typically a stream stores only a small percentage of the nutrients which enter the catchment over a year, mainly as a result of organic debris accumulating phosphorus. The amount of retention is controlled by vegetation, discharge and temperature (Munn and Meyer 1990). Figure 2.2 shows how phosphporus is transported through the environment.

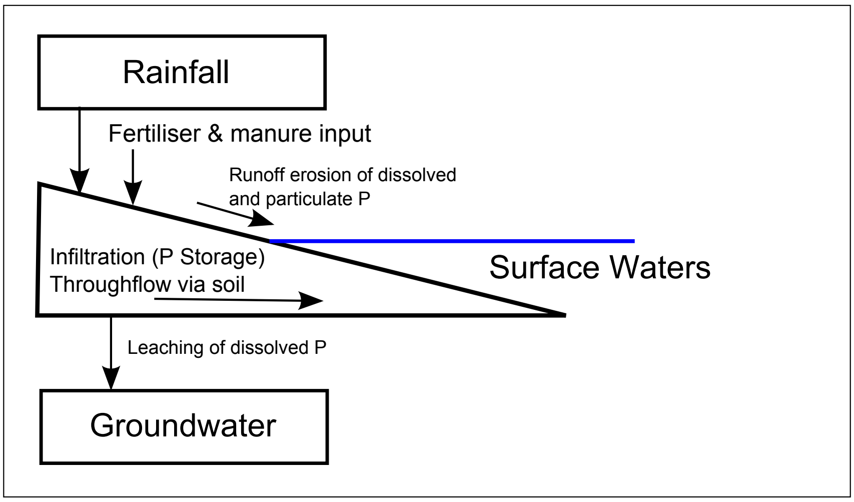

FIGURE 2.2: Schematic diagram of P transport and pathways

Phosphorus is transported predominantly with organic matter (Meybeck 1982). Soil erosion by rainfall and overland flow are widely recognised as the main pathways of phosphorus loss to channels (Haygarth *et al.* 2005). Agricultural activities such as soil structure degradation and drainage systems, as well as the obvious large areas of soil and additions of phosphate rich fertilisers, mean that diffuse pollution from agriculture is taken very seriously by government (Water-UK 2007).

2.4 Source-Mobilisation-Transport of Phosphorus

Figure 2.3 shows a conceptual model of how phosphorus can be viewed as a pollutant in terms of transport and sources.

2.5 Phosphorus: the pollutant

The nutrient enrichment of waterbodies has become a major environmental issue (Heathwaite *et al.* 2003, Withers & Haygarth 2007, Withers & Sharpley 2008). Phosphorus is widely regarded as the main contributing nutrient to eutrophication; the process of nutrient enrichment leading initially to a plant growth boom, followed by a degradation in water quality as excess material decays and rots in the waterbody leading to decreased dissolved oxygen levels. Eutrophication also causes algal blooms which in turn deoxygenate the water which results in fish and shellfish deaths. The water then becomes cloudy, colored a shade of green, yellow, brown, or red. Human society is impacted as

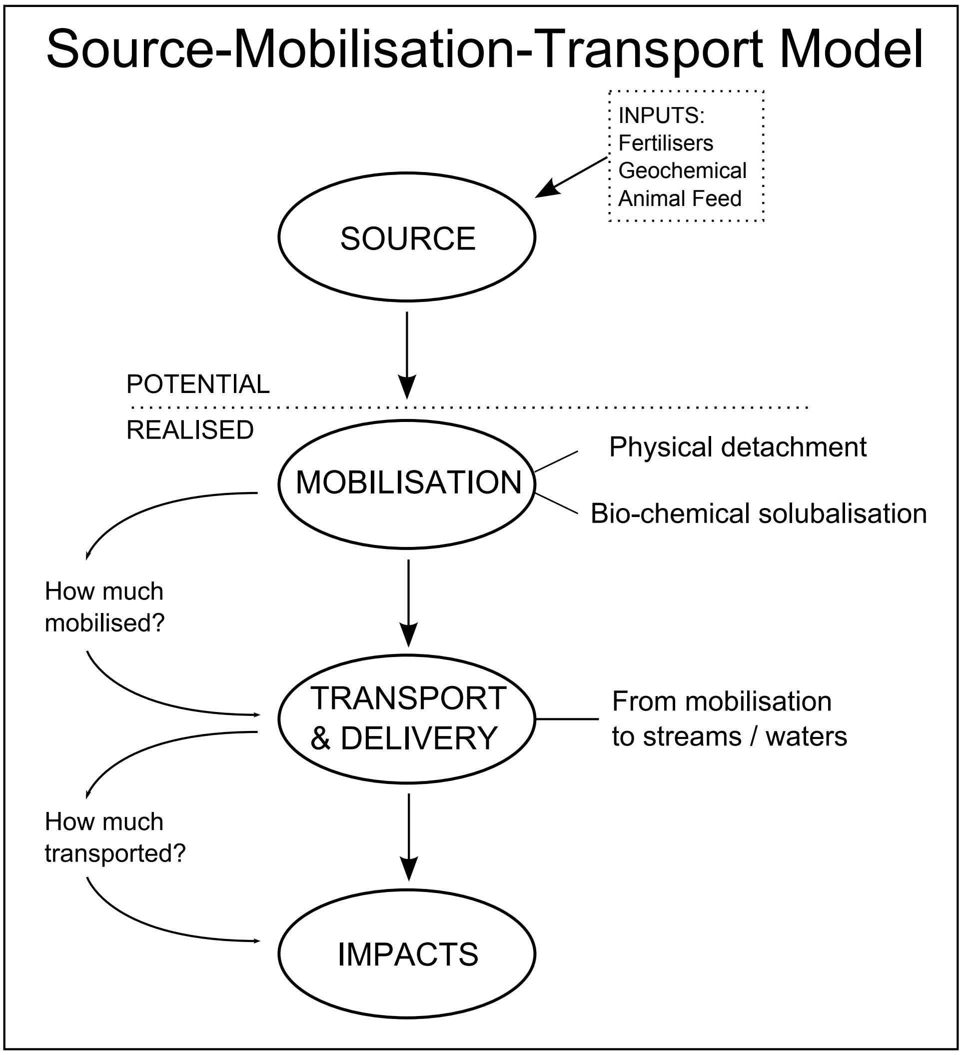

FIGURE 2.3: Conceptual model of phosphorus Source-Mobilisation-Transport

well: eutrophication decreases the resource value of rivers, lakes, and estuaries such that recreation, fishing, hunting, and aesthetic enjoyment are hindered (Bowes *et al.* 2005).

2.6 Agricultural sources of phoshphorus in the North of England

Since 1945 and the end of the Second World War there has been a shift in agricultural practices away from mixed farming systems towards intensive specialist farms which are loacted in the areas most suitable for them (Robinson & Suterland 2002). As a result of this, England can be roughly divided by region by its agricultural characteristics. For example the East of England is dominated by arable farming and the West is known for its intensive cattle farms. This shift has caused changes in pressures on water bodies (DEFRA 2002) including:

- Much greater stocking densities.
- Increased use of inorganic fertilisers especially in arable regions.
- Increased slurry use instead of straw based manures.
- Reduced crop rotation patterns leading to increased pesticide use.

The North West of England is characterised by naturally nutrient poor rivers and lakes which has resulted in plants and animals adapting to these conditions and thus being especially vulnerable to changes in nutrient levels or increased siltation (DEFRA 2002). It is for this reason that the River Eden, as studied in this research, is especially important in the UK in terms of diffuse pollution research.

2.7 How to deal with the phosphorus problem

In order to comply with the Water Framework Directive it is clear that action will need to be taken. DEFRA are approaching this problem with the Catchment Sensitive Farming (CSF) program. This is a initiative to promote "best practice" land mangement techniques to farmers to attempt to reduce the contribution of agriculture to water quality degradation. DEFRA describe the project as:

> "Catchment Sensitive Farming is land management that keeps diffuse emissions of pollutants to levels consistent with the ecological sensitivity and uses of rivers, groundwaters and other aquatic habitats, both in the immediate catchment and further downstream. It includes managing appropriately the use of fertilisers, manures and pesticides; promoting good soil structure and rain infiltration to avoid run-off and erosion; protecting watercourses from faecal contamination, sedimentation and pesticides; reducing stocking density; managing stock on farms to avoid compaction and poaching of land; and separating clean and dirty water on farms." (DEFRA Website Definition, Accessed December 2008)

Currently there are forty priority catchments within England and Wales in which the CSF program is being delivered. The mitigation activities which are being encouraged include:

1. Livestock Management
   (a) Keep stock tracks and paths well drained and away from watercourses.
   (b) Avoid overgrazing.
2. Yard management
   (a) Roof livestock holding areas to intercept rainfall from roofs prior to precipitation being allowed to flow across heavily polluted floor areas and transport high levels of pollutants to the watercourse (Hilton 2002). This technique is known to be highly effective when considering phosphorus as the pollutant (Dwyer *et al.* 2002).

A full breakdown of mitigation measures and suggested Best Management Plans (BMPs) for agricultural land and building areas is given in the report by Hilton hosted by English Nature (now known as "Natural England") (Hilton 2002).

It is clear that even in a small catchment it would be completely unfeasible to introduce these measures over the entire area of agricultural land. As well as costing money it is important to convince landowners and farmers that the measures being introduced are necessary. This identifies two needs of the ideal program delivery: (1) measures must only be introduced where they are necessary to avoid wasting money and (2) there must be a clear and simple way of identifying priority areas within a catchment to identify risky areas. It has been shown that the blanket targeting of entire catchments or even large reaches of river in order to improve water quality is neither cost-effective (Schleich et al 1996) or even successful in reducing the amount of pollutants entering the channel (Jokela et al 2004, Granlund et al 2005).

2.8 Costs of treating diffuse pollution

A recent estimate of the total external environmental costs of agriculture in the UK was between £141 and £300 million per year (Environment Agency 2007). The approximate annual costs of treating drinking water for pesticides are about £120 million; for phosphate and soil £55 million, for nitrate £16 million and for microorganisms £23 million. Monitoring water supplies and supplying advice on pesticides and nutrients costs around £11 million; off-site damage from soil erosion is put at £14 million (DEFRA 2002).

2.9 Conclusion

It is clear that phosphorus causes in-stream water quality degradation which has impacts both environmentally (i.e. eutrophication and associated problems) and socio-politically (forthcoming Water Framework Directive deadlines). If the costs of treating both diffuse pollution (Environment Agency 2007) and introducing mitigation measures (Hilton 2002) are considered then it is obvious that the key to successfully managing the problem is to be able to pinpoint within a catchment the areas which should be the focus of mitigation techniques and BMPs. This project will develop existing techniques and model frameworks and aims to highlight the areas of the River Eden catchment which should be the focus of future diffuse pollution mitigation work. The next chapter looks at the history of diffuse pollution modelling, how different models can be categorised and briefly introduces the model which this project uses as the base for the modelling framework.

Chapter 3

Diffuse Pollution Modelling

This chapter will look at the history of diffuse pollution models and their broad categories with some detail on specific models which have been developed with different approaches. It will then introduce what could be refered to as the "fourth generation" of diffuse pollution models, an example of which this project later uses for the River Eden catchment.

3.1 Introduction

Traditionally models of phosphorus transfers from soil to watercourses were developed for two main reasons; (1) to aid understanding of the processes associated with P transfer and (2) to enable quantitive predictions of phosphorus transfers (Krueger *et al.* 2007). More recently modelling approaches have been developed to attempt to mitigate against the impact of diffuse pollutants (particularly phosphorus) on rivers and streams. This is a result of: (1) the increased legislation concerning the status of freshwater in the UK and beyond (i.e. the Water Framework Directive) (Water Framework Directive 2000) and; (2) because of the increasing costs of treating diffuse pollution (Environment Agency 2007).

More recently the focus of modelling has shifted in an attempt to identify the Critical Source Areas (CSAs) (Heathwaite *et al.* 2005) in order to effectively target areas within a catchment likely to be responsible for causing the observed in-stream water quality problems. The SCIMAP model which will be introduced later in this chapter is an example of this approach, and the model framework which this project uses.

3.2 Empirical modelling of phosphorus transfer

The original phosporus transfer models were built on the concept of relating the characteristics of a site with the phosphorus loss from that site, usually on an annual basis (Krueger *et al.* 2007). The broad category of empirical models can be further sub-divided into the following groups:

1. Statistical Models.
2. Export Coefficient Models.
3. Multi-Layered Approach.

The following section will cover the properties and approaches used in these different types of models and detail some of their problems.

3.3 Statistical Models

Statistical models are the simplest form of empirical phosphorus transport model. They work by examining statistical relationships between the data available (i.e. land use, soil type and soil phosphorus loss) (e.g Daly *et al.* 2002, Andersen *et al.* 2005). This means that they cannot be transfered and applied between different catchments, or even applied to years which haven't been included in the statistical analysis, hence they have limited application for mitigation purposes.

3.4 Export Coefficient Models

The earliest attempts to model non-point source pollution began in the late 1970s and early 1980s with the development of a simple export-coefficient based approach. The early models were developed in the United States of America and adopted by the ODEC (Organization for Economic Co-operation) in their research into eutrophication in standing waters (Ormernik 1976, Vollenwieder *et al.* 1982).

The North-American model used one generalised export coefficient for all agricultural land, which when applied in the UK was found to be far too insensitive for the huge spatial heterogeneity in land use (stocking densities and farm management practices) characteristic of lowland farming in the United Kingdom. The UK has also seen marked changes in agriculture in the last 70 years, particularly after 1945 and also after joining the European Community in 1972 (Johnes 1996).

The development of the export-coefficient (E-C) approach (for reasons above) for the UK took place in the mid 1990s (Johnes *et al.* 1994, 1996) and aimed to use loading totals from above any given point in the catchment landscape to give a nutrient loading figure for that specific point. The basic approach to the model is as follows:

- Collect data on spatial distribution of land use (and therefore associated nutrient applications).
- Include other sources of nutrient input (nitrogen fixing, atmospheric deposition).
- Derive export coefficients from field data and literature to determine loss from each land use found in the catchment.

The model then uses discharge data at the given point to give mean annual concentrations of total N and P in the stream. The E-C approach does not differentiate between different species of phosphprous and thus means it is a more suitable indicator of change in loading from year to year (Johnes & Burt 1993, Johnes *et al.* 1994a, Heathwaite & Jones 1996).

In conclusion, the E-C approach to modelling of non point source pollution is very useful. It is simple enough to be coded using a standard spreadsheet software package, and its data inputs come from pre-existing monitoring databases. The grouping of individual nutrient species into two totals for N and P means many of the problems associated with the prediction of the unstable fractions and species are eliminated instantly (Johnes 1996).

The wide scale at which E-C modelling has been applied (from laboratory to plot scale) means that this approach can be tested in a uncertainty estimation framework to evaluate and assess the usefulness of the models (Murdoch *et al.* 2005, Khadam & Kaluarachchi 2006).

3.5 Multi-Layered Approach

Further increasing the complexity from simple statistical relationships via Export-Coefficient modelling led to the multi-layered approach to phosphorus transfer modelling (Krueger *et al.* 2007). The most well known example from this generation of models is the Phosphorus Index Tool (PIT) (Heathwaite *et al.* 2003).

The Phosphorus Indicator Tool was developed at Sheffield University and followed on from the previous generation of diffuse pollution models (the export-coefficient type).

An important part of the model is the "indicator" aspect; the aim of the work was to "identify appropriate factors determining P loss from agricultural land to surface waters' (Heathwaite *et al.* 2003). Indicators are used in this sense to "simplify information that can help reveal a complex phenomenon" (Heathwaite *et al.* 2003).

The project also aimed to develop a tool suitable for use regarding the implementation of the WFD to "demonstrate that UK policy on P loss from agriculture is rational' (Heathwaite *et al.* 2003). The tool makes use of proxies to act as indicators for P loss, for example using erosion from cultivated land as an indirect indicator of sediment movement and therefore P loss (Fraser *et al.* 1999). A key part of the PIT project is the transparency and modular approach to developing the model, enabling the tool to be adapted for slightly different tasks in the future (Heathwaite *et al.* 2003).

3.6 Physically based transfer modelling

Physically based phosphorus models include all aspects of the Source-Mobilisation-Transport (STM) model (Krueger *et al.* 2007) (Figure 2.3). Attempts are made to represent each process as accurately and fully as possible (Krueger *et al.* 2007) and are often broadly based on the equations first used in the procedural models from the 1980s (the ANSWERS model) (Beasley *et al.* 1980, Kreuger *et al.* 2007) and the CREAMS model (Kinsel 1980).

Physically based transfer models are usually designed for specific end goals and designed to be operated at the scale at which they were designed (Kreuger *et al.* 2007). This means that attempting to use such models for a different purpose, or at a different scale leads to problems; (1) current data measuring strategies mean that it is impossible to properly measure the model paramaters and inputs / outputs (Kavetski *et al.* 2003) and, (2) many of these parameters are not actually measurable (Beven 1989).

There are numerous examples of physically based phosphorus transfer models from various countries. A comprehensive review of several leading models can be found in Lewis & McGechan which covers the AMINO (Netherlands), GLEAMS and DAYCENT (USA) and MACRO (Sweden) models (Lewis & McGechan 2002). These models include many of the processes surrounding the complicated area of phosphorus mobilisation and transport from agricultural soils including soil evapo-transpiration, soil erosion, crop growth rates, soil carbon, soil nitrogen, applications of manures and slurries and soil temperatures (Lewis & McGechan 2002). Kavetski (2003) and Beven (1989) both discuss the problems associated with data availability for complicated physically based models (Kavetski *et al.* 2003, Beven 1989) and therefore it is suggested for this project, and indeed any work aiming to identify CSAs within river catchments, that they are not best suited. This is because they are data and parameter dependent, and there is often a lack of suitable data to resolve all the demands of the model.

3.7 Diffuse Pollution Modelling: The Next Generation

As discussed in the introduction to this research a critical concept in modern diffuse pollution is the notion of CSAs within catchments. These are important because we know that not all points in a landscape (even with the same landuse) contribute equally to observed water quality problems (Gburek *et al.* 2000, Heathwaite *et al.* 2000, Quinn 2004). Connectivity (i.e. a hydrological flow path) must combine with significant nutrient inputs in order for there be a problem (Lane *et al.* 2006). In order to identify these points in the landscape, and therefore be able to begin focused mitigation techniques against diffuse pollution, it is key to be able to model at catchment scale whilst capturing the much smaller scale (sub 10m) hydrological processes driving connectivity (Lane *et al.* 2006). This comes after research has suggested that nutrient delivery is dependent on extremely fine scale hydrology (Burt *et al.* 1999, Quinn 2004).

Two methods have been used in order to include the key aspect of nutrient delivery to watercourses in models; (1) combining physical transfer models with multi-scaled process complexity (e.g. Quinn 2004, Lane *et al.* 2006) and, (2) a risk based approach to identify the CSAs within a catchment (Lane *et al.* 2006). The first method was successful in transcending scale and was able to include open ditches (i.e. small scale) in a larger resolution framework (Dunn & Mackay 1996). However such models are still let down by their need for calibration and the lack of suitable data to undertake this.

The second approach takes a step back from complicated physical models which attempt to include all aspects of the Source-Mobilisation-Transport model and instead combines the risk of material production (i.e. soil erosion and therefore phosphorus transport) at different points in the landscape with the risk of those areas being hydrologically connected and therefore resulting in the threat of nutrients being transported to the watercourse. An example of this type of modelling is SCIMAP (Sensitive Catchment Integrated Modelling Analysis Platform) developed jointly at Durham and Lancaster Universities.

3.8 Conclusion

This research project aims to identify the areas in the River Eden catchment which should be targeted for mitigation methods to reduce the impacts of phosphorus pollution on water quality. After reviewing the broad categories of diffuse pollution model, the risk based approach seems an ideal choice as it includes a comprehensive treatment for hydrological connectivity and will provide an output map identifying potential problematic land units (Lane *et al.* 2006). Therefore the SCIMAP model will be used for this research.

Chapter 4

SCIMAP

This chapter looks at the core science behind the SCIMAP model which will be used to attempt to identify the key areas within the River Eden catchment which could be responsible for the degradation in water quality.

4.1 The SCIMAP model: classification

As discussed in Chapter 3, the SCIMAP model is one of the newer generation of diffuse pollution models described by Lane *et al.* (2006). These models combine the risk of diffuse pollution being generated at a particular point in the landscape with the risk of that area of land being hydrologically connected to rivers and streams, hence being a possible contributor to in-stream water quality problems (Lane *et al.* 2006).

4.2 Process Representation

Reaney *et al.* (In review), and the project website (SCIMAP) give a comprehensive overview of the SCIMAP modelling approach but a summary of the model development is given here for completeness, and as an introduction to the new inverse modelling work which features later in the document (Chapter 6). The following list summarises the key processes which are included in the SCIMAP modelling framework;

- Generation risk of material to be entrained (${p_i}^g$)
- Delivery index (connection probability) of the entrained material (${p_i}^c$)
- Locational risk (combination of ${p_i}^g$ and ${p_i}^c$ to give ${p_i}^{gc}$)

- Routing of location risk to give risk loading (Lj)
- Transformation of risk loading to risk concentration (Lc)

4.3 Risk Generation

SCIMAP focuses on risks to water quality which can be entrained as opposed to those which are dissolved (e.g. fine sediment). The risk generation parameter (${p_i}^g$) is calculated by:

$$\mathrm{p}_i{}^g = {p_i}^h \, . \, {p_i}^e$$

Where ${p_i}^h$ is the risk of there being sufficient energy to erode the material and ${p_i}^e$ is the risk of that material actually being erodible. Energy available for erosion is assumed to be a function of the area draining through a point in the landscape which enables water depth and therefore soil erosion potential (Ai) to be estimated and also the local slope β_i which is represented by stream power index (Ω_i); which is calculated by $\Omega_i = A_i$ tan β_i.

The risk of available energy is defined by scaling the stream power to give a hydrological risk of erosion combined with the topographic data layer (DEM) which is used to determine the upslope contributing area (UCA) and local slope (β_i).

In order to calculate ${p_i}^e$ (the risk of the material being erodible and therefore possessing a risk) there are two methods which are available; the first is to use expert logical judgement and set the values *a priori* after making basic assumptions about how different land uses are likely to affect erodability. The second is to use inverse modelling techniques to give a new set of land use weightings, tailored specifically for the organism or nutrient which the investigation is being focused on. SCIMAP has so far been applied with a set of logically assumed land use weightings (risks) and also after inverse modelling of salmonid fry data in the Eden catchment (Reaney *et al.* In review).

4.4 Risk delivery

The next step in the modelling process is concerned with the delivery of entrained material to the channel. Reaney *et al.* (2007) and Lane *et al.* (2009) describe a method of conceptualising a catchment's connectivity as a series of points, each one of which can be envisaged as having either a connected or disconnected state at any one time.

If the temporal scale in which this connected state is being considered is expanded then the frequency and length of the connected time period of each point in the landscape can be used to infer a distribution of connectivity. This distribution enables estimates of the amount of material that will reach the channel if combined with knowledge of how the temporal distribution is spatially structured around the catchment (Reaney *et al.* 2007, Lane *et al.* In review).

In the case of entrained material delivery then the spatial structure refers to the method of transport (predominantly overland flow) and thus means that if a point on the flow path does not generate overland flow then this point becomes the controlling location for all points upstream of this.

An established approach (Kirkby 1975, Beven & Kirkby 1979) for topographic wetness index (TWI) is used within the SCIMAP framework to factor in the propensity of each point in the landscape to generate overland flow. The Kirkby TWI expresses this propensity as a ratio of UCA per unit contour length draining through a single point in the landscape and the tangent of the slope and is assumed to represent the hydraulic gradient (Beven & Kirkby 1979). Lane *et al.* (2004) refer to the lowest value of the TWI along a flowpath as the "network index" (NI) (a measure of the propensity to vertical flow in a catchment) (Lane *et al.* 2004).

During rainfall events, a greater number of points in the landscape catchment will become connected (more runoff will be generated), the reverse happens as the catchment dries up after the rainfall event. To this end, a point with a higher network index value is likely to be connected for longer periods of time than those points with lower network index values (Reaney *et al.* In review).

SCIMAP assumes that the controlling factor in overland flow generation is the local topography and therefore there are two options for mapping the network index onto the duration of connection (Reaney 2008); firstly a probability density function of all the network index values within the catchment (using the rank NI value as a delivery index). Secondly the percentage of time that a point in the catchment landscape is generating overland flow can be used to assign this point a relative risk value. Both these methods have their disadvantages; the first method represents only relative network index values and does not consider the frequency or magnitude of connection periods. The second method as well as being more computationally intensive brings a return to the problems of calibrating models without sufficient data (Reaney *et al.* in review). A third method is used within the SCIMAP framework; it is assumed that there is a linear relationship when mapping NI to connection duration between the largest and smallest 5% of NI values (i.e. always connected and never connected values).

This connection probability is used as the delivery index (DI) for entrained material. Lane *et al.* (in review) have shown that this method performs admirably when compared to a distributed physically based hydrological model and contains significant information about the probability and duration of hydrological connection periods.

The next stage in the SCIMAP model is to calculate a location risk ${p_i}^{gc}$ for each point in the landscape.

$$\mathrm{p}_i{}^{gc} = {p_i}^g \cdot {p_i}^c$$

This at-a-point locational risk is assumed to be the sum of all upstream risks and thus a risk loading value can be calculated as:

$$L_j = \sum_{i=1}^{j} p_i^g \cdot p_i^c$$

Where j is the number of upslope contributing cells at that point. This calculation however does not factor in the differences in environmental impacts which could result from differences in UCAs (i.e. large loading values from a small UCA could have more serious environmental impacts than similar loading values from larger UCAs) (Reaney *et al.* in review). Each loading value is then scaled to the UCA to calculate risk loading per unit area (Cj):

$$C_j = \frac{\sum_{i=1}^{j} p_i^g \cdot p_i^c}{\sum_{i=1}^{j} a_i \cdot r_i}$$

Where a_i is the cell size and r_i is the rainfall weighting factor. UCAs are weighted according to upslope contributing rainfall based on the Met Office Long Term Average (LTA) datasets (Perry & Hollis 2004).

4.5 SCIMAP: Process flow diagram

Figure 4.1 shows the different stages involved in producing a risk map using the SCIMAP model framework. The three required data inputs (DEM, landcover and rainfall data) are shown at the top, and each of the boxes below represents a data layer which is computed by the model. Each of these layers is produced and saved as a readable file, which can be inspected using a GIS package. Thus, in the event of surprising results, or simply for extra analysis it is possible to interrogate each of these layers individually at any time.

4.6 SCIMAP: Current status of model

To date the SCIMAP model has been applied in the River Eden catchment with two datasets; (1) the original application of SCIMAP used expertly judged *a priori* values for erodability of different land uses within the catchment which were then combined with the connectivity analysis to generate risk maps and, (2) inverse modelling of salmonid fry populations in the catchment were used to generate a different set of land-use weightings (Reaney *et al.* in review). This research takes observed water quality data and uses inverse modelling techniques to pinpoint which parts of the catchment are likely to be responsible for the observed nutrient data. The next chapter examines the characteristics of the River Eden catchment which will be used for testing the modelling approaches developed in this research.

‘

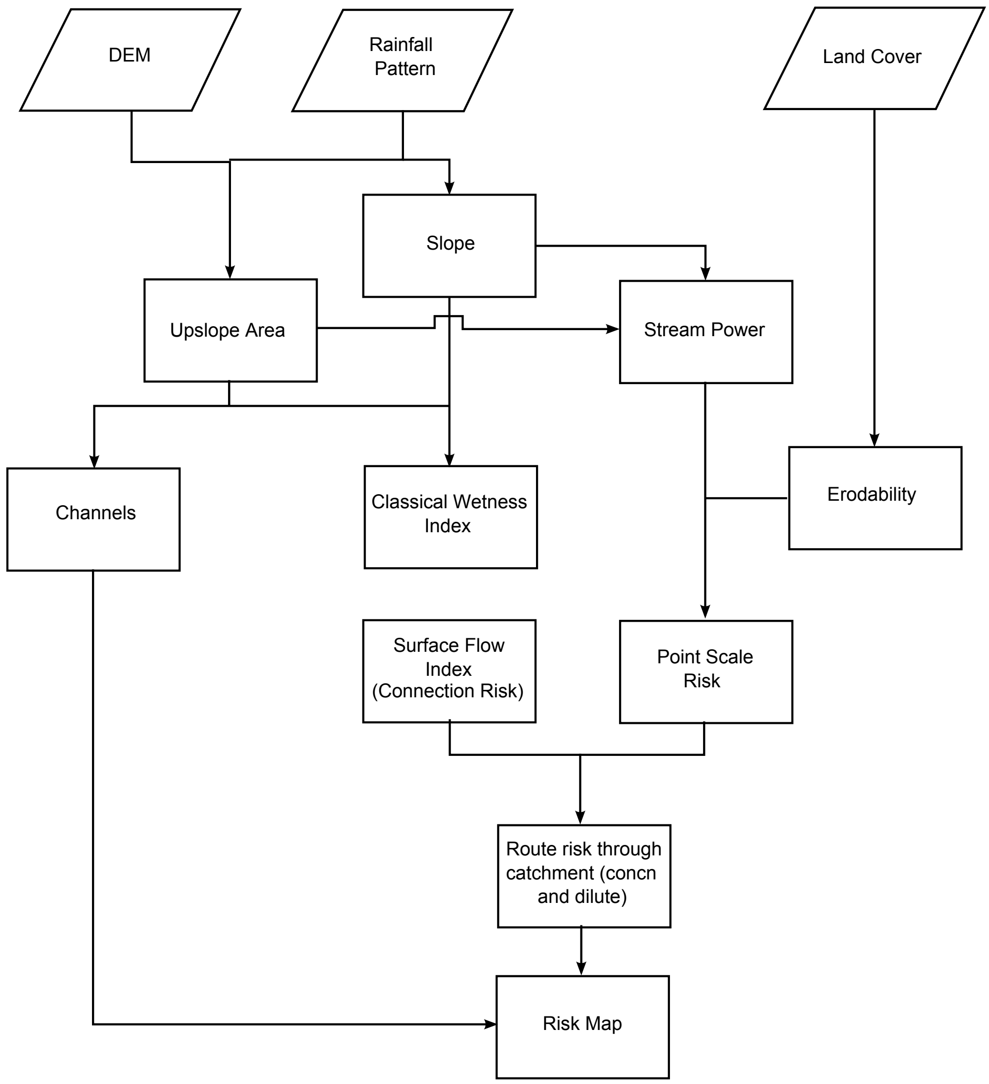

FIGURE 4.1: SCIMAP model stages (SCIMAP project website: http://www.scimap.org.uk)

Chapter 5

The Eden Catchment

This chapter looks at the characteristics of the River Eden catchment in the North of England and concludes that it is an ideal testing ground for the SCIMAP diffuse pollution modelling framework.

5.1 Overview and land use

The total area of the catchment is 2400km^2 and includes the major rivers of the Eden, Eamont, Irthing, Petteril and Caldew. The Eden catchment is predominantly rural with 95% of the total area classified as agricultural or rural, and only 1% classified as urban according to the CEH land cover map 2000. The total population of the catchment is approximately 240,000 in 2005 according to the Environment Agency Eden Catchment Flood Management Plan with the main population centres being Carlisle and Penrith. Figure 5.1 shows the boundaries of the River Eden catchment, the major rivers and the population centres within the catchment.

5.2 Topography

The upper reaches of the catchment are characterised by the steep slopes of Skiddaw and surrounding fells. The lower reaches towards Carlisle are characterised by wide, shallow valleys. The fells exceed 950m at the upstream extents of the catchment and remain high along the catchment boundary. Kirkby Stephen marks a change in the rivers character and the steep tributatries disappear and the river valley widens as it travels through Appleby and onto Carlisle.

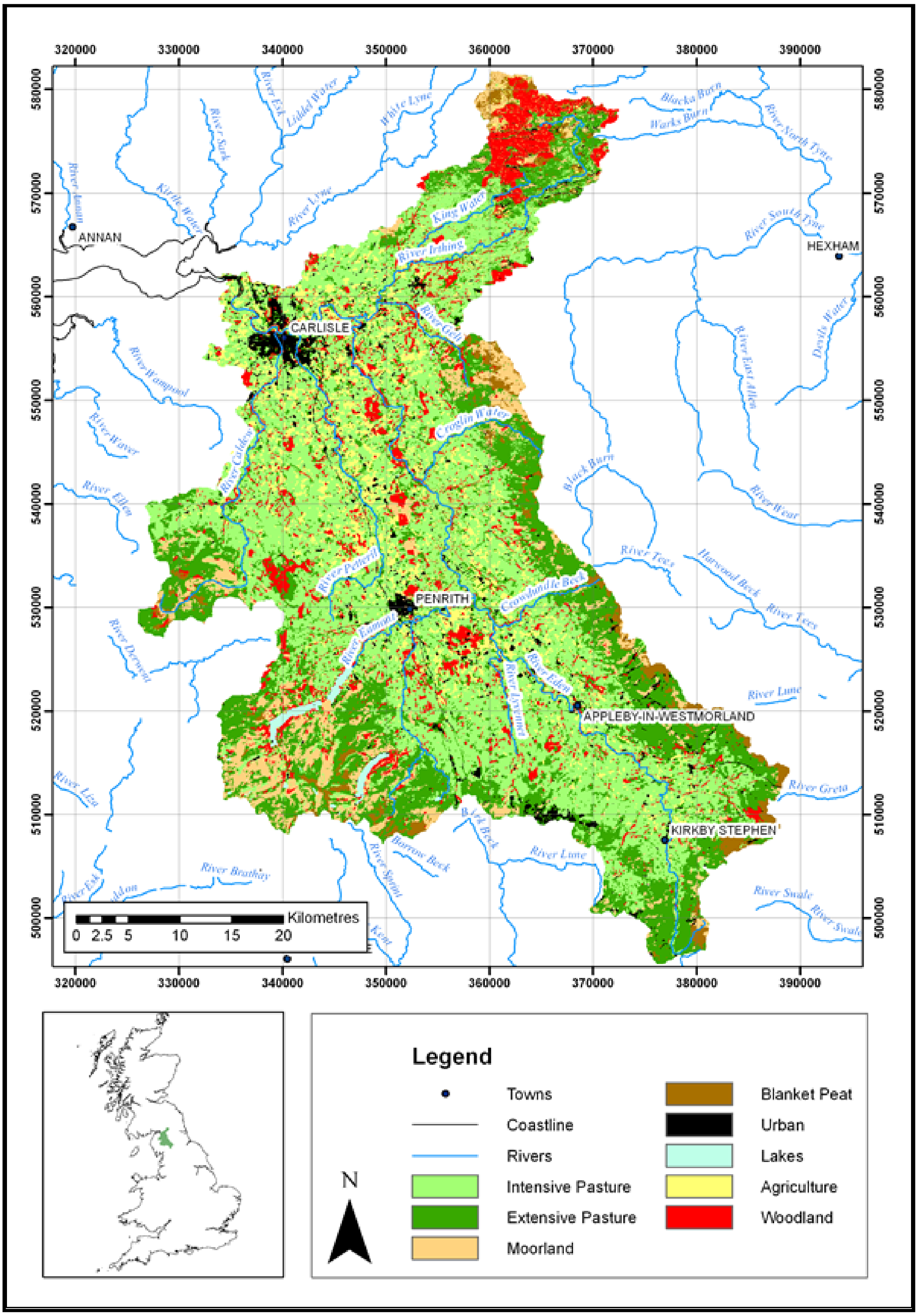

Figure 5.1: River Eden Catchment

5.3 Geology

At the upstream end of the catchment, the higher moorland areas are underlain predominantly by Millstone Grit. Between Kirkby Stephen to Carlisle, this changes with two main lithologies; the areas to the West of the main Eden channel is made up mainly of sandstone and mudstones. The areas to the East of the River Eden and also surrounding Carlisle consist mainly of Carboniferous limestones.

Geology plays an important part in the response of the catchment to rainfall events. In the south and east of the catchment there are numerous steep tributaries which are underlain by sandstone. This promotes rapid run-off and can cause the rivers to rise quickly after heavy rain. Downstream of Carlisle, the low relief is underlain by mudstones from various eras. The western part of the catchment is dominated by the metamorphics of the Borrowdale Volcanic Group and Skiddaw slates. This geology would normally promote rapid run-off, however the large lakes (Ullswater and Haweswater) dampen this potential and thus the contribution to catchment hydrology made by the lithology is not as large as it might otherwise be (Environment Agency 2008).

5.4 Catchment hydrology

In the upland parts of the catchment, such as the source of the Eamont at Hellvellyn, the average annual rainfall exceeds 2800mm. Around Carlisle this figure is much lower than than the rainfall recorded in the mountainous areas and averages 760mm per year. The combination of plentiful rainfall and rapid run-off caused by the topography and geology (especially between Kirkby Stephen and Carlisle) results in quick responses from the channels after heavy rainfall. The tributaries upstream of Kirkby Stephen are particuarly steep and their rapid run-off regimes can cause the main channel to rise quickly after heavy rain. Figure 5.2 shows the gradients of the main river channels within the Eden catchemt.

The geology and hydrogeology of the catchment combines with the topographic characteristics resulting in rivers with water levels that rise quickly after rainfall. In the upland tributaries upstream of Penrith in the Eamont catchment, average annual rainfall exceeds 2800mm on Helvellyn. Around Carlisle and on the coastal fringe, this is reduced to about 760mm. The average annual rainfall for England and Wales is 920mm (Environment Agency 2008).

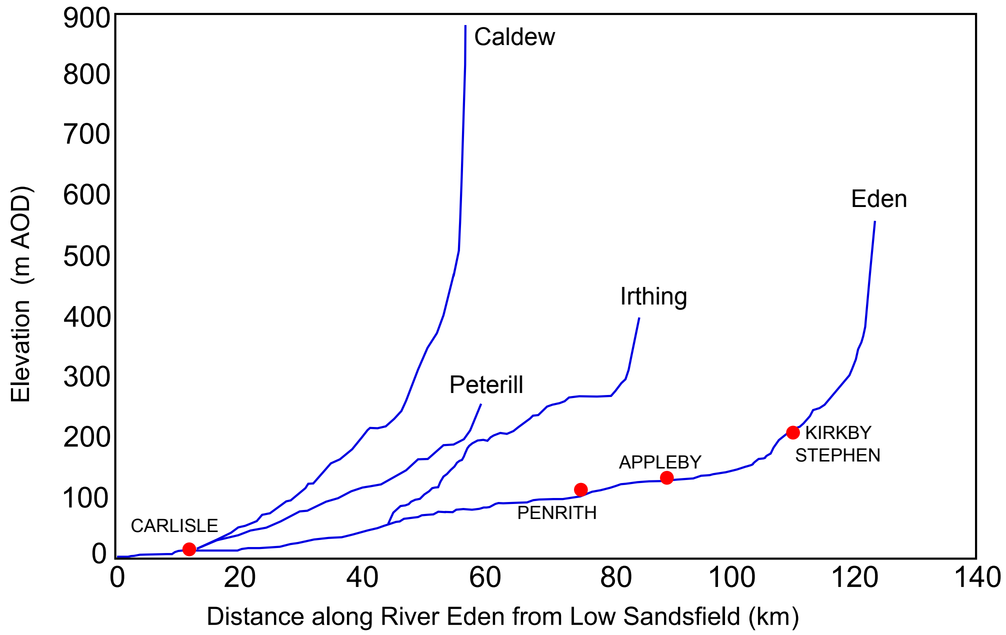

FIGURE 5.2: Gradients of main channels

5.5 Conclusion

The River Eden catchment was chosen for this research project for two reasons; (1) the SCIMAP modelling framework has already been applied to the catchment (Reaney *et al.* in review) and thus this research can be directly assessed against existing model outputs and, (2) the large spatial area of the catchment combined with the variety of land uses (in a largely rural environment) make it an ideal testing ground for the SCIMAP risk modelling approach for pin-pointing the sources of diffuse pollution within the catchment.

Chapter 6

Creating a dataset for inverse modelling

The SCIMAP framework offers two methods of assigning risk weightings to different land uses within a catchment; the first being a priori values based on logical assessment of the erodability potential of differing land uses, and the second by using a nutrient dataset to calculate weightings based on an inverse modelling approach. The end result is the same for both methods; a series of risk values ranging between 0 and 1 for the land uses within the catchment, which are used to drive the SCIMMAP model.

6.1 Previous research

Research has already been undertaken in the Eden catchment using inverse modelling (Reaney et al in review). This study used fish population data (salmonid fry) to infer the land uses which were responsible for the spatial distribution of the fry population within the catchment.

6.2 Background to research

Central to the operation of SCIMAP is the assumption that there is a spatially distributed land use signal which controls at-a-point water quality, whether related to sediments, nutrients or indeed both. However, upstream of any signal monitoring point will be a mosaic of land uses, each with differing levels of potential hydrological connectivity to the receiving waters, and each with differing levels of potential risk export.

Through time, this mosaic pattern will exert a complex impact upon the at-a-point time series of phosphorus concentrations and loadings.

The aim of the inverse modelling is to determine how to calculate the weights that each land use should be give, such that the SCIMAP model can adequately represent time-integrated phosphorus loadings and concentrations in the channel network.

Immediately, this defines the twin data needs of the inverse modelling; (1) the data must be representative of time integrated phosphorus concentrations and loadings; and (2) it must be spatially distributed, in order to best resolve the land use weightings that are most effective in discriminating the phosphorus characteristics at different locations within the catchment.

6.3 Data sources and collection methods

The Environment Agency (EA) operates a national monitoring scheme for rivers and canals known as the General Quality Assessment (GQA) scheme. Samples are collected from 7000 sites which cover approximately 40,000km of rivers and canals. The GQA scheme is designed to collect one sample per month however problems with data collection equipment mean that sometimes there are less than twelve per year collected. The sites where data are collected are chosen based on local conditions at different reaches of river and are often based on pressing ground characteristics such as where decisions need to be made on land use changes or abstraction programs. Phosphates in rivers are measured using the flow injection colorimetric method which is applied to unfiltered water samples and processed with ammonium molybdate and potassium antinomyl tartrate and ascorbic acid as the reducing agent. The results are given as concentration of orthophosphate in mgl^{-1}. The EA also collect extra data in specific surveys or after serious pollution incidents. However, the GQA database does not contain these extra measurements in order to try and avoid bias.

In order to use the discrete phosphorus data collected by the EA within the SCIMAP modelling framework, it is necessary to have a time integrated, spatially distributed dataset of phosphorus concentrations throughout the catchment. The GQA database provides excellent spatial distribution of data, however these data are not continuous.

Averaging these data has the potential to produce a result biased to the sampling period, which may not be entirely suitable for use in the SCIMAP framework. This is due to the sampling strategy of the GQA scheme, which is not designed to sample in such a way as to be representative of hydrological variability.

This research will use the GQA phosphorus dataset in two ways; (1) to drive an inverse model to attempt to produce a flow weighted dataset for SCIMAP; and (2) the average values for concentrations of phosphorus at each site will be used to drive the model (without statistical modelling or improvement).

6.4 Calculating load (traditional method)

Concentrations of phosphorus vary depending on the discharge in the channels within the basin. Simply averaging a series of concentration measurements is likely to lead to unrealistic averages if major P fluxes are missed. Therefore a discharge weighting must be factored in. The simplest equation for loading is:

$$\mathrm{L} = \mathrm{C} \times q$$

where L is the load in gs^{-1}, C is the concentration of nutrient (mgl^{-1}) and q is discharge (m^3s^{-1}).

If sufficient data are available then it is possible to use the instantaneous discharge and concentration measurements to calculate the loading value. If both flow and concentration are measured continuously then the loading estimate will be minimally affected by the changing discharge in the river over time. However if the sampling resolution is lower (even weekly) then the error involved with this approach can be extremely large (Johnes 2007). This is due to the highly episodic nature of phosphorus transport within rivers which is controlled by high flow conditions. Therefore the probability of missing a major nutrient (pollutant) flux event could be very high. Missing the event would happen if the timings of the measurements did not coincide with high flow events. For example Walling and Webb (1985) state that 60% of the overall sediment load was transported in 2% of the time in their review of the discharge of contaminants to the sea in the River Exe catchment. A similar study in the Cessnock catchment of Scotland reported that conventional spot sampling would have missed 87% of the phosphorus transfers that their designed nutrient study collected (CIWEM 2004).

Statistical regression models can be used as an alternative basis of loading calculation where there are not sufficient data for the direct estimation technique to be applied (e.g. Asselman 2000). An empirical relationship - a rating curve - is derived from the temporal point samples where both flow and concentration are recorded and used to predict the concentrations where only flow was recorded. A load prediction can then be calculated by using the direct estimation method with the predicted data for concentration and the measured data for the discharge.

However, as with all modelling techniques, caution must be applied to ensure that accuracy is improved and not reduced with use of statistical techniques. Firstly and perhaps most importantly is the choice of regression model. A poorly fitting model (i.e. linear model on non-linear relationship) could produce a severely biased load estimate (e.g. Ferguson 1986). Where flow is the dependent variable it is important that the model is tested for extreme highs and lows to see whether the predicted concentrations produce anomalously high or low concentration estimates.

In order to achieve maximum accuracy with this modelling an amended version of the Tarras-Wahlberg & Lane (2003) simulation will be used. Using a Monte-Carlo based simulation should prevent the issues of unknown bias which arise when fitting rating curves between discharge and a variable (e.g. sediment, nutrient concenrtation) in log-log space.

6.5 The Tarras-Wahlberg and Lane method

Tarras-Wahlberg and Lane (2003) applied a novel combination of traditional rating curves with a Monte-Carlo simulation to estimate suspended sediment yields within the Puyango river basin in Ecuador in their study of transport of contaminated sediments as a result of mining operations in the basin. This approach was necessary because of the lack of frequent and good quality data on sediment yields; a similar problem faced in the UK as a result of the monthly measurements of nutrients in the GQA dataset.

The term "Monte-Carlo method or model' refers to a large ranging and widely used variety of approaches to mathematical and physical modelling. The technique can be broadly summarised as follows:

- Define a domain of possible inputs.
- Generate inputs randomly from the domain, and perform a deterministic computation on them.
- Aggregate the results of the individual computations into the final result.

In the Ecuador research, each of the discharge measurements were much more frequent than the concentrations of suspended sediments to LOWESS (also known as locally weighted scatter plot smoothing) rating curves were applied to each daily discharge in log space giving an estimated value of $ln(C)$. The normal distribution of the residuals allowed the standard deviation (SD) for each estimated $ln(C)$ to be calculated.

A Monte-Carlo simulation (250 iterations) was used to simulate $ln(C)$ values for each day based on the recorded discharge; this enabled the uncertainty to be propagated through to the annual estimates of suspended sediment yield and also holds the advantage that no correction factors need to be applied when the $ln(C)$ value is converted into real space, because the Monte-Carlo simulation takes place in log-space (Tarras-Wahlberg & Lane 2003).

Figure 6.1 summarises the Monte-Carlo methodology used in the Ecuador study;

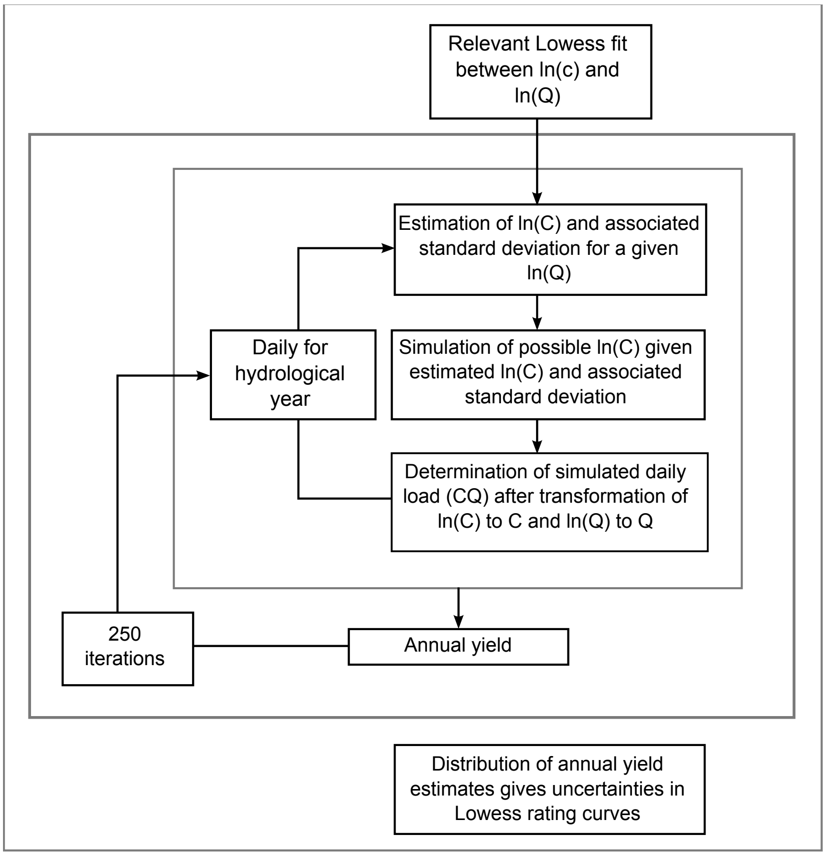

FIGURE 6.1: Monte Carlo method used in Puyango basin study (Tarras-Wahlberg and Lane 2003)

- Measured field data was used to establish relationships between suspended sediment concentration (C) and discharge (Q) for the Puyango basin were established with the LOWESS (locally weighted scatterplot smoothing) method used as the framework for the relationship. This method was chosen in the Ecuador research because it (1) makes no assumptions as to the form of the relationship (Lane *et al.* 2003) and (2) is ideal for use in a non-linear relationship (Hicks *et al.* 2000).

- The established relationships between lnQ and lnC enabled an estimate of lnC (and the associated SD) to be made for any value of lnQ which in turn allowed the simulated daily load to be calculated (based on 250 iterations of sampling lnC).
- This process was repeated for the hydrological year and resulted in a simulated annual suspended sediment yield.
- A range of annual yields could then be produced which took into acount the uncertainty associated with the LOWESS method (Lane *et al.* 2003).

6.6 Data sources for Monte-Carlo simulation in the Eden Catchment

The data requirements for the loading calculations are two fold; firstly a series of phosphorus concentration measurements distributed around the catchment with the date, time and concentration in mgl^{-1} for each site and secondly a set of discharge measurements which can be used to match the phosphorus data consisting of date, time and discharge in mgl^{-1}. In order to calculate discharge weighted phosphorus values (either loads or concentrations) it is necessary to have discharge data for the same point in the river.

In an ideal situation both measurements would be taken from exactly the same point in the channel, however in reality this is not the case. In the Eden catchment the majority of the discharge measuring stations are located on the main channels of the Eden, Eamont, Petteril, Irthing and Caldew whereas there are many nutrient measuring points located on the smaller channels and tributaries. This immediately presents a problem of whether the data from those P stations can be used in the simulation because of a lack of suitable associated discharge data. The map below (Figure 6.2) details a section of the upper part of the catchment showing the difference in concentration between sites where nutrient data are available, and those where discharge data are available.

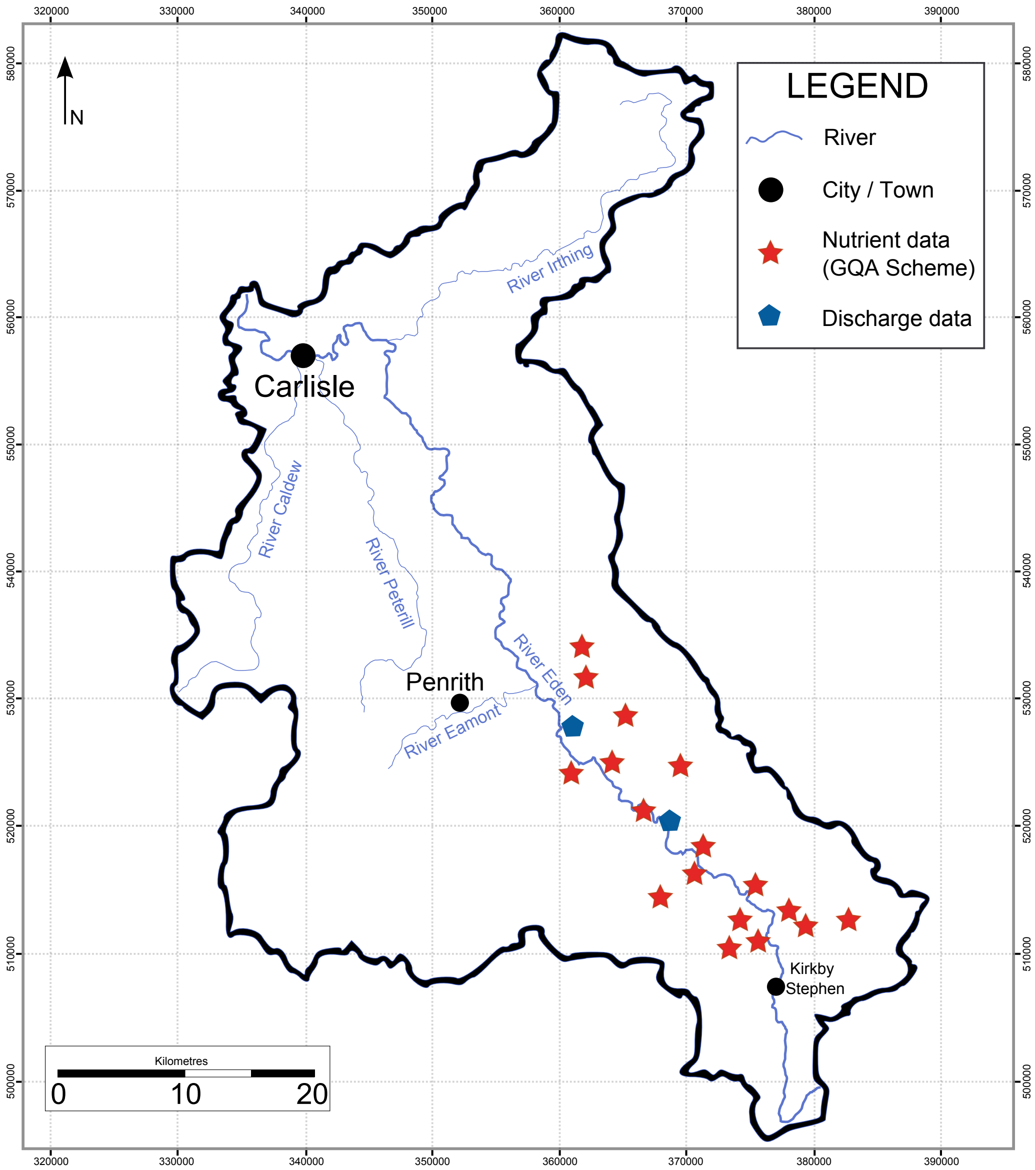

FIGURE 6.2: Comparison of nutrient and discharge data availability

The following sites were selected for testing with the Monte-Carlo simulation because of their proximity to available discharge data: (See Figure 6.3 for their locations within the catchment).

1. River Eden upstream of Kirkby Stephen.
2. Swindale Beck at Hall Garth.
3. River Belah at Belah Bridge.
4. Scandal Beck at Soulby.

5. Brampton Beck D/S Brampton.
6. River Irthing at Newby East.
7. River Eden at Warwick Bridge.
8. Eden at Beaumont.
9. River Eden at Sheepmount.
10. River Eden at Eden Bridge.

Figure 6.3 shows the locations of the points in the catchment where the Monte Carlo simulation was applied. (The numbers on the map relate to the numbered list of locations above).

This low number of sites was used because of the lack of availability of suitable discharge data for using on the multiple points where nutrient data is available which were located on smaller tributaries. When these sites were tested with the Monte-Carlo simulation they often produced negative values for estimated concentration. Although it was clear early in the study, that the small number of suitable sites would mean that this dataset would not be appropriate for using to drive an inverse model across the entire Eden catchment, it was decided to use the simulation for the ten sites where good data was available. By assessing the technique here, it is available for use in the future when more discharge data are available.

The next task in the simulation methodology was to extract the data from the water quality (GQA) database into a useable form for the modelling. The database is originally in the Microsoft Access format and so firstly all the records for the Eden catchment were exported into a CSV (comma separated value) file. In the interests of platform compatibility and not being restrained by specific software, a MySQL database was created and the data file was uploaded into this database.

MySQL was chosen because as well as being licensed as Open Source, it interacts very well with the PHP scripting language and thus the data can be accessed, manipulated and exported by PHP driven web-pages, meaning that the system can easily be adapted and used for other catchments in the future. To access the data required for the inverse modelling work a series of dynamically driven PHP based web pages were constructed which populate a drop down box with the names of all the sites for which data is available in the catchment. Upon selecting one of the locations a query is executed on the database which extracts all the available records for that site, selects the required columns (date, time, and value in mgl^{-1}) and exports them as a CSV file for use in modelling. This

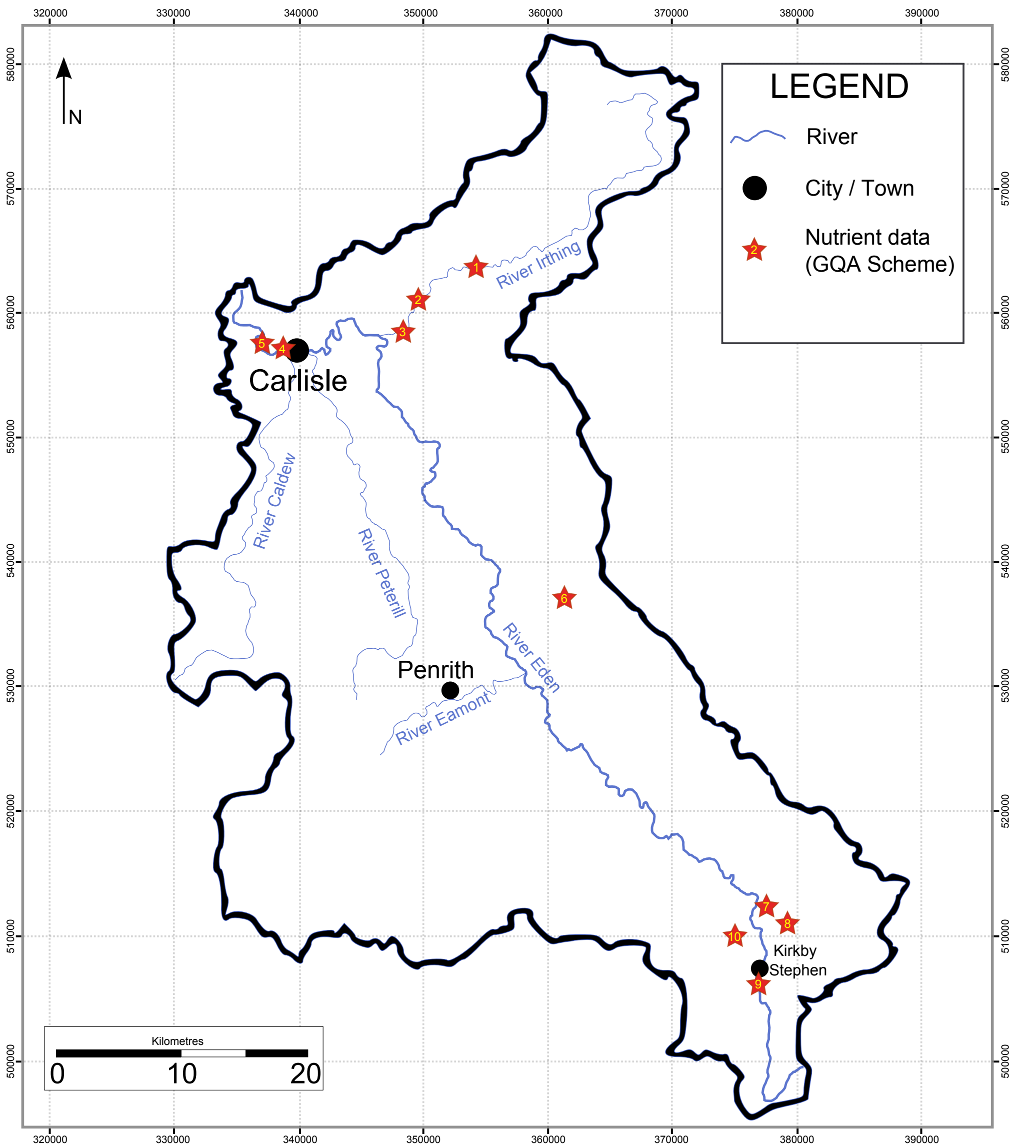

FIGURE 6.3: Locations used in the Monte Carlo simulation

technique facilitates easy access to a dataset which was once both locked to a platform, and proprietary software, a situation not conducive to modern scientific research.

The discharge data were also supplied by the EA and were prepared by Ian Pattison at Durham University as part of an ongoing research project into high and low flows and flooding in the River Eden catchment. The data are collected at 15 minute periods, although there were some gaps in the time series where faults with monitoring equipment, or communication breakdown resulted in missing data.

To correct these and get a complete record some interpolation was needed. A Visual-Basic macro was written which first dealt with creating the missing time slots by way of a loop which examined the time between each value. If this was not 15 minutes then a row was inserted and the time filled in (i.e. 15 minutes after the preceding value). The loop was broken when the next time in the dataset was the correct value. The macro then looked at the two known values either side of the missing values and averaged them to enable a simple interpolation to fill in the gaps.

6.7 Flow weighting methodology

The technique applied in the Eden catchment essentially takes the same form as the Tarras-Wahlberg & Lane (2003) method with some modifications. Firstly the LOWESS regression method was not considered suitable because of the lack of data available; instead a log-log curve was fitted to the phosphorus concentration and discharge data and then the prediction uncertainty around that curve was used to repeatedly sample for possible values of phosphorus concentrations for a given (measured) discharge value (i.e. the Monte-Carlo method). Figure 6.4 shows the stages in the Monte-Carlo simulation used in this study.

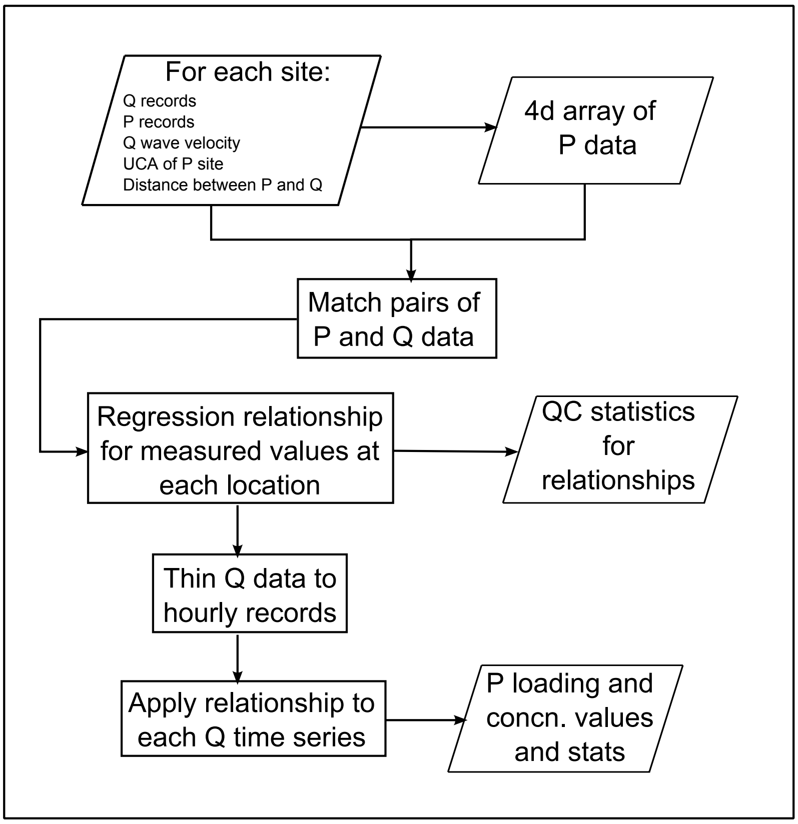

FIGURE 6.4: Monte-Carlo method used in the study

Table 6.1: Measured versus simulated means (P)

Site	**Measured mean P concn.**	**Simulated mean P concn.**
Eden U/S Kirkby Stephen	0.016	0.003
Swindale Beck at Hallgarth	0.016	0.003
Belah at Belah Bridge	0.016	0.003
Brampton Beck D/S Brampton	0.016	0.003
Irthing at Newby East	0.016	0.003
Eden at Warwick Bridge	0.016	0.003
Eden at Beaumont	0.016	0.003
Eden at Sheepmount	0.016	0.003
Eden at Edenbridge	0.016	0.003

The resulting range of estimates for phosphorus at that given discharge value were then either averaged for a mean concentration value or multiplied by the discharge and subsequently averaged to give a mean load and standard deviation. This all takes place in log space thus avoiding the problems of traditional rating curve methods of not knowing the exponent that Q is raised to *a priori* which leads to a biased rating curve.

A key part of the simulation is determining the number of iterations needed to obtain the most accurate estimates of phosphorus loads and concentrations. The number of iterations is a parameter which is set in the MATLAB code before the simulation is run. The MATLAB code was modified to write the values of estimated phosphorus load of each iteration into a variable for 1-100 runs in order to assess the numer of iterations which were needed.

The residuals of the means for the four tested sites were calculated and plotted (Figure 6.5) to show the change in variation of the estimated loads as the number of iterations increased. Even though the variations shown in the early stages of the plot (i.e. from 1 to 40) are negligible there is no discernable variation after forty iterations. As the processing time and computing power needed to run forty simulations is also perfectly feasible on a reasonably powered Desktop PC there is no reason to run at less than this number of iterations, and certainly no reason to perform the simulation more times as there would be no discernable improvement in data output.

6.8 Results of Monte-Carlo simulation

The results of applying the Monte-Carlo simulation to the ten sites where suitable data are available are shown in Table 6.1. The sites can be referenced on the map in figure 6.3.

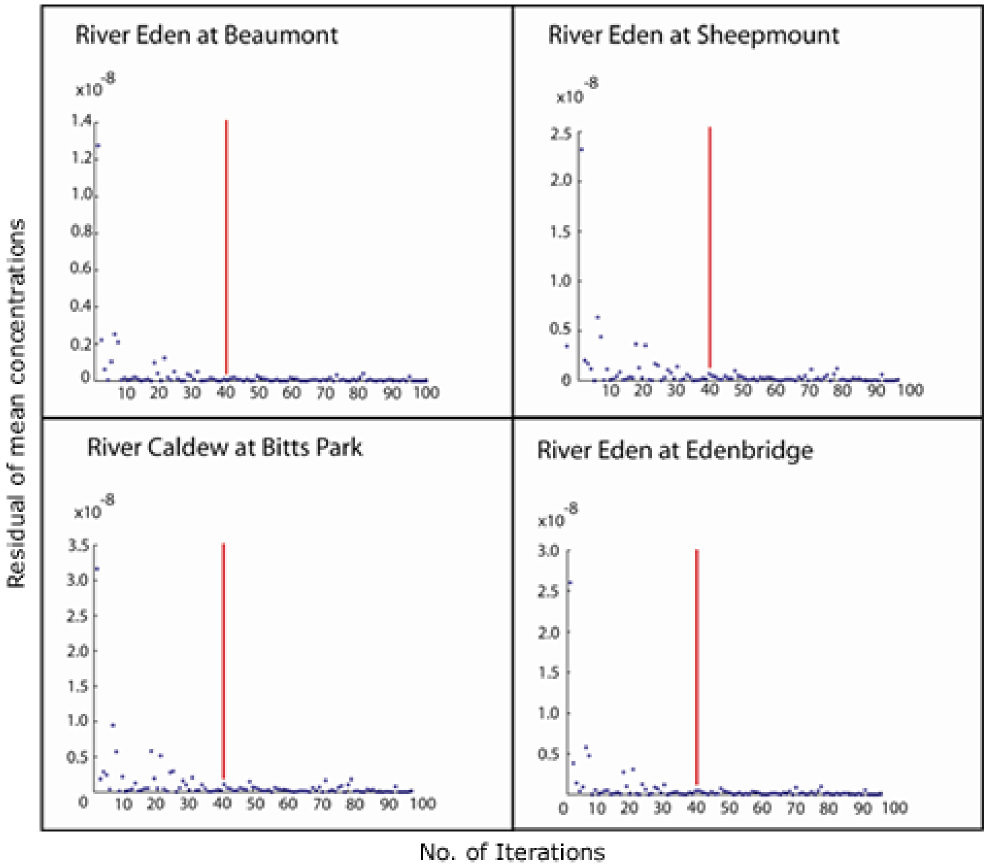

FIGURE 6.5: Residuals of means (testing the Monte Carlo simulation)

With such a small sample size, it is difficult, and perhaps not very useful to over analyse these results. The issue of how phosphorus behaves during and after rainfall events is also a pertinent factor here. Research has shown that high rainfall intensity storm events (¿9 mm per hour) can account for the majority of annual phosphorus loss from arable land (Fraser *et al.* 1999). What can be done however, is to look at the technique and potential future application. Tarras-Wahlberg and Lane used a similar technique (one which formed the basis for this study) to great success in the Puyango basin in Ecuador, and there is no reason why this simulation (now developed and adapted for nutrients) can not be used in the future, should the data availability issue be overcome. In the future, if this technique is to be developed and continued, the issue of ensuring that the raw data captures any such major storm events will be important.

6.9 Data Availability

Spatial distribution: The GQA database provides excellent spatial coverage of measuring sites within the catchment (see Figure 6.2 for an example in the upper part of the catchment where this is highlighted). As discussed in Section 6.4, the two data points for C and q would be in exactly the same place, both temporally and spatially. However, the current available discharge dataset has poor spatial coverage of the catchment, altough it is at an excellent resolution for the modelling.

UCA ratios: The Monte-Carlo simulation takes the lack of spatial correlation between the q and C measurements into account by using the relationships between the upslope contributing area (UCA) for the data pairs. The ratio between the two UCA values is used as a conversion factor when estimating the exponent that Q needs to be raised to. The simulation was tested on data where there were multiple nutrient data points on smaller tributaries to the main channels in the catchment. The UCA ratio is very small in these cases, which can lead to the simulation calculating negative values for the concentrations. This returns to the problem of a lack of discharge data.

6.10 Conclusion

Unfortunately there was a lack of suitable discharge data to apply the Monte-Carlo simulation across the whole of the River Eden catchment and make use of the nutrient data set which has excellent spatial coverage of the whole catchment. However, in the interests of applying the technique in the future, and also to assess the quality of the raw GQA measured data it was decided to use the Monte-Carlo simulation on ten suitable sites.

The simulation produces reasonable results for the sites where it was possible to test it, and some issues with the process were highlighted. Thus, should more discharge data become available, this technique could potentially provide a valuble method of improving theEA GQA dataset.

In the absence of a suitably spatially distributed flow weighted dataset the next chapter details how the GQA dataset was used to drive an inverse model to determine which land uses within the River Eden catchment are responsible for the observed in-stream nutrient levels.

Chapter 7

Application of SCIMAP in the River Eden catchment

This chapter looks in more detail at the data sources (and associated processing) which are required for the SCIMAP risk modelling framework. It also details and examines two different SCIMAP model runs: (1) using logical erodability weights (the a priori expert judgment method), and (2) using land use weightings obtained from inverse modelling of measured water quality data in the form of phosphorus measurements from the EA GQA database.

The SCIMAP risk modelling framework requires the following data sources: (1) topographic data, (2) landcover data and (3) rainfall intensity data. One of the underlying ethoses of the SCIMAP project is that the data must be readily available at a national scale so that the framework can be distributed and applied in different catchments across the UK without the need for extra costly, time intensive and unrealistic data collection programs to be implemented before the mapping framework can be used.

7.1 Data sources and pre-processing

The topographic data used are the Interferometric Synthetic Aperture Radar data produced by InterMap (more commonly known as NEXTMAP data). This is supplied at a resolution of 5m with a stated vertical precision of +/- 1m. The InterMap DTM (Digital Terrain Model) product was selected for use in the research, as opposed to the digital surface model (DSM) product they supply. The DTM processing has a tendency to smooth topography, but it is a much more suited product for lowland areas (Milledge *et al.* 2009). This is essential in a diffuse pollution investigation in the River Eden, as these are the areas where agriculture is likely to be concentrated, and thus a likely major contributing area to diffuse pollution.

Several different processing steps are required before this topography data can be used. It is necessary to calculate the upslope contributing areas (UCA) of each point in the landscape so that the diffuse pollution can be diluted: to do this the pits and depressions in the DEM need to be filled in order to enable both the UCA calculations and also flow routing to take place. The Planchon-Darboux (2001) algorithm is used for this. An important note to add is that the processed DEM (i.e. pits filled) is only used for UCA and flow routing and not for other variables which are calculated during the model run (such as slope). This important differentiation means that the role of depressions and pits in the landscape as water storage mechanisms can be included whilst ensuring that their UCAs and flowpaths are accurate (Reaney *et al.* in review).

It is also necessary to clip the DEM tiles to the catchment boundary to ensure that streams and channels which are not part of the River Eden catchment do not remain in the topography grid. If they remain in the grid then a risk of 0 is often diluted back into the main river catchment and gives erroneous results. To ensure this does not happen a 50m buffer was added to the catchment boundary layer in ESRI ArcMap and the DEM clipped to this.

The landcover data were supplied by the Centre for Ecology and Hydrology (CEH) in the form of the Landcover Map of Great Britain (2000). While this dataset has its problems (the main one being it is now eight years old) and most likely does not accurately represent agricultural land use (which of course is the main scope of the SCIMAP risk mapping framework), the lack of a suitable alternative means the LCM2000 is used in the model. The data comes at a spatial scale of 30m and is resampled to the same scale as the model is to be run at (10m in the case of the River Eden catchment) using the nearest neighbour algorithm.

The rainfall dataset is taken from the UK Meteorological Office long term average (LTA) annual dataset (Perry and Hollis 2004). Again the nearest neighbour algorithm is used to interpolate the coarser scale which it is supplied at to the finer model run scale.

7.2 SCIMAP computing: hardware / software

The SCIMAP risk mapping framework is coded in the C++ language and takes the form of a custom module in the SAGA (System for Automated Geographical Analysis) GIS framework (http://www.saga-gis.org). SAGA GIS was chosen for several reasons: (1) it has excellent grid handling capabilities, (2) it comes with key hydrological processing tools (e.g. Planchon-Darboux 2001) which are used in the pre-processing of the topographic data and (3) it is open source and therefore free to use making it an extremely viable option in any large scale future rollout of the SCIMAP risk mapping tool across the UK. The ESRI ArcGIS suite is also used to manipulate and process the data for displaying and outputting in the form of maps, although there is no reason why other GIS software (including open source alternatives) could not be used for this stage.

7.3 Modelling procedures

The first stage in applying the SCIMAP mapping framework is to prepare the input datasets. As discussed earlier the initial stage is to resample the topography, landcover and rainfall data to the same cell size, in the case of the River Eden catchment this is 10m as a balance of being at a fine enough resolution to represent hydrological processes whilst remaining computationally realistic with the hardware and software available.

Each of these data layers is then converted to ASCII file format using the inbuilt tools in ESRI ArcMap. This is the format required by SAGA GIS. The three input files are then loaded into the SAGA GIS software (the SCIMAP tool is a bespoke module for this software). The different landuses in the LCM2000 layer are classified by a three digit number which is assigned by CEH. The SCIMAP tool requires these landuses to be reclassed into a risk weighting of between 0 and 1. In the first model run logical erodability weights were assigned based on the expert judgment values previously used in the River Eden (Reaney *et al.* in review). For example urban areas are given a weighting of 0.01 and improved pasture is given a value of 0.3.

The SCIMAP module within SAGA GIS tool then takes the three input layers and creates new layers by analysing various topographical features and multiplying layers together.

The different stages involved in computing the final risk maps are detailed in Figure 7.1 (Reaney *et al.* in review).

7.4 Using inverse modelling within the SCIMAP framework

As discussed in Chapter 6 the lack of suitable discharge data meant that undertaking Monte-Carlo simulations to give estimates of flow weighted nutrient concentrations over

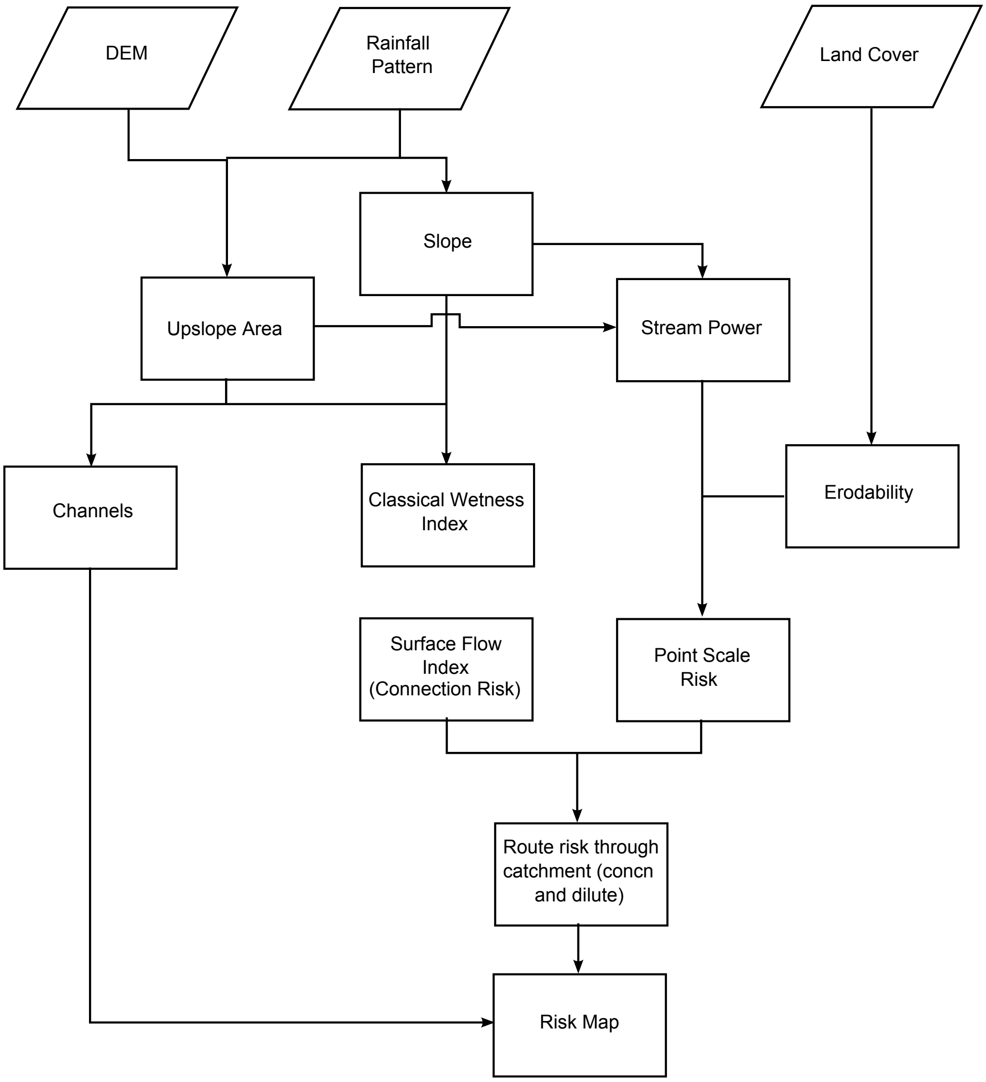

FIGURE 7.1: SCIMAP model stages (SCIMAP project website: http://www.scimap.org.uk)

the entire catchment was not possible. Despite the potential problems of the raw observed GQA data not being temporally rich enough to suitably capture major nutrient fluxes it was decided to use this dataset and compare it to the results obtained when using the logical erodability weights.

7.5 What is inverse modelling?

Inverse modelling is a technique which is commonly used in the physical sciences to estimate quantities that are directly or indirectly related to the measured quantity (Tarantola 2005). It is commonly used in climate science research as a tool to estimate the concentration of atmospheric trace gases from the absorption features of the atmosphere (Huang *et al.* 2008, Meirinik *et al.* 2008).

Lane (2008) suggests that inverse modelling (IM) may have an important role to play in future hydrological models. In his text "What makes a fish (hydrologically) happy? A case for inverse modelling" he suggests that by using IM techniques we can allow the organisms to tell us which processes matter to them and thus use observed data to conceptualise and also validate the process cascade in a new way (Lane 2008). Inverse modelling provides an ideal technique for the case of modelling diffuse pollution in the River Eden as it allows work to be carried out with the data which is available, rather than attempting to find enough suitable data for a more complicated model.

7.6 Applying an inverse modelling technique to phosphorus data

The aim of the inverse modelling process for use in the SCIMAP mapping framework is to use different randomly assigned land use weightings and determine which best match the observed water quality data in the River Eden. The technique follows the simplified version of the one used in the study of salmonid fry populations and diffuse pollution in the River Eden (Reaney *et al.* in review).

The diagram below (Figure 7.2) shows the steps involved in the inverse modelling. The measured chemistry data is included in the form of a mean value averaged over the entire duration of the sampling period from the GQA database which is added to the X-Y location of the sampling site. All this inverse modelling was undertaken in The Math Work MATLAB software package. Firstly each different landuse unit within the catchment area is assigned a random risk weighting (between 0 and 1) (the random landcover grid). This layer is added to the connection grid to create a combined landcover and connection grid. This is done for 30,000 different random land use weightings.

The risk concentration value (i.e. after the combined risk and connection grid has been routed and diluted according to the connection grid) can then be calculated for each point in the catchment. The correlation between these points, and those where there is measured chemistry data (i.e. all the sites where mean phosphorus concentration has been sampled is then plotted for different land uses within the catchment. The resulting plots then show the contribution that different land uses within the catchment have made to the observed water quality.

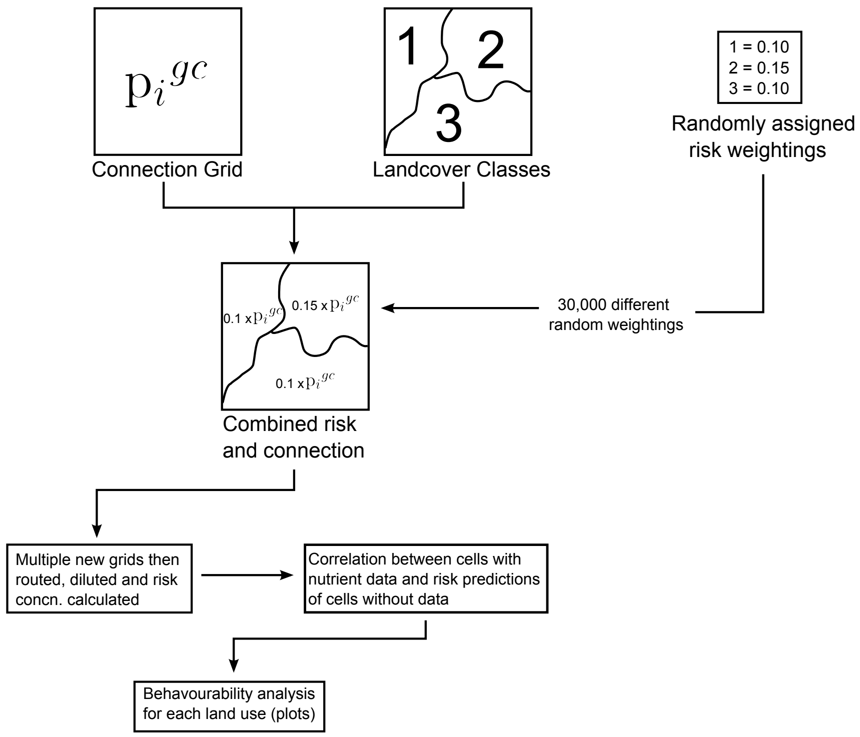

FIGURE 7.2: Inverse modelling of phosphorus data

7.7 Accounting for uncertainty

In order to account for uncertainty within the inverse modelling process, the standard deviations of the means of the GQA dataset were used. At the point where the correlation between cells containing measured nutrient data and those containing predicted risk (the second to last stage in Figure 7.2) the model takes into account the value of the standard deviation divided by the mean for each cell with observed data ($\sigma/\overline{x}$).

A low ($\sigma/\overline{x}$) value relates to a low standard deviation relative to the mean, and thus this point is given more weighting in the final plots and the land uses associated with this observed data point are subsequently rated more highly.

7.8 Inverse Modelling results

The outputs from applying the inverse model to the GQA phosphorus dataset and including the uncertainty weightings are below. Analysis and explanation of these plots is given in Section 7.9.

7.9 Inverse modelling: analysis and explanation

The inverse modelling process produces two visual outputs; (1) a dotty plot and (2) a behavourability analysis plot. The dotty plot shows predicted variables against individual parameters (in this case different land uses) for each simulation. The behavourability analysis is much more useful for this research as it can be used to assign a relative risk value to the different land uses within the catchment.

The plots are used to derive relative risk values for each different land use which form the erodability risk parameter in the SCIMAP framework. The plots are interpreted as follows;

(1) The x axis shows the correlation between observed nutrient concentrations and predicted nutrient concentrations of individual data cells (computational units) within the catchment. At $x = 0$ it can be said that the whole sample of cells has a correlation better than 0 and thus each starts with a generic mean and wide SD (corresponding to the dotted lines) bands on the y axis.

(2) The relative risk value should be read off as the correlation between the observed and predicted increases, however a caveat does apply; as the correlation increases, the sample size decreases, resulting in the SD bands also becoming narrow to the point of matching the mean value. This is a quirk of the modelling code and presents no issue as long as it is noted when interpreting the plots.

(3)The general trend of the line should be used as an indicator of what the relative risk value should be (i.e. any step changes as correlation increases should be treated with caution and a likely due to anomalies within the dataset).

(4)Broad SD bands (around 2.66) suggest the land use is unimportant and thus if they remain broad as correlation increases (noting the caveat above in (2) above) then it can be interpreted that the landuse in question is not a major contributor to the observed levels of nutrients in the river.

Based on the interpretation guidelines outlined above, the following interpretations are made;

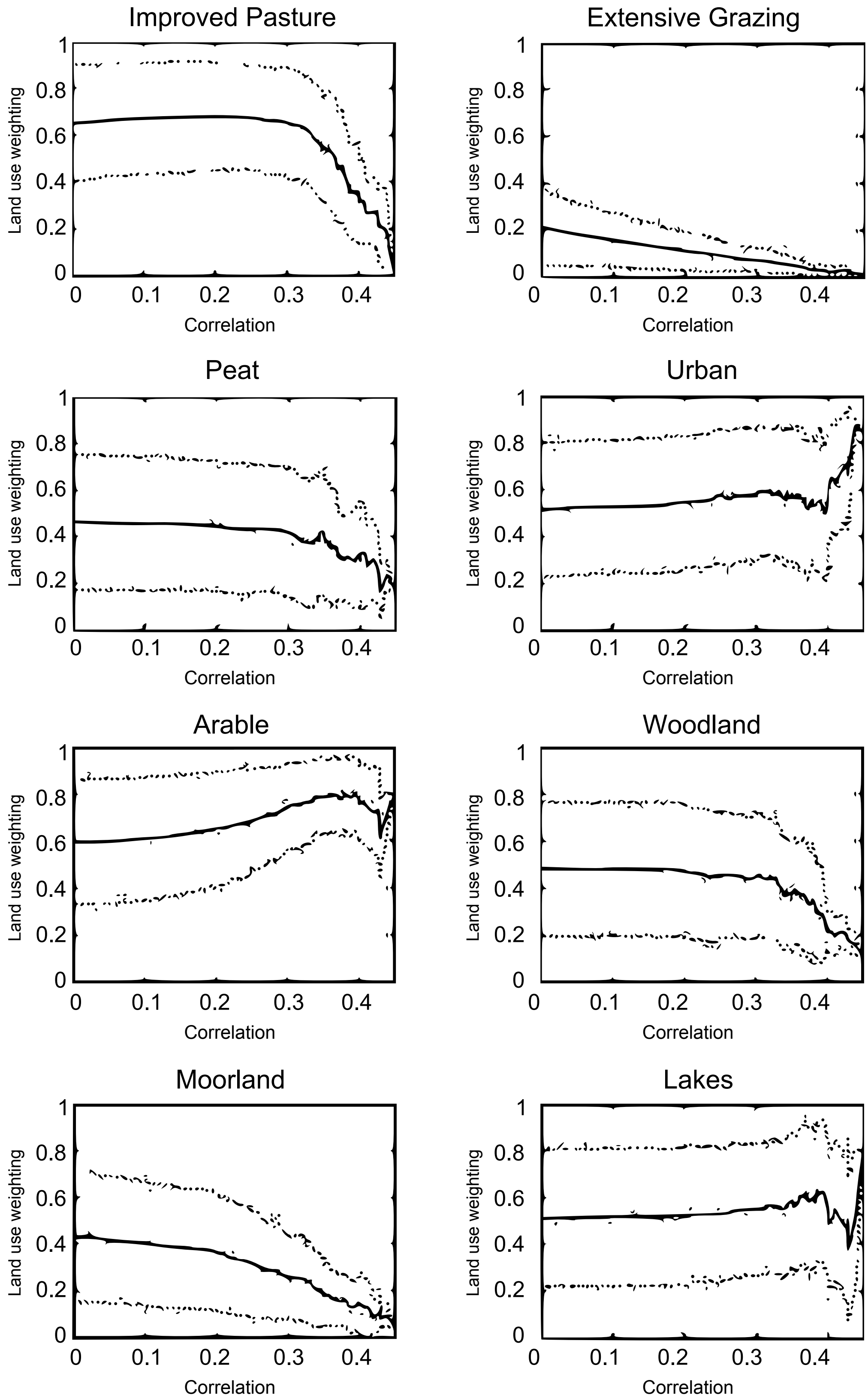

FIGURE 7.3: Outputs from inverse modelling of GQA data using uncertainty weightings

(1)Improved pasture: the plot shows a largely flat mean line as correlation increases from 0 to 0.3 where it sharply decreases towards a low land use weighting. However the SD bands remain wide as correlation approaches 0.45 and thus suggest this land use is not as important in the catchment. A low risk value can therefore be used in the SCIMAP model.

(2)Extensive grazing: this plot contrasts with the previous plot in that the SD bands quickly narrow as correlation increases which suggests that there is an excellent confidence in how the model is treating extensive grazing. The mean line has a general increasing trend with the SD bands becoming very narrow from around 0.35 correlation. A low relative risk weighting will be given to this land use in the SCIMAP model. The same applies to moorland as it follows the same pattern.

(3)Peat: this plot is very similar to that of improved pasture although it results in a slightly higher land use weighting of 0.2. The step change and broad SD bands as correlation increases suggests a slightly lower confidence in the models treatment of peat, and thus a slightly lower value will be assigned to this value.

(4)Urban areas: this plot shows a gradual increase in the mean land use weighting as correlation increases however the SD bars remain broad. The large step change when the correlation reaches 0.45 should be treated with caution. It is most likely a point source signal from sewerage treatment works located near urban areas which is contained in the nutrient data and carried through into the inverse modelling. This fits in with the extremely low logical a priori risk value.

(5)Arable: this plot shows an increasing mean land use weighting value as correlation increases, and although there are some small step changes the SD bands do narrow, and thus the suggested high land use weighting will be used. This fits with high logical risk weighting assigned and makes sense; land which has been ploughed and probably treated with additional nutrient based fertilisers or manures is potentially a major contributor to in-stream nutrient increases.

(6)Woodland: this plot displays a decreasing trend with associated narrowing of the SD bars before the inevitable narrowing as correlation approaches 0.5. Therefore a low risk weighting will be used in the SCIMAP model.

(7)Lakes: this plot shows similar trends to the urban plot and although results in a high land use weighting, such a value will not be used in the SCIMAP model. This is due to the broad SD bands and the mean line on the plot not deviating from the original generic mean of 0.5. Expert judgement suggests lakes should be given a relative risk weighting of 0 and so agrees with the interpretation of this plot.

TABLE 7.1: Table to show the risk weightings used in the SCIMAP framework.

Land Use	A Priori Value	Inverse Model	Change
Improved Pasture	0.30	0.01	-0.29
Extensive Grazing	0.15	0.00	-0.15
Urban	0.00	0.01	+0.01
Arable	1.00	0.80	-0.20
Woodland	0.05	0.10	+0.05
Moorland	0.05	0.10	+0.05
Lakes	0.00	0.00	0

Thus, the final risk weightings to be used are as follows;

7.10 SCIMAP Risk Maps

The following figures are example outputs from different stages of the SCIMAP modelling framework; figures 7.4 and 7.5 show intermediate data layers created during the modelling process, and figures 7.6 and 7.7 are the final outputs from the model.

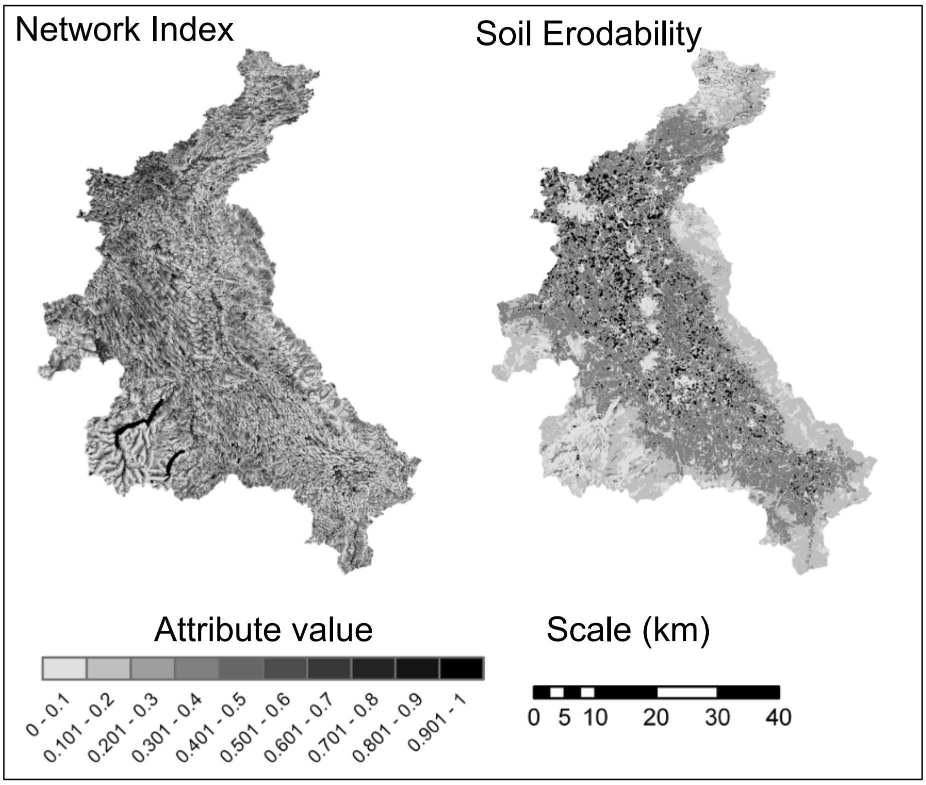

FIGURE 7.4: Network index and soil erodability - intermediate SCIMAP data layers

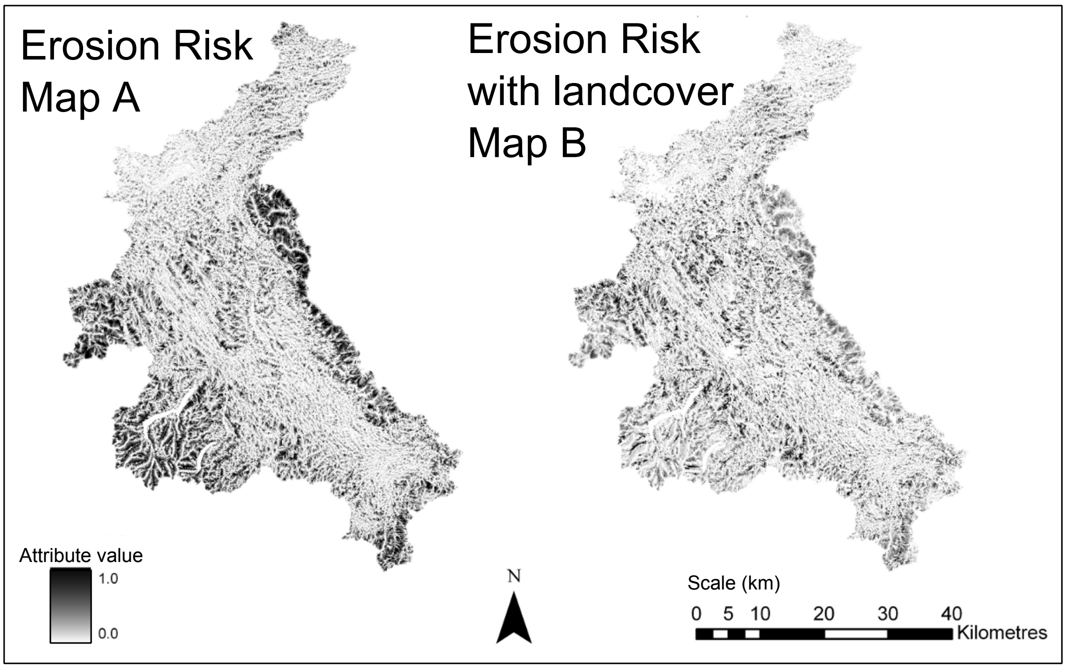

FIGURE 7.5: Erosion risk maps - intermediate SCIMAP data layers

7.11 Analysis of results

Even before the final outputs from the SCIMAP modelling tool are examined it is possible to use information from the data layers created during the modelling process to infer some of the properties of the catchment which might make certain areas more likely to be responsible for generating or transmitting diffuse pollution around the catchment. From the Network Index grid (Figure 7.4) which is in effect a representation of surface flow connection risk (Reaney *et al.* in review) it is clear that the most highly connected areas are in the low lying areas in the west of the catchment. The least connected areas are found in the eastern hillslopes of the Pennines and the south-western Lakeland fells.

The soil erodability data layer (Figure 7.4) also reveals some clues about the likely problem areas: again the greatest risk of soil erosion (and therefore associated nutrient delivery via fine sediment) occurs in the lowland areas of the Eden catchment. The areas with the greatest energy available to them for soil erosion (i.e. the stream power index data layer) are found in the Pennine and Lakeland hillslope. This means that the potentially serious combination of high erosion susceptibility meeting high potential erosive energy is avoided within the catchment.

Further analysis of the data layers created by the SCIMAP tool can give us more information: Figure 7.5 Map A is a combination of the landscape controls and connectivity of the catchment (stream power, flow paths, rainfall) and shows which areas are most likely

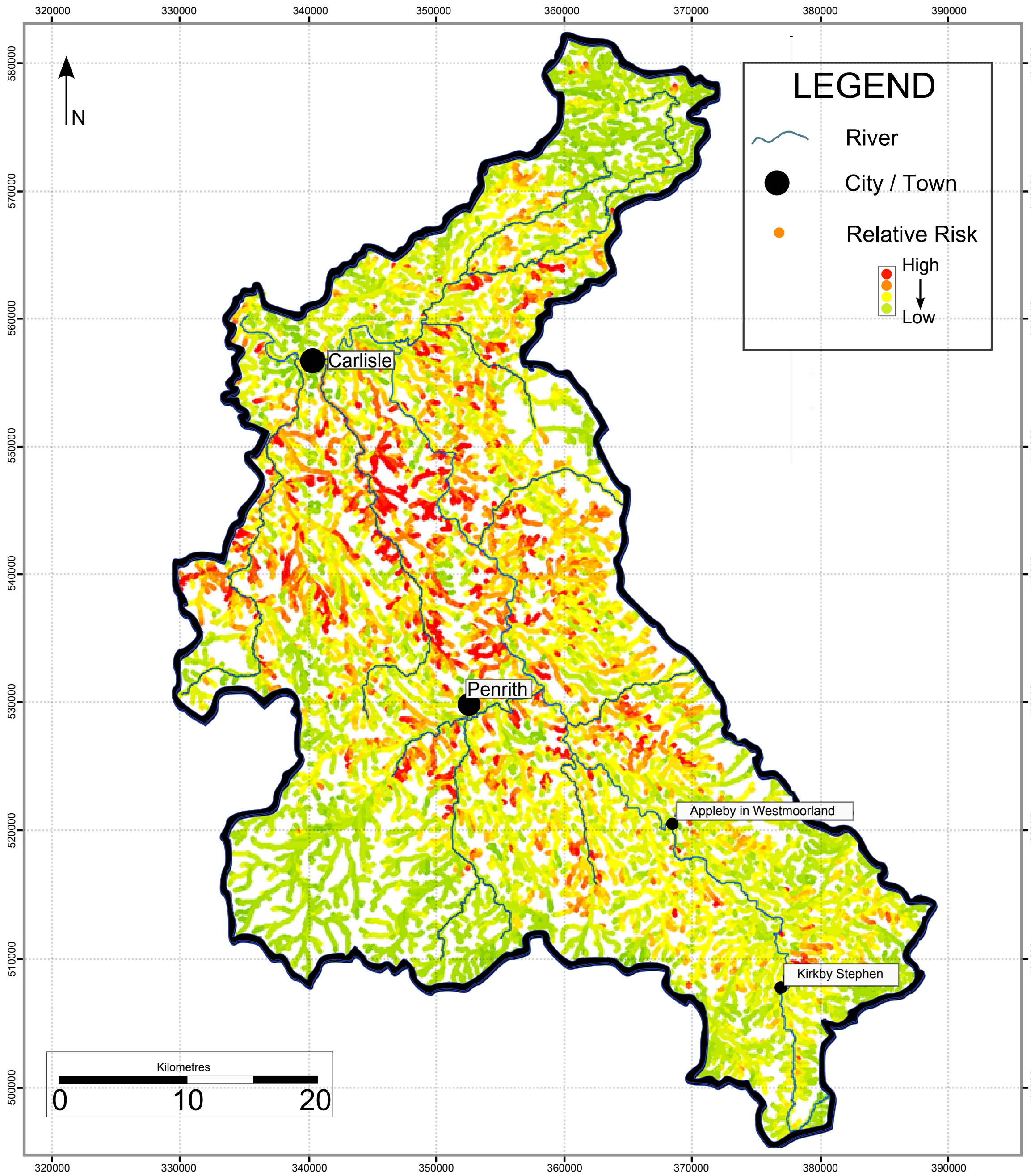

FIGURE 7.6: SCIMAP output based on logical erosion risk

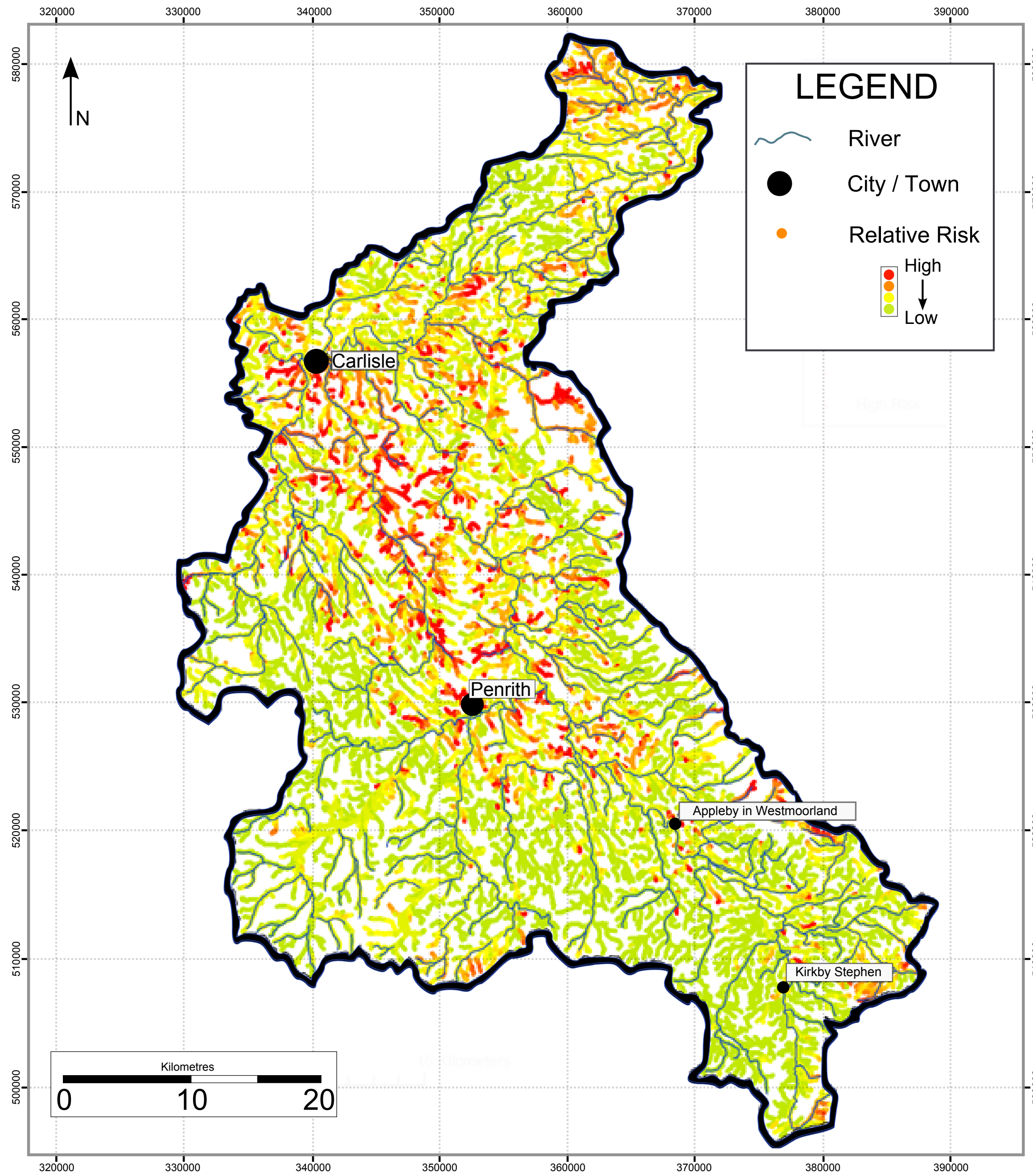

FIGURE 7.7: SCIMAP output based on inverse modelling results

to be susceptible to soil erosion based purely on topographic analysis (i.e. no land cover effects). These areas are mainly in the headwaters of the catchment. When landcover is also included (Figure 7.5 Map B) then the areas most at risk of soil erosion are shown to be within the main valley (River Eden) where there are large areas of arable farmland which are well connected to the main channels.

The final stages of the SCIMAP modelling process add a dilution effect to the accumulated risk to effectively give a risk concentration. The scalebar on the map runs green through red in multiples of the standard deviation of the mean of the risk value. Where this value is more than the mean (i.e. towards the red end of the spectrum) this shows that the risk increases faster than the dilution effect can alleviate it, therefore the area is identified as having a risky input to it. Conversely if the risk value is a lower multiple of a standard deviation of the mean (going towards the green end of the spectrum) it shows there are low risk inputs to the channels.

As seen in Section 7.9, the risk values for different land uses which are inputted into the SCIMAP system are extremely similar, and match each other in terms of the overall relative risk of each land use. For example, arable land is downgraded from 1.00 to 0.80 after inverse modelling, however this land use is still deemed to be the most risky. Therefore, both maps produced from SCIMAP (figures 7.6 and 7.7) using the two sets of risk values are very similar.

In order to quantify the differences between the two outputs some further processing was undertaken. Firstly, each of the 195,686 individual risk points which make up the maps were exported from the GIS package into a dataset of location (X and Y) and a risk value. Once this operation was completed on both the maps a third map could be generated, based on the differences between the risks at each location (Figure 7.8).

The area with most change (although it should be noted that the changes are of a small magnitude) is between Carlisle and Penrith located on a small tributary of the River Eden. This is shown as a cluster of dark red at approximately 348000,545000 on the map. The CEH Landcover map of the area shows this section of the catchment is made up predominantly of improved grassland, intensive grazing or hay / silage cut. This land use was given a lower risk weighting in the values based on the inverse modelling, and this area of change could reflect the reduction in risk which occurs as a result of reclassing a large area of risky land. However, this pattern has not been reflected more widely across the catchment.

Statistical analysis of the changes in risk between the two SCIMAP outputs shows just how small any changes between the two maps are (Figure 7.9). This shows the majority of changes between the risk values were very small (+/- less than 0.05).

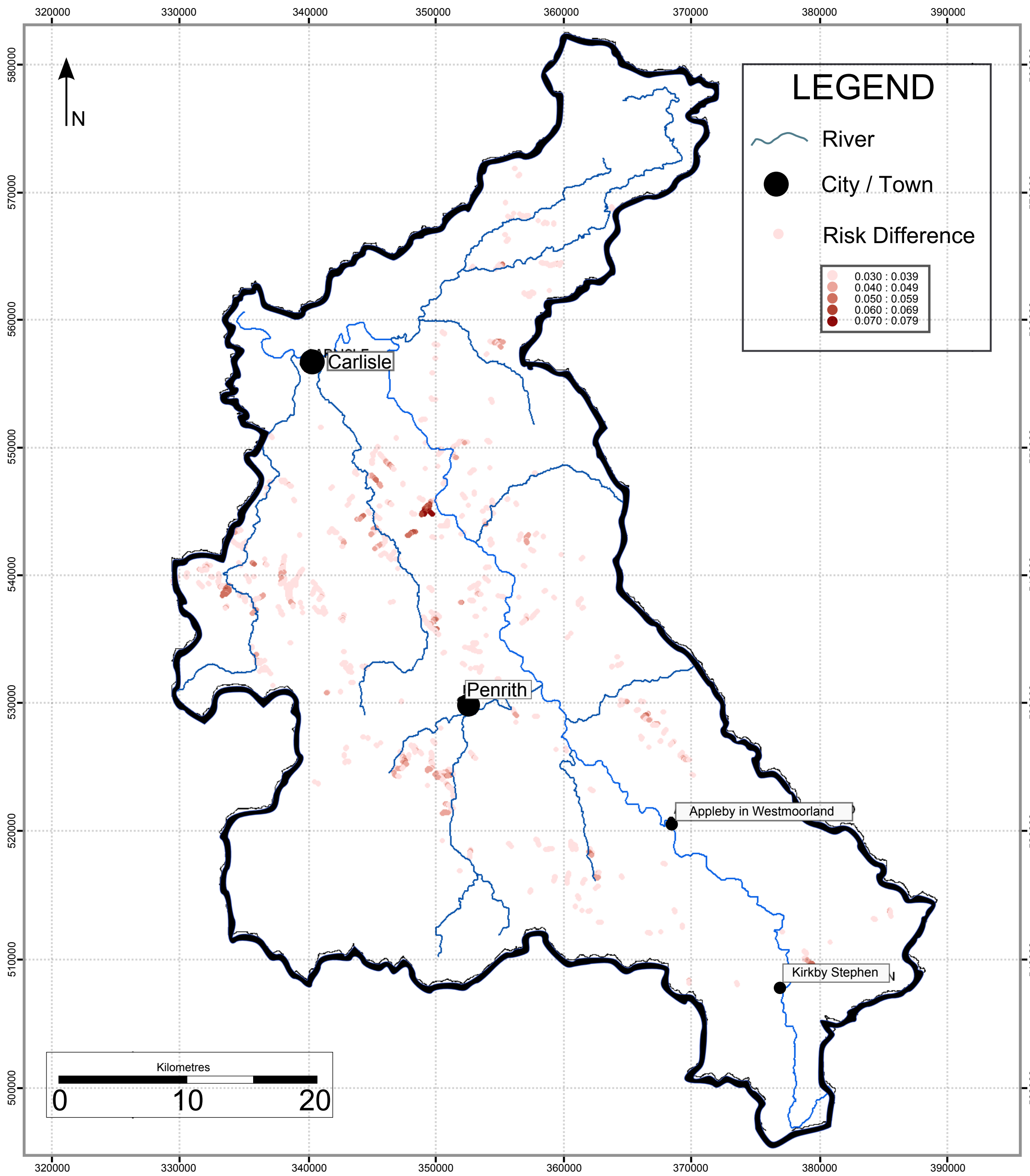

FIGURE 7.8: Risk difference between logical and inverse modelling

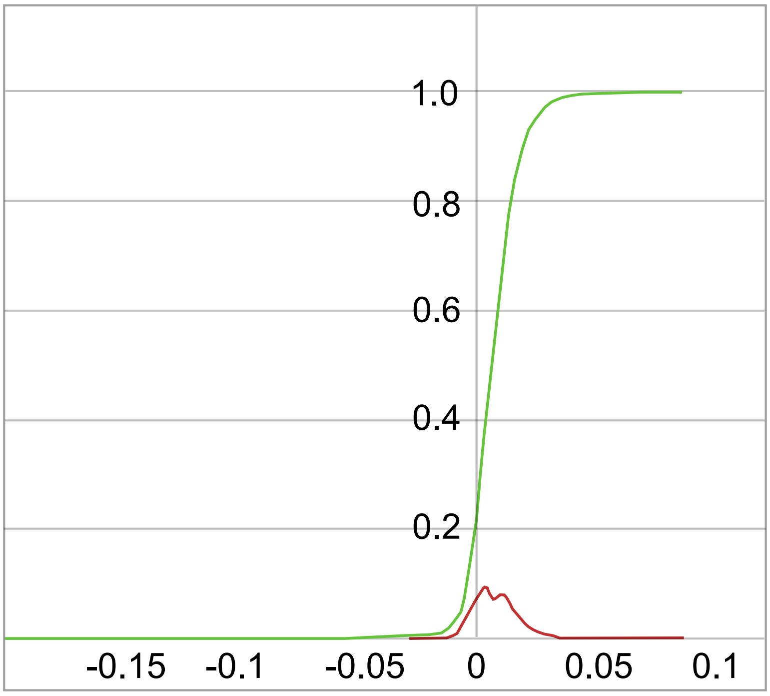

FIGURE 7.9: Comparing logical and inverse modelling relative risk change

In order to assess the differences or similarities between predicted risk within the catchment and observed field data from the GQA scheme a plot was made of the means of all the observed data points against their nearest predicted risk value. To calculate this the NEAR command in ESRI ArcWorkstation was used to calculate and extract the nearest predicted risk point to each of the measured data sites. A scatter plot was then plotted and is shown in Figure 7.10. Although the R^2 is low (0.14) there is a broad positive correlation between the two. This is to expected as the plot does not compare like for like; one axis is the means of raw field data (with associated uncertainty) and the other is a relative risk value which has been scaled to between 0 and 1. The broad positive correlation however, would be expected and suggests that the patterns of higher risks, match the areas with higher nutrient levels.

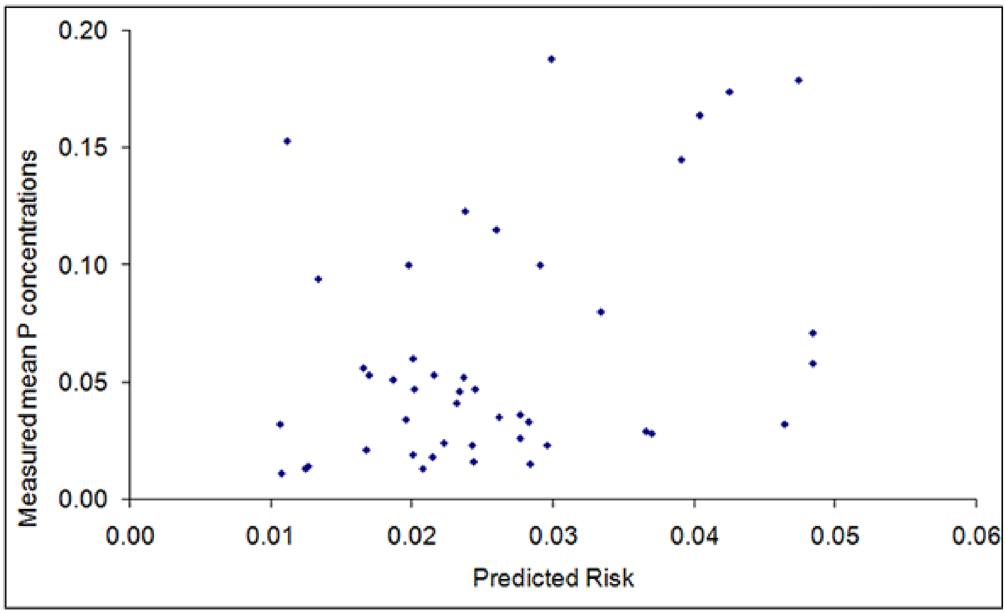

FIGURE 7.10: Relative risk versus observed field data

Chapter 8

Discussion and conclusion

The research objectives cited in Chapter 1 were as follows:

- Develop a technique to improve the quality of the available water quality data in the River Eden catchment (Chapter 6).
- Develop and apply inverse modelling techniqes to water quality data to generate risk values for land uses in the catchment (Chapter 6).
- Use the newly developed land use risk weightings within the SCIMAP modelling framework to produce risk maps for the River Eden catchment (Chapter 7).

These objectives will be discussed in the following sections:

8.1 Improving the quality of nutrient data

The Monte-Carlo based simulation developed and used in this research project was based on the approach by Tarras-Wahlberg & Lane (2003) in the study into suspended sediment yield in Ecuador. In the case of this project the simulation aimed to estimate flow weighted phosphporus concentrations from the Environment Agency datset which consists of sporadic (monthly at best) but spatially rich (90 sites across the River Eden catchment) phosphorus concentrations. These sites cover the main channels in the River Eden catchment (Eden, Eamont, Irthing, Petteril and Caldew) as well as multiple smaller tributaries.

This work was undertaken as previous investigations have shown that the majority of phosphorus transfers take place in a short period of time (67% in 2% of the time in the

River Exe catchment (Walling & Webb 1985)). Thus by matching known concentrations with discharge at the sites across the River Eden, developing a regression relationship for each of these and then applying the relationship across the duration of the discharge time series it is possible to estimate the flow weighted concentrations (CIWEM 2004).

This method was successfully applied to ten sites across the River Eden catchment and the results in Figures 4.8 and 4.9 show a clear reduction in variation (as measured by standard deviation). Unfortunately there was not sufficient discharge data available to apply this method across the whole of the River Eden catchment and make use of the nutrient data for the remaining 80 sites. However the software code and background infrastructure (database) for storing the nutrient data remain in the hands of the SCIMAP project team and thus can be applied relatively easily in the future.

This research finds that, despite some evidence in the literature (e.g. Walling % Webb 1985, CIWEM 2004) that suggests the regular time interval sampling which forms the GQA dataset could miss many phosphorus transfer events, such data can provide a suitable alternative for driving an inverse model.

8.2 Applying the inverse model

Inverse modelling techniques have already been used in an application of the SCIMAP modelling framework in the River Eden catchment (Reaney *et al.* in review). Mean values for phosphate concentration for each site where nutrient data was available (90 sites) were used and applied to a simplified version of the inverse modelling setup used by Reaney *et al.* in review).

This method assigned risk weightings to all cells within the catchment based on their landuse and connection probabilities and correlated the risk values for cells without nutrient data, and those with nutrient data (i.e. where the Environment Agency measuring stations are). This method produces a series of risk values for different land uses within the catchment over a simulation period of 30,000 iterations to determine which estimated risk values best match the observed characteristics in the field. These risk weightings can then be used as risk weightings within the SCIMAP modelling framework. Whilst Lane (2008) puts forward a case for inverse modelling, due to issues surrounding data availability and the principles of: (1) letting the affected organisms or environmental indicator prescribe what hydrological processes need to be included in models; and (2) keeping the modelling framework as simple as possible so nationally available data can be used; it is worth noting here that the inverse model used in this research will only be as effective as the data which is used in it; which returns us once again to the issue of whether the data from the EA GQA scheme is suitable.

However this GQA data was the best available within the timeframe and cost parameters of this research and so has been included. Results from the inverse modelling match very closely those risk values derived from expert logical judgment. For example, arable land was asigned a slightly decreased risk value of 0.8 from its expert assigned value of 1.0. This in many ways is reassuring. The expert judgment deemed this land most risky and thus assigned it the highest possible value due to the likelehood of the land being ploughed and exposed to nutrient applications. Inverse modelling of the phosphorus data appears to have included the clear risk which arable land poses and has assigned it a suitably high risk value.

8.3 Applying the SCIMAP model

With the new risk values for various land uses within the catchment generated from inverse modelling it was possible to run the SCIMAP model with both the original logical risk weightings and the new modelled risk weightings. The resulting risk maps are very similar and do not show any obvious areas where there is a significant risk change.

The SCIMAP framework has some clear attributes. Firstly it can be run using data which is readily available to scientists at a national scale. Secondly once the pre-processing of these datasets is completed the time taken to apply the model to the River Eden catchment was less than an hour. The constraint on not using a finer spatial resolution within the SCIMAP framework was not time or cost; it was software buffer problems which caused consistent crashes. Preliminary investigations indicated that this could be resolved by recompiling the SAGA-GIS and SCIMAP software to function on more powerful 64bit computers.

The SCIMAP model outputs suggest that the most risky areas for in-stream water quality within the River Eden catchment are in the main valley of the River Eden. The model enables some characteristics of the catchment to be inferred before the SCIMAP model run was completed, by analysing some of the grids produced during the earlier stages of the modelling run. This analysis suggested the most likely areas of diffuse pollution contribution were in the headwaters of the catchment (if landscape controls on erosion and connectivity characteristics, stream power, flowpaths and average rainfall intensity are considered). Once landcover was included in the analysis then the areas shown to be most at risk are the arable farmland areas surrounding the main channel of the River Eden. This strongly suggests that landuse is the key controlling factor in determining risky areas within the River Eden catchment.

8.4 Conclusions

Several conclusions can be drawn from these results:

(1) If these outputs were being used by stakeholders in water pollution, for example Environment Agency or DEFRA officials, local rivers trusts and landowners, these people would be advised by the maps to look at the same fields if they looked at either the map driven by logical erosion weights, or the map driven by inverse model derived risk weightings. This suggests that obtaining water quality data, processing it and then undertaking statistical modelling may not be necessary, and land use combined with local topography and rainfall data can be enough to highlight areas which need closer examination.

(2) The lack of available suitable discharge data for the catchment and the forced use of potentially unreliable phosphorus concentration data could mean that the results of inverse modelling are unreliable and do not capture a vast amount of phosphorus transfers. The resolution of nutrient concentration data that is available on the River Eden pales into insignificance when compared to that which can be obtained using bank-side monitoring equipment. Work has been undertaken in Northern Ireland (Jordan *et al.* 2005) where equipment installed in a sub catchment of Lough Neagh is capable of recording total phosphorus (TP) concentrations at a 10 minute resolution.

(3) Conversely the good correlation between the logical erodability and inverse modelled risks could suggest that the low resolution data collected by the Environment Agency as part of their GQA scheme captures enough data on phosphorus transfers to enable risk mapping to be undertaken using the inverse modelling approach adopted here.

The next stage in applying the research conducted here, and future catchment studies undertaken using the SCIMAP modelling framework should be to take the SCIMAP output into the field and engage with stakeholders (the Catchment Sensitive Farming officers, landowners, famers and Rivers Trusts) and examine the areas which the model has deemed risky. Groundtruthing will enable us to determine what it is on the ground which is causing the high risk levels and should provide better model validation data than any computerised dataset.

If the SCIMAP framework is successful in identifying the areas at major risk to the quality of water in the channels then there is little doubt that it provides an extremely cost-effective and realtively simple approach to identification of land use hotspots. With the deadline for meeting the standards prescribed by the Water Framework Directive fast approaching, I have no doubt that the SCIMAP tool could prove invaluble to Goverment in their efforts to improve the quality of rivers in the UK.

Chapter 9

Recommendations for future research

9.1 Nutrient data

As discussed the conclusion, both SCIMAP maps produced would result in people visiting the same areas of the catchment to carry out further investigations in the field. It would be useful to carry out the same processes that I have undertaken in the River Eden catchment on a few more catchments (preferably with different hydrological and topographic characteristics) to compare the risk map outputs and see if the patterns remain broadly the same. This would perhaps negate the need to use measured nutrient data to pinpoint land use hotspots, and instead use a simple combination of landuse data and topography data, which is readily available at a national scale to highlight areas which need further investigation.

9.2 Flow weighted concentration estimation

A second reccomendation relates to the flow weighted concentration work (Chapter 6). In order for this technique to be fully assessed and applied within the SCIMAP modelling framework it is necessary to obtain more discharge data. This would enable the remaining 80 sites within the catchment to be included, and thus a new set of land use weightings to be generated by the inverse modelling technique.

9.3 Public availability

Thirdly I would suggest that the outputs from the SCIMAP framework are integrated with a new user-interface. It would be extremely viable to export the X and Y locations from the SCIMAP model with their associated risk values and use simple software to integrate them with the Google Maps system. Google Maps provides an unrivalled public GIS system which includes satellite and terrain data as well as road and street maps at a click of a button. Thus if the project was rolled out and used by stakeholders and end-users it would be extremely easy to discuss the results and investigate further from the office, house or meeting room.

References

Andersen, H.E., Kronvang, B. & Larsen, S.E., 2005. Development, validation and application of Danish empirical phosphorus models. *Journal of Hydrology*, 304(1-4), 355–365.

Asselman, N.E.M., 2000. Fitting and interpretation of sediment rating curves. *Journal of Hydrology*, 234(3-4), 228–248.

Beasley, D.B., Huggins, L.F. & Monke, E.J., 1980. ANSWERS: A model for watershed planning. *Transactions of the ASAE*, 23(4).

Beven, K., 1989. Changing ideas in hydrology: the case of physically-based models. *Journal of Hydrology,* 105 (1-2).

Beven, K.J. & Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1).

Bowes, M.J. et al., 2005. The relative contribution of sewage and diffuse phosphorus sources in the River Avon catchment, southern England: implications for nutrient management. *Science of the Total Environment,* 344(1-3), 67–81.

Brady, N.C. & Weil, R.R., 1990. Nature and properties of soils. Prentice Hall, 393

Burt, T.P. et al., 1999. Denitrification in riparian buffer zones: the role of floodplain hydrology. *Hydrological processes*, 13(10), 1451–1463.

Carpenter S R, Caraco N F, Correll D L, Howarth R W, Sharpley A N and Smith V H, 1998. Nonpoint pollution of surface sources with phosphorus and nitrogen. *Ecological Applications.* 8 (3), 559-568.

Cartz, L. et al., 1979. Effect of pressure on bonding in black phosphorus. *The Journal of Chemical Physics*, 71, 1718.

CIWEM, 2004. The Water Framework Directive: Integrating Approaches to Diffuse Pollution.

Daly, K. et al., 2002. Modeling Phosphorus Concentrations in Irish Rivers Using Land Use, Soil Type, and Soil Phosphorus Data: Landscape and Watershed Processes. *Journal of environmental quality*, 31(2), 590.

DEFRA, 2002. Directing the flow: priorities for future water policy. Department for Environment, Food and Rural Affairs, London. 9-31.

Directive (91/676/EEC), The Nitrate Directive. *Official Journal L 375 , 31/12/1991 P. 0001 – 0008.*

Directive (92/43/EEC), The Habitats and Birds Directive. *Official Journal L 206 , 22/07/1992 P. 0007 – 0050.*

Directive 2000/60/EEC,, Water Framework Directive. *Official Journal L 327, 22.12.2000, p. 1–73.*

Directive 91/271/EEC, Habitats and Urban Waste Water Treatment Directive. *Official Journal L 135, 30.5.1991, p. 40–52.*

Dunn, S.M. et al., 1996. The hydrological component of the NELUP decision-support system: an appraisal. *Journal of Hydrology*, 177(3-4), 213–235.

Dwyer, J. et al., 2002. *Policy mechanisms for the control of diffuse agricultural pollution, with particular reference to grant aid*, English Nature. 26-32.

Environment Agency, 2007. The unseen threat to water quality: Diffuse water pollution in England and Wales report. Published by the Environment Agency, Bristol, UK 2-16.

Environment Agency, 2008. River Eden, Cumbria: Catchment Flood Management Plan, Published by the Environment Agency, Warrington, UK, 14.

Environment Agency, 2008b. Water Quality Data Library information. http://www.environment-agency.gov.uk/research/library/data/34383.aspx Accessed December 2008.

Ferguson, R.I., 1986. River loads underestimated by rating curves. *Water Resources Research*, 22(1).

Fraser, A.I., Harrod, T.R. & Haygarth, P.M., 1999. The effect of rainfall intensity on soil erosion and particulate phosphorus transfer from arable soils. *Water Science & Technology*, 39(12), 41–45.

Gburek, W.J, Sharpley A N, Heathwaite A L and Folmar G L., 2000. Phosphorus management at the watershed scale: a modification of the phosphorus index. *Journal of Environmental Quality*, 29, 130-144.

Harris, G. & Heathwaite, A.L., 2005. Inadmissible evidence: knowledge and prediction in land and riverscapes. *Journal of hydrology*, 304(1-4), 3–19.

Harris R C & Skinner A C, 1992. Controlling Diffuse Pollution of Groundwater from Agriculture and Industry, *Water and Environment Journal,* 6 (6), 569-574.

Haygarth, P.M. et al., 2005. The phosphorus transfer continuum: Linking source to impact with an interdisciplinary and multi-scaled approach. *Science of the Total Environment,* 44(1-3), 5–14.

Heathwaite, A.L. et al., 2005. A tiered risk-based approach for predicting diffuse and point source phosphorus losses in agricultural areas. *Science of the Total Environment,* 344(1-3), 225–239.

Heathwaite, A.L. & Johnes, P.J., 1996. Contribution of nitrogen species and phosphorus fractions to stream water quality in agricultural catchments. *Hydrological Processes*, 10(7), 971–983.

Heathwaite, A.L., Quinn, P.F. & Hewett, C.J.M., 2005. Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. *Journal of Hydrology*, 304(1-4), 446–461.

Heathwaite, L. et al., 2000. A conceptual approach for integrating phosphorus and nitrogen management at watershed scales. *Journal of Environmental Quality* 29, 158-166.

Hilton, J., Buckland, P. & Irons, G.P., 2002. An assessment of a simple method for estimating the relative contributions of point and diffuse source phosphorus to in-river phosphorus loads. *Hydrobiologia*, 472(1), 77–83.

Johnes, P., Moss, B. & Phillips, G., 1994. Lakes-classification and monitoring A strategy for the classification of lakes. R&D Report: National Rivers Authority (Bristol), 286/6/A.

Johnes, P.J., 1996. Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: the export coefficient modelling approach. *Journal of Hydrology*, 183(3-4), 323–349.

Johnes, P.J. & Burt, T.P., 1993. Nitrate in surface waters. In: *Nitrate: Processes, Patterns and Management*. Wiley-Blackwell, Chapter 8 269-321.

Jokela, W.E. et al., 2004. Effectiveness of agricultural best management practices in reducing phosphorus loading to Lake Champlain. *The Science and Management of Soil and Water: Lake Champlain*, 14.

Kavetski, D., Kuczera, G. & Franks, S.W., 2003. Semidistributed hydrological modeling: A “saturation path” perspective on TOPMODEL and VIC. Available at: http://www.agu.org.ezphost.dur.ac.uk/pubs/crossref/2003/2003WR002122.shtml [Accessed November 12, 2009].

Khadam, I.M. & Kaluarachchi, J.J., 2006. Water quality modeling under hydrologic variability and parameter uncertainty using erosion-scaled export coefficients. *Journal of Hydrology*, 330(1-2), 354–367.

Kinsel, W.G., 1980. CREAMS, A Field Scale Model for Chemical Runoff an Erosion form Agricultural Management Systems. *Conservation Report*, 26, 640 pp.

Kirkby, M., 1975. Hydrograph modelling strategies. *Processes in physical and human geography: Bristol essays*, 69.

Krueger, T. et al., 2007a. Processes affecting transfer of sediment and colloids, with associated phosphorus, from intensively farmed grasslands: a critical note on modelling of phosphorus transfers. *Hydrological Processes*, 21(4), 557–562.

Krueger, T. et al., 2007b. Processes affecting transfer of sediment and colloids, with associated phosphorus, from intensively farmed grasslands: a critical note on modelling of phosphorus transfers. *Hydrological Processes*, 21(4), 557–562.

Lane, S.N., 2008. What makes a fish (hydrologically) happy? A case for inverse modelling. *Hydrological Processes*, 22(22), 4493–4495.

Lane, S.N. et al., 2006. Surveillant science: Challenges for the management of rural environments emerging from the new generation diffuse pollution models. *Journal of Agricultural Economics*, 57(2), 239–257.

Lane, S.N. et al., 2004. A network-index-based version of TOPMODEL for use with high-resolution digital topographic data. *Hydrological Processes*, 18(1), 191-201.

Lane, S.N., Reaney, S.M. & Heathwaite, A.L., 2009. Representation of landscape hydrological connectivity using a topographically driven surface flow index. *Water Resources Research*, 45(8), W08423.

Lewis, D.R. & McGechan, M.B., 2002. SW—Soil and Water : : A Review of Field Scale Phosphorus Dynamics Models. *Biosystems Engineering*, 82(4), 359-380.

Melack, J.M., 1995. Transport and transformations of P, fluvial and lacustrine ecosystems. 245–254. *Phosphorus in the global environment. Scientific Committee on Problems of Environment*, 54.

Meybeck, M., 1982. Carbon, nitrogen, and phosphorus transport by world rivers. *American. Journal of Science* 282(40), l–450.

Milledge, D.G., Lane, S.N. & Warburton, J., 2009. The potential of digital filtering of generic topographic data for geomorphological research. *Earth Surface Processes and Landforms*, 34(1), 63–74.

Munn, N.L. & Meyer, J.L., 1990. Habitat-specific solute retention in two small streams: an intersite comparison. *Ecology*, 71(6), 2069–2082.

Murdoch, E.G., Whelan, M.J. & Grieve, I.C., 2005. Incorporating uncertainty into predictions of diffuse-source phosphorus transfers (using readily available data). *Water science and technology: a journal of the International Association on Water Pollution Research*, 51(3-4), 339.

Omernik, J.M., 1976. The influence of land use on stream nutrient levels. *Report EPA 600/3-76-014 January 1976. 114*, 29.

Perry, M. & Hollis, D., 2005. The development of a new set of long-term climate averages for the United Kingdom. *Int. J. Climatol*, 25, 1023–1039.

Planchon, O. & Darboux, F., 2002. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, 46(2-3), 159–176.

Quinn, P., 2004. Scale appropriate modelling: representing cause-and-effect relationships in nitrate pollution at the catchment scale for the purpose of catchment scale planning. *Journal of Hydrology*, 291(3-4), 197–217.

Reaney, S., 2008. The use of agent based modelling techniques in hydrology: determining the spatial and temporal origin of channel flow in semi-arid catchments. *Earth Surface Processes and Landforms*, 33(2), 317-327.

Reaney, S. et al., (in review). Modelling the risks of fine sediment delivery to river catchments based on an explicit treatment of hydrological connectivity. *Ecological Modelling*.

Reaney, S.M., 2008. The use of agent based modelling techniques in hydrology: determining the spatial and temporal origin of channel flow in semi-arid catchments. *Earth Surface Processes and Landforms*, 33(2), 317–327.

Reaney, S.M., Bracken, L.J. & Kirkby, M.J., 2007. Use of the Connectivity of Runoff Model (CRUM) to investigate the influence of storm characteristics on runoff generation and connectivity in semi-arid areas. *Hydrological Processes*, 21(7), 894–906.

Robinson, R.A. & Sutherland, W.J., 2002. Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology*, 157–176.

Schleich, J., White, D. & Stephenson, K., Cost implications in achieving alternative water quality targets. *Water Resources Research*, 32(9).

SCIMAP project website: http://www.scimap.org.uk (Accessed December 2008)

Smith, V.H., Tilman, G.D. & Nekola, J.C., 1999. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution*, 100(1-3), 179–196.

Tarras-Wahlberg, N.H. & Lane, S.N., 2003. Suspended sediment yield and metal contamination in a river catchment affected by El Nino events and gold mining activities: the Puyango River Basin, southern Ecuador. *Hydrological Processes*, 17(15), 3101–3123.

Vollenweider, R.A. & Kerekes, J., 1982. Eutrophication of waters, monitoring, assessment and control. *Organization of Economic Cooperation and Development, Paris, France*.

Walling, D.E. & Webb, B.W., 1985. Estimating the discharge of contaminants to coastal waters by rivers: Some cautionary comments. *Marine Pollution Bulletin*, 16(12), 488–492.

Water UK, 2007. Managing agricultural diffuse pollution: Time for a change, Policy Statement, Water UK, London. 2-3.

Withers, P.J. & Haygarth, P.M., 2007. Agriculture, phosphorus and eutrophication: a European perspective. *Soil Use and Management*, 23, 1.

Withers, P.J. & Lord, E.I., 2002. Agricultural nutrient inputs to rivers and groundwaters in the UK: policy, environmental management and research needs. *Science of the Total Environment,* 282, 9–24.

Withers, P. & Sharpley, A., 2008. Characterization and apportionment of nutrient and sediment sources in catchments. *Journal of Hydrology*, 350(3-4), 127-130.