Spatial Aspects of Chemical Exposure Assessment: A Tool for River Networks

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften am Fachbereich
Mathematik / Informatik der Universität Osnabrück

vorgelegt von
Jan-Oliver Wagner
geboren am 2. März 1971 in Hamburg

Osnabrück, den 18. Juni 2001

Summary

Chemical exposure assessment has gained increasing attention in recent years. Its methodologies have enabled scientists and policy-makers to understand exposure paths and to identify environmental compartments of concern. Mathematical models are used for the prediction of a chemical's concentration in a certain compartment and in some cases also for predicting the duration or time of highest load. With the Geo-referenced Regional Exposure Assessment Tool for European Rivers (GREAT-ER) spatial aspects of regional exposure assessment are addressed for the "down-the-drain" path of consumer chemicals such as detergents.

On the basis of a carefully developed simulation model (Boeije, 1999), this thesis describes the concept and realization of the developed software tool GREAT-ER. With data composition and processing on the one hand and application and analysis on the other hand, two crucial aspects in spatial exposure assessment are identified and discussed.

Geo-referenced real-world data are not readily available in a usable form. An intermediate format is defined to separate the tasks of an initial preparation of raw data from the final aggregation leading to a directly usable data set. It is shown that the latter step can be fully automated and thus efficiently supports an iterative procedure of data quality improvement.

The application of GREAT-ER to the substances LAS (readily degradable) and boron (inert) in four Yorkshire catchments demonstrates the ability to predict mean final effluent and in-stream concentrations with an average error of less than a factor of 2. Furthermore, regional summaries and risk characterization add useful information to judging a regional response to the (potential) release of a substance.

In conclusion, the development and application of GREAT-ER has proven that geo-referenced exposure assessment is possible with regard to both quality and practicability. Future activities should focus on gaining further experiences in performing simulations, improving the tool itself and extending its abilities. Finally the integration of further models should be evaluated.

Contents

Chapter 1

Problem description

1.1 Introduction

In recent decades industrial societies have started to make the protection of the environment and human health an integral part of their policies, economy and science.

At the beginning, ad-hoc plans and actions to deal with pollution that posed an immediate and direct threat dominated. Popular examples are types of pollution that can easily be seen or smelt such as air pollution in the German Ruhrarea (early 1960s, law implementation in 1964) or the eutrophication of German lakes by phosphate from detergent products (1970s, law implementations in 1975 and 1980).

Besides these, other chemicals, which were detectable only using high-tech devices or after the harmful effects on humans and environment had become apparent, were identified (partly by chance) as posing a high and unacceptable risk (e.g. dioxins).

Awareness to avoid future problems increased and lead to the demand for a variety of prediction tools. These were initially used to improve understanding of the dynamics of chemicals and their effects on the environment, and were later used on a regulatory level (e.g. for the evaluation of new substances).

With the current EU chemical legislation only the local assessment of ecotoxic

and human-toxic risks is considered based on one generic environmental release. Additional background concentrations are calculated by a regional multimedia box model without spatially explicit information.

The applied generic-based methodologies have proven sufficient for some types of initial screening (i.e. problem identification), but have also revealed a number of disadvantages (e.g. problematic validation, can not be validated for explicit sites) and missing features (e.g. not viable for the managment of identified problems).

These drawbacks are addressed by introducing spatially explicit methods. These are based on comprehensive (spatially explicit) data and a powerful software tool with regard to usability and performance (response time).

The tremendously rapid development of information technologies rendered the use of chemical fate models as software tools widely usable and well accepted. Theoretical concepts became applicable with the given information technology. One example is comprehensive uncertainty and variability analysis by methods with high computational efforts (e.g. Monte-Carlo approach). The present thesis approaches a further level of complexity: the spatial refinement of environmental simulation models and its feasibility.

1.2 Environmental Risk Assessment

Legislation required a measurable, quantitive method for balancing the need of chemical products and the hazard they posed to human health and the environment. The EU directive 93/67/EEC on Environmental Risk Assessment (ERA) describes a methodology which is mandatory for new substances on the European market.

A risk posed by a chemical to the environment is characterized by its toxicological effect on organisms and the concentration of the substance within environmental compartments. The substance is considered to pose a risk if the environmental concentration exceeds the no-effect concentration on organisms (figure 1.1).

For new substances both concentrations (effect and exposure) need to be estimated (EU, 1996): The predicted no-effect concentration (PNEC) is usually determined on the basis of results from monospecies laboratory toxicity tests or, in a few cases, established from concentrations determined from model ecosys-

tem tests, taking into account adequate safety factors. The PNEC is regarded as a concentration below which an unacceptable effect will most likely not occur. The predicted environmental concentration (PEC) is derived from monitoring data or estimated by applying simulation models. In the first stage of exposure assessment where exposure models are used, generic exposure scenarios are applied. It is assumed that substances are emitted into non-existing environments with predefined agreed environmental characteristics. These characteristics can be average values or reasonable worst-case values depending on the parameter in question.

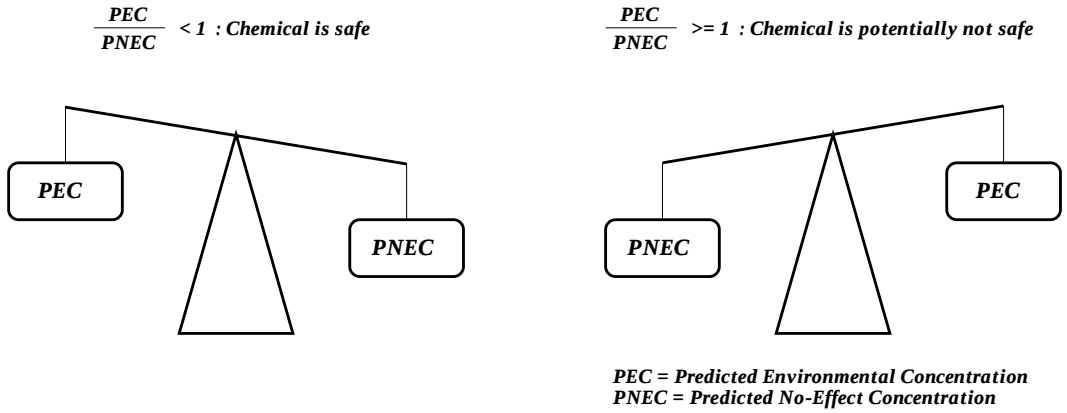


Figure 1.1: *Environmental Risk Assessment: PEC/PNEC ratio*

The ERA process involves a number of uncertainties, safety factors and stochastics. The primary rule is to not classify an unsafe substance as safe. Chemicals that do not pass the first screening phase are not necessarily unsafe. Refinement with reliable data on chemical release / emission (place, mass, duration) and physico-chemical properties (degradation, sorption, etc.) forms the second tier of the ERA process. The present thesis focuses on refinements of the ERA exposure part. The effect part is not directly linked to questions of environmental exposure assessment and should be regarded as a separate task.

1.3 Chemical Exposure Assessment

Exposure assessment estimates the dose or the quantity of risk agents (e.g. toxic chemicals) received by individuals or the environment (Louvar & Louvar, 1998). It is used

- to identify agents that are potentially hazardous to individuals or the environment,
- to identify specific populations at risk
- to identify paths of exposure, and
- to quantify the exposure.

Exposure assessment can be performed by analogies, monitoring and modeling. Applying modeling to chemical exposure assessment, a further objective can be added to the list (Trapp & Matthies, 1998):

- to determine present and future concentrations in abiotic and biotic environmental segments.

With including natural variability in time (e.g. flow duration curves), mathematical modeling with its ability to formulate complex processes and the opportunity to implement these as computer programs, gets a valuable method of exposure assessment.

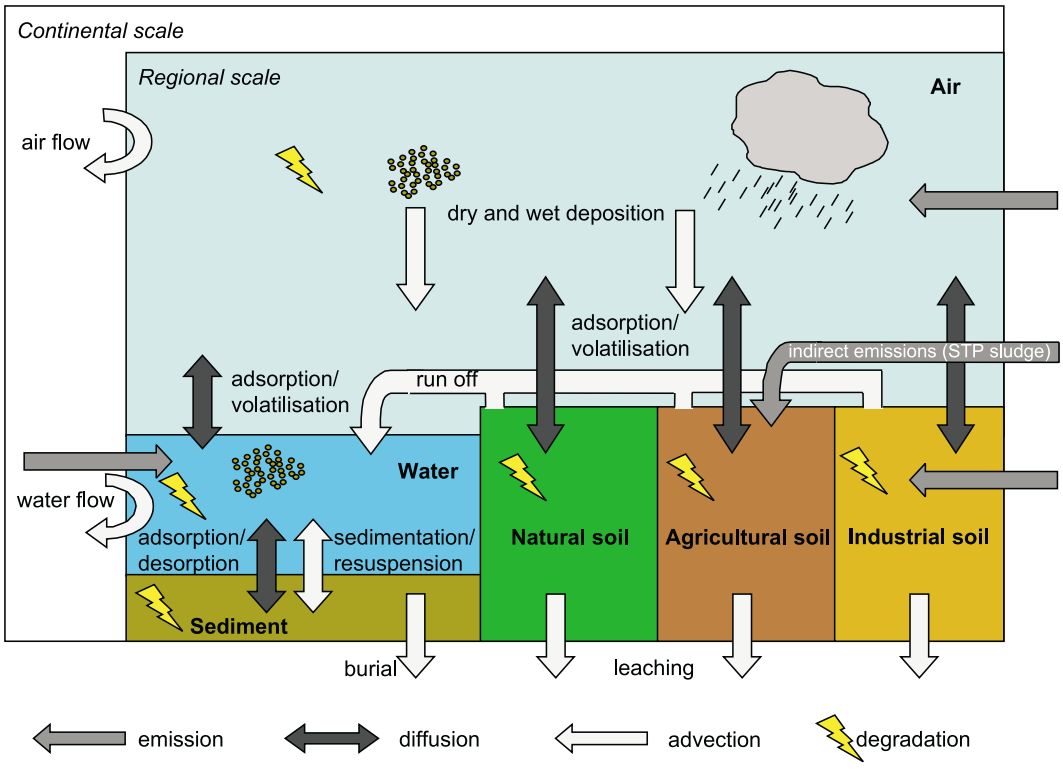


Figure 1.2: *Multi-compartment approach*

Processes such as advection, degradation, partitioning, diffusion and dispersion form the mathematical basis of models for chemical exposure assessment. No matter how detailed such models are, they always remain simplified views of the real environment.

A commonly used approach is the multimedia model (figure 1.2) in which the environment is divided into a number of compartments reflecting a reasonable real-world ratio of water, soil, air, sediment etc. (Mackay, 1991). Models of different complexity estimate chemical fate and behaviour in this simplified world (Mackay, 1979):

- Level 1: Equilibrium, no reactions, closed system
- Level 2: Equilibrium, open system, reactions, steady state
- Level 3: Non-equilibrium, open system, reactions, steady state
- Level 4: Non-equilibrium, open system, reactions, non-steady state

The EU Technical Guidance Documents (TGDs) have adopted the multimedia approach and describe a generic EU region for the evaluation of substances (EU, 1996).

While this methodology is suitable for the screening tier, averaging for the generic region implies unrealistic scenarios for many chemical release and exposure patterns.

One example of this is the ”down-the-drain” path (figure 1.3) of consumer products like detergents (Feijtel et al., 1997): The generic EU region assumes 70% of all waste water discharges as treated and 30% as untreated. This causes two problems: Actual regions such as Germany treat over 95% of their waste water and hence the in-stream concentration of substances would be over-estimated. On the other hand, for regions with a low level of treatment (e.g. Italy: less than 40%) the in-stream concentration is under-estimated.

Problems similar to the above-described scenario can be dealt with within the given TGD methodology and its associated software European Union System for the Evaluation of Substances (EUSES), by repeated use with different refined data. Applying the same model formulations, downscaling from a one-box generic EU region to a 25x25 km grid with local data for each cell, comparison with the calculated maximum values lead to values differing by factor of 10 (air compartment) to 100-1000 for water and soil comparment (Klepper

& den Hollander, 1999). It has been shown that this simple box model may not be able to provide good local predictions, but is able to provide spatial distributions of the concentration.

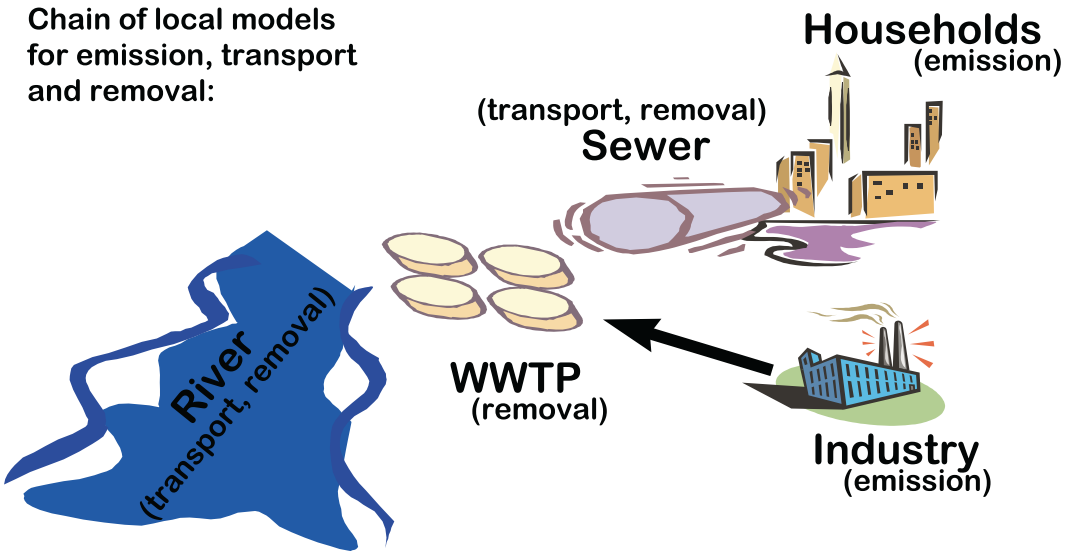


Figure 1.3: *"Down-the-drain" path of consumer products*

However, the ultimate goal to have better support for the management of such situations is only reached with a level of refinement being equal to a spatial explicit representation of geographic objects of the region under consideration. This basically means that the environment of a region is split into its geographical elements such as rivers, agricultural soil, etc. The chemical fate is then simulated with explicit models for each element type of mutually exchanging mass flow. Interconnection is defined by the actual geographic position and / or extent. For this refinement, there is no pre-defined separation into elements (such as a squared grid). The separation must be chosen with regard to the given task.

Concerning the "down-the-drain" example, only such spatially explicit models offer the potential to identify the presence of single sites at high risk ("hot spots"). Naturally, also the location of the sites is identified and hence this modeling approach provides good support for decision-making.

1.4 Generic vs. spatially explicit simulation

In the present situation, tools for risk analysis have the potential to evolve into tools for the powerful management of risk and related issues. The EU directive for establishing a framework for Community actions in the field of water policy (EU, 1999) implies a strong need to make use of the potential and to create an integrated complex management tool. The development of such a management tool would lead to numerous single tasks that were previously separate being joined together. A practical management tool relies on site-specific rather than averaged information. In consequence, a migration from generic to spatially explicit data appears to be a fundamental requirement. The benefits and disadvantages of both simulation approaches need to be considered in order to ensure reasonable migration.

”Generic” data usually describe a situation that was intended to be non-existing. In many cases, averaged data are applied with the intention to gain most adequate data, but (where this is done) they are often not declared as generic. The term ”generic” refers to data only. Model formulations are not intended to reflect a non-existent situation (though they naturally do due to the general limitation of modeling the real world) and hence there are no ”generic models”. Nonetheless this term is often used because the term ”model” can represent a complete entity covering model formulations and (generic) data.

In this context generic data are usually applied with two types of models, regional and local ones (EU, 1996 and ECETOC, 1994). Generically driven regional box models estimate large-scale spatial averages, maximums and distributions. Generically driven local models are based on average site properties and result in site-typical averages, maximums and distributions. In general, generically driven models can be applied where

1. generic results of simulations are sufficient (e.g. identification of compartments of concern, comparison of substances),
2. reality reveals no extreme variation in parameters that are averaged,
3. insufficient detailed information is available or
4. information might be available, but cannot be adequately aggregated.

A spatially explicit simulation depends on highly detailed and adequately aggregated information. The need for highly refined geographical data is increasingly satisfied by the current development of digitally mirroring the real

world. Spatially explicit exposure assessment is already usable in practice for regions where several interests have jointly resulted in comprehensive data collections (e.g. surface- and groundwater data to support water management in dense population areas). Future plans will extend the data collection to a harmonized country level and ultimately cover the entire globe within a few years. In general, spatially explicit simulations can be applied where

1. validation via monitoring for explicit sites is desired,

2. precise management support is required (e.g. what-if scenarios: adding a sewage treatment plant at a certain site),

3. vulnerable regions rather than vulnerable compartments need to be identified or

4. a low number of point emissions needs to be simulated adequately.

1.5 GREAT-ER

The GREAT-ER project (Geography-referenced Regional Exposure Assessment for European Rivers) was aimed at developing a refined method for exposure assessment of ”down-the-drain” chemicals. On the basis of a river network topology, rivers are divided into a number of segments. The processes of emission (via sewage treatment plants), dilution, advection and several eliminations are considered for each segment. The core of the system consists of a steady-state, deterministic model which computes the simulated concentrations (C_{sim}) for each segment. For this a hybrid approach with a stochastic simulation on top of the deterministic model is applied (figure 1.4).

Real-world data with their spatial and temporal variability, as well as their uncertainty are used. By multiple applications of the model with randomly varied input data (Monte-Carlo simulation), geo-referenced distributions of the chemical concentrations in the environment are obtained.

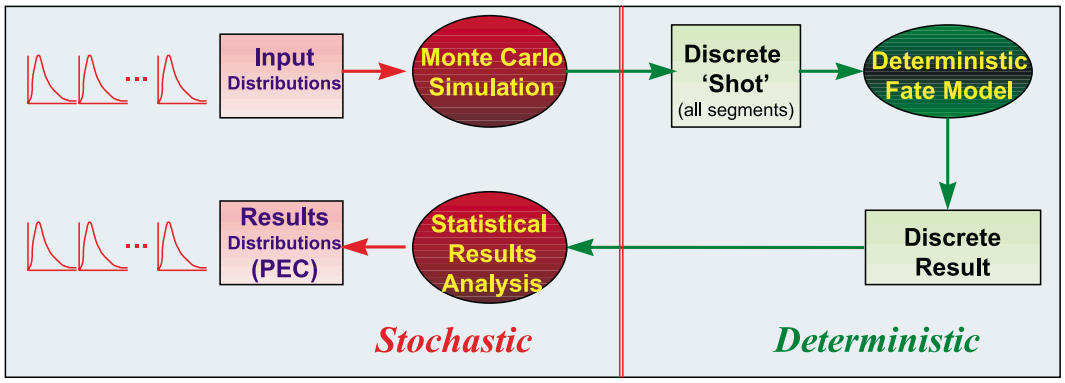


Figure 1.4: *Stochastic simulation on top of deterministic models (Boeije, 1999)*

The pilot study areas in the United Kingdom (Yorkshire, sub-catchments of the River Ouse) and in Italy (Milan, Upper River Lambro) are digitally processed and comprehensive monitoring campaigns for the detergent substances LAS and boron are performed. The monitoring data are used for calibrating and analysing the model system.

Additionally, a methodology is developed to aggregate local C_{sim} results to Predicted Environmental Concentrations (PECs) characterizing the investigated regional scenario (Boeije et al., 2000): $PEC_{initial}$ (unweighted aggregation of concentrations just downstream from waste-water emissions) and $PEC_{catchment}$ (weighted aggregation of all average stretch concentrations).

The GREAT-ER project is conducted by the following partners:

- Institute of Environmental Systems Research, University of Osnabrück, Germany: Geographical data methodology (GIS data processing), development of the software "GREAT-ER 1.0", incorporation of other partners modules.
- University of Ghent, Belgium: Chemical fate modeling.
- Institute of Hydrology, Wallingford, United Kingdom: Hydrological modeling, hydrological data collation.
- University of Milan, Italy: data collation and monitoring for the Lambro catchment.
- UK Environment Agency: Monitoring for Yorkshire catchments.
- European Center for Toxicology and Ecotoxicology of Chemicals (ECETOC), Brussels, Belgium: Project management, support for monitoring.

1.6 Aims of this thesis

The present thesis complements the joint work undertaken in the GREAT-ER project. With the topic of *chemical fate prediction for use in geo-referenced environmental exposure assessment* (Boeije, 1999) the actual simulation methodology is developed. The present study covers the software prototype design, the composition and processing of geographic and associated data and, most essentially, application and analysis with the developed tool. The forthcoming thesis of Frank Koormann will focus on selected problems of this type of geo-referenced exposure assessment and discuss add-on developments.

The approach of a geographically refined chemical exposure assessment in support of the ERA scheme is firstly undertaken with the GREAT-ER project. Besides the model adequacy itself, the huge data demands and data outputs require both a sound data preparation routine and a sound results analysis methodology.

A number of questions arise in respect to the development of a tool that considers the spatial aspects of chemical exposure assessment for river networks:

1. How can a *usable* tool be constructed for the aspired purpose?
2. How can input data of a certain quality be prepared?
3. Which questions can be answered with this tool?
4. Which questions will remain unanswered?
5. How can simulation results be interpreted?
6. How can the quality of simulation results be improved?

Chapters 2 and 3 of this thesis deal with the first question. Chapters 4 and 5 describe the data incorporated in GREAT-ER 1.0 and discuss the pre-processing of geographic and corresponding attribute data, providing an answer to the second question. The sixth chapter deals with a full application of the tool and an analysis of the simulation results, thus answering questions 3 to 6.

Chapter 2

Concepts of and requirements for GREAT-ER

The objective of the development of the GREAT-ER prototype was to combine a scientific research tool with an applicable decision support system for use in the refinement tier within environmental risk assessment.

This chapter provides an overview of the concept and requirements of the GREAT-ER software, based on the intention to construct an open system that can be further developed as a research tool. At the same time this software tool is a specification for a later professional software product proving its scientfic soundness, reliability and applicability.

2.1 Introduction

The handling of all aspects of exposure assessment on a spatial scale requires a tool that is powerful and complex enough to adequately cover the given task, but which is also simple to use. Whether such a tool is deemed applicable depends on the type of application. A regulatory or commercial decision support system does not need extensive options to substitute modules of the system or perform in-depth data changes. On the other hand, a research tool does not necessitate a standardized, fixed and visually perfect user interface.

The design and development of GREAT-ER aims at considering both types of application. Obviously, this can not be accomplished entirely. Instead, a

prototype needs to be developed which is open to further research needs and at the same time forms a basis to derive a professional tool. The intention of this hybrid approach is to enable the tool to support gaining of new scientific knowledge and understanding. Concurrently it is intended to have it provide stable and scientifically sound results for political discussion and decisions.

There are numerous tasks that a spatially explicit exposure assessment tool should cover in general:

- a user interface for specification of non-georeferenced data
- a user interface for visualizing and exploring geographic information and all data associated with geographic objects
- options to manipulate any of the geo-referenced data
- analysis tools to support further processing of results
- provide transparency in terms of usability and comprehensive results
- an open approach for later incorporation of further models.

In GREAT-ER an overall modular approach separates the principle tasks and specifies the scope of the end-user tool (figure 2.1): It mainly provides a desktop GIS and the simulation models. The preparation of huge and static (not user-editable) data is kept external.

The end-user system consists of three major components: a desktop GIS (including the user interface), model and data. An appropriate interconnection of all modules and sub-modules needs to be enabled in order for GREAT-ER to fulfil all tasks listed above.

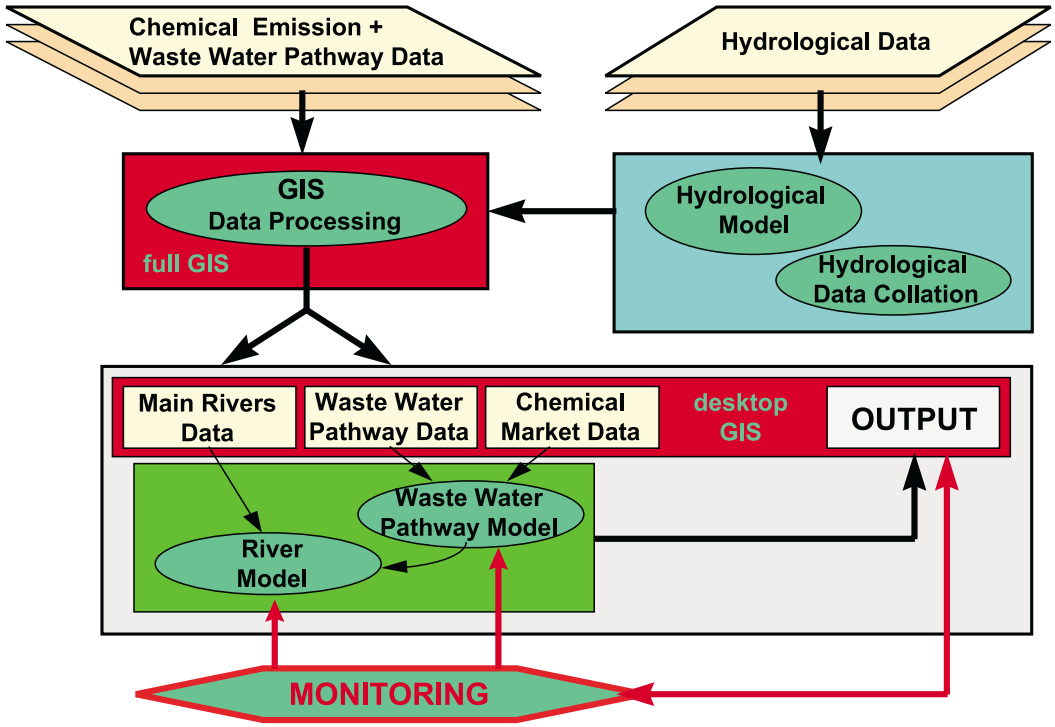


Figure 2.1: *GREAT-ER project: modular approach (Feijtel et al., 1997)*

2.2 Other geo-referenced simulation tools

Other approaches to combine GIS and environmental models have primarily addressed hydrology and water quality issues: BASINS (Better Assessment Science Integrating Point and Nonpoint Sources by the U.S. Environmental Protection Agency), ATV-Gewässergütemodell (Abwassertechnische Vereinigung, German Association for Water Environment), NOPOLU (Beture-Cerec, France), MONERIS (IGB Berlin, Germany) and GESREAU (GEStion des Ressources en EAU, Switzerland). BASINS appears to be the most advanced of these systems. It has incorporated a simple chemical fate model for river networks, TOXIROUTE.

A set of policy and decision support systems with spatial aspects (Geonamica, WadBOS, RamCo and MODULUS) has been developed by RIKS (the Netherlands). The systems share the same basic software modules.

The applications of both groups have their roots in addressing problems different to those addressed by the GREAT-ER project. In principle, several of

these tools have the potential to cover the tasks requested for GREAT-ER. The reasons why GREAT-ER is not based on one of these systems are:

- At the time GREAT-ER was designed and launched, most other tools were not available or not sufficiently advanced.
- GREAT-ER is meant to be a lean prototype, open for further development as a scientific tool as well as a policy product. Implication of further interests which were focused on more remote questions might have reduced the flexibility of the development.
- GREAT-ER is meant to be open in terms of transparency and code availability. This could not be fully attained as part of a proprietary product.

Review of the GREAT-ER concepts is performed by direct communication with project representatives of BASINS and NOPOLU. With an accomplished GREAT-ER prototype a tool focusing on exposure assessment is available. It should then be evaluated whether to directly couple GREAT-ER with other software tools that have proven their suitability for specific tasks (e.g. water quality, hydrology).

2.3 GIS-model coupling

The coupling of a GIS and simulation models is one of the most challenging parts of the software engineering phase.

In the literature several categorisations of coupling methods for GIS and environmental models can be found (Wagner, 1996). In principle they differ only in how precisely the categories are chosen. Most approaches have set up three main categories for the coupling intensity:

The coupling intensity is divided into three levels: *ad hoc integration*, *partial integration* and *complete integration* (Tim et al., 1994). In the first level, GIS and the model are independently developed and the data transfer is realised as a simple exchange of files. At the second level, both components use (at least partly) a common database which belongs to one of the components. This includes the need to internally adapt one component. The third level is an integrated system in which both components can fully access the commonly

used database. In (Wagner, 1996), level 2 is applied to couple a simple water model for chemical exposure with the GIS ArcInfo.

Another categorisation approach distinguishes between *shallow coupling*, *deep coupling* and *embedded GIS* (Fedra, 1996). This graduation is rather based on the user view, cf. the user interface. At an initial level, the GIS and the model are partly independent of each other. Both offer their own user interface but use the same database. At the next level one user interface integrates the two components for the user view and internally direct communication between the components is established. For the highest level the model is a further developed method of the GIS. The definition of the coupling intensity levels is not very detailed and leaves any solution across the categorization conceivable. In several cases level 3 is realised (Fedra, 1996).

Resulting from a review of the literature, a list of five main types of coupling were identified (Wagner, 1996, figure 2.2):

(a) Model realized in GIS built-in language

(b) Coupling interface as a program of its own

(c) Interface between GIS built-in language and model (direct coupling)

(d) Interface between model and database

(e) Embedding

A component approach is recently being favorized for many information technology projects. Software components are modules wrapped up with a standard internal interface for flexible use without need to access the source codes of the applied modules. It can be described as a further abstraction level introducing interfacing technology of its own. Examples are COM (Component Object Model) and CORBA (Common Object Request Broker Architecture). This technology can be applied to types (b)-(e) in order to gain further flexibility.

(a) Model realized in GIS built-in language

The model is reimplemented in the language given by the selected GIS. The connection to the GIS data is established directly.

This means a dependency on the GIS used. In principle a reimplementation does have the potential for error, especially in the mathematical part and hence it must be rechecked carefully. The built-in language is believed to be sufficient for simple models, but complex models may be less efficient or implementation may not be possible at all.

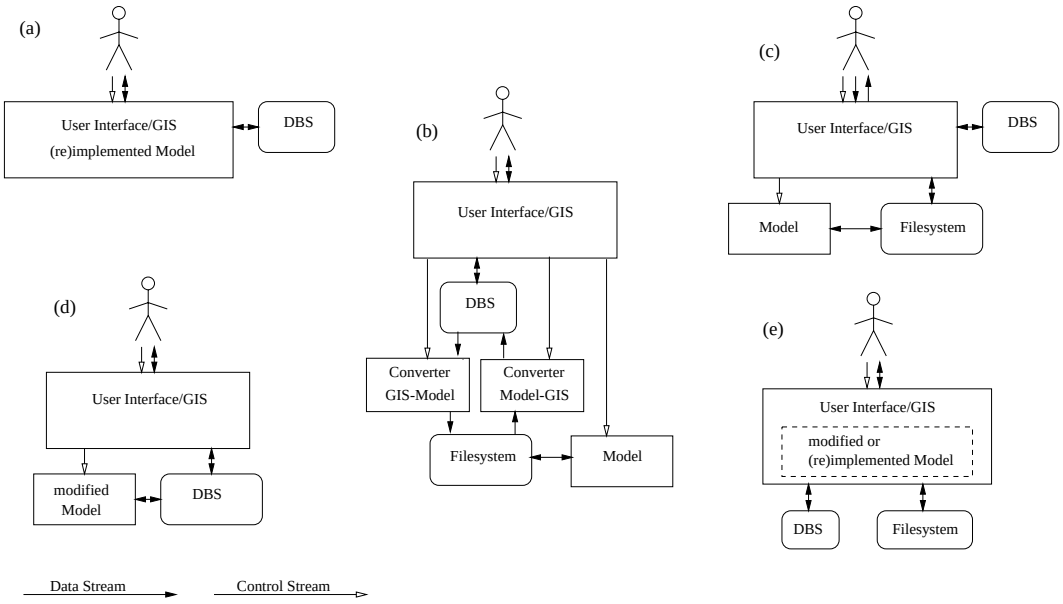


Figure 2.2: *GIS-Model coupling (Wagner, 1996): (a) Model realized in GIS built-in language, (b) Coupling interface as program of its own, (c) Interface between GIS built-in language and model (direct coupling), (d) Interface between model and database, (e) Embedding*

(b) Coupling interface as a program of its own

Two conversion tools prepare the GIS data for the model and vice versa. This concept offers several opportunities for adaptations and has often been realized (Han et al., 1995).

The model implementation is not changed, but any change to the model interface must follow an adaptation of the corresponding conversion tool.

(c) Interface between GIS built-in language and model (direct coupling)

All routines for exporting model input data and for reading model results are implemented in the GIS built-in language. The conversion components are part of the GIS.

The model implementation is not changed. The built-in language must be powerful enough to handle the model's input and output.

(d) Interface between model and data base

The model is extended by routines to access a database which is shared with the GIS. The conversion components are part of the model.

The model implementation is changed and hence the source codes are required. The type of software development tools used for the model may reveal problems in establishing data base access.

(e) Embedding

The functionality of the model is directly linked to the GIS (e.g. as a library file). Embedding can be loose (only one funcion call) or very tight (model split up into several small routines (i.e. sub-models, processes).

Depending on the intensity of embedding the alteration of the model may vary from marginal to almost entirely. In the latter case the model would become an integral part of the GIS.

Coupling method for GREAT-ER

For the implementation of the GREAT-ER prototype, direct coupling (type (c)) was choosen for its simplicity to exchange input / output data between GIS and model component. This solution supports independent development of both components. Furthermore, implementing a prototype with limited ressources should not involve too much development of flexible technolgies, but concentrate on the prove of concept.

2.4 Simulation models

2.4.1 Overview

The simulation models which are integrated into GREAT-ER are each associated with geographical objects and in sequence describe the wastewater pathway from household emission to the end of a river network (figure 2.3).

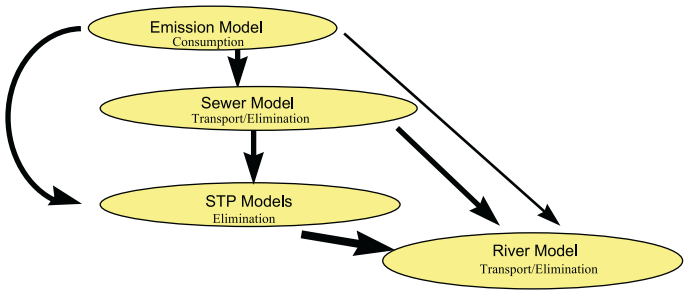


Figure 2.3: *Sequence of simulation models*

All submodels (except emission) share a simple mode in which a lumped first-order elimination or percentage removal efficiency is applied. Higher levels of complexity are available to refine the physico-chemical processes within the modeled compartments.

In general, with each additionally considered process, data requirement increases. Data gathering is usually time- and cost-intensive and should be postponed until a simple approach does not suffice for the required quality. All models / sub-processes are linear approaches with widely known and applied equations (e.g. Trapp & Matthies, 1998). GREAT-ER combines present modeling approaches rather than introducing entirely new model formulations.

For the simulation of the Sewage Treatment Plants (STPs), two models are available: one for trickling filter and one for activated sludge plants. Both plant types are typical for the European situation, although activated sludge plants are more efficient. Some discharges in Europe even remain untreated.

On top of these deterministic models a stochastic module allows to perform Monte-Carlo simulations (figure 1.4). Some of the input parameters are usually given as distributions (e.g. average annual river flows). In GREAT-ER almost all input parameters can be assigned to a distribution. Consequently, the

computed results are also given as distributions of concentration (over time) for each geographical object. It is assumed that these concentrations are distributed log-normally and the models return mean and standard deviation for each object.

The model implementation has a command-line / text file interface that allows either manual or batch execution. The formats of the input / output text files are fixed. Commands and options are partly specified as command-line parameters and partly specified in an initialization file.

The models, their stochastic part and the model implementations are described in detail in (Boeije, 1999).

2.4.2 Substance data

The substance data set contains all substance-specific parameters. The link to a chemical has the highest priority in grouping parameter sets. This essentially means that also any model-specific referring to a chemical property also belongs to the substance data rather than to the model parameters. In consequence, the substance data set is a comprehensive collection of properties that partly overlap in their meaning but which refer to different models under specific assumptions.

For the current GREAT-ER submodel set, the substance data are divided into seven sections: substance identification, physico-chemical properties, partitioning, degradation, sewage treatment removal, in-stream removal and use pattern / market Information.

2.4.3 Emission model

The emission model is a straight-forward calculation of the total chemical mass entering the sewer systems. Three types of chemical release are considered: Regional consumption based on per-capita use and population served by treatment plants, site-specific consumption (replacing regional consumption) and additional input as mass flow (e.g. industrial emission).

2.4.4 Sewer model

Complexity mode 1 assumes no removal in the sewer system. Modes 2 and 3 apply an overall removal percentage.

No sewer properties are used (e.g. travel times, combined or separate sewer system) and hence in cases where site-specific in-sewer removal is required, elimination needs to be incorporated into an overall site-specific STP removal.

Applying the site-specific actual daily flow of the treatment plant, the sewer end-point concentration is provided, which at the same time is the influent concentration for the STP models ($C_{sim,influent}$).

2.4.5 STP models

In complexity mode 1 both, trickling filter and activated sludge models apply an overall removal percentage.

For trickling filter plants no refined model is available. Modes 2 and 3 are identical to mode 1.

For activated sludge plants a modified version derived from SimpleTreat (Struijs, 1996) is used for modes 2 and 3 (Boeije et al, 1998). It is a mechanistic, steady-state model including a primary and secondary settler. Modification to SimpleTreat includes the consideration of single-sludge biological nutrient removal.

Besides the actual STP model, a fraction of direct emission can be specified for any site. The corresponding fraction is added to the output of the treatment plant ($C_{sim,effluent}$).

2.4.6 River model

A first-order in-stream removal with fixed rate coefficient is used for complexity mode 1.

In mode 2, chemical sorption is considered. The steady-state sorption / dissolved fraction is used to specifically calculate the effects of the subprocesses

volatilization, degradation and sedimentation. All subprocesses are dealt with by fixed (pseudo) first-order rate coefficients.

With mode 3 individual mechanistic models can be selected for different processes (biodegradation, hydrolysis, photo-degradation, sedimentation and volatilization). A selection can be made in order to model a specific substance adequately (e.g. volatilization only for highly volatile substances).

The river model outputs are the concentrations for the beginning of each river segment ($C_{sim,start}$), the end of each segment ($C_{sim,end}$) and an average concentration for the entire segment ($C_{sim,internal}$).

2.4.7 Stochastic Monte-Carlo module

Most parameters of the GREAT-ER simulation models can be given as distributions. The river flow statistics can be regarded as one of the most essential distributions. Discrete samples are taken from the distribution curve for each simulation. For the river flows one sample is taken as a percentile to simulate the entire river network to ensure consistency of flows (e.g. the 37th flow percentile is applied throughout the whole catchment). Applying numerous discrete samples results in distributions for the simulated concentrations ($C_{sim,x}$). This number of Monte-Carlo samples is an overall parameter applied to all submodels.

A set of correlation factors as part of the catchment properties supports the stochastic module. Correlations can be defined for temperature $\leftrightarrow$ flow, windspeed $\leftrightarrow$ flow, suspended solids $\leftrightarrow$ flow, dissolved oxygen $\leftrightarrow$ flow, biological oxygen-demand $\leftrightarrow$ flow and suspended solids sedimentation $\leftrightarrow$ flow.

2.5 User interface

2.5.1 Analysis and visualization requirements

The enormous amount of data required and produced by GREAT-ER sets forth needs for adequate methods to analyze and to visualize these data. Some user requirements are also incorporated into the following list:

- visual analysis via interactive browsing through 2-dimensional maps of the area of interest (e.g. zooming, visibility selection of geographic layers, shape / color selection for geographic objects),
- opportunities to visualize temporal variability of data (river flows, chemical concentrations),
- standard plots and diagrams (e.g. scattered plot, line plot),
- standard table operations,
- downstream profiles (e.g. concentration, flow, flow velocity),
- aggregation of results to express regional summaries,
- detailed review of any simulated site / geographic object and related information / attributes.

2.5.2 Selection of software components

Operating System

The potential user group of GREAT-ER (authorities, chemical industry, academia) should easily be able to apply the system without needing to change their primarily used operating system. Academia is usually the most flexible group using a variety of systems. The other two groups have less opportunities and at the start of the project these groups used primarily Microsoft Windows-operating systems (versions 3.1 and 95). With a 3-year project duration it was believed that Windows NT 4.0 (or compatible systems) would be the standard. Hence Windows NT 4.0 was selected as the prime requirement for an operating system.

GIS

GIS capabilites within the GREAT-ER project are required in two different ways. On the one hand, geographic data have to be prepared for use. This processing stage is more complex for raw data than for initially processed data. In most cases further processing is needed and requires comprehensive GIS capabilities. On the other hand, a desktop GIS primarily offering a map-based visualization takes over parts of the user interface.

As for the complex GIS software, most of the commonly available products are sufficient. The system can be chosen regardless of the aspired end-user operating system because only transferable data and not programs are created. ARC/INFO was selected as a readily available system. The Free Software product Geographic Resource Analysis Support System (GRASS) is currently an adequate alternative but in 1996 it was undergoing a process of reorganisation.

In 1996 the availability of powerful desktop GIS systems for Windows NT 4.0 was quite limited. In absense of comparable alternatives, the ArcView software by ESRI was felt to be the most suitable solution. ArcView is broadly used in educational institutions, governmental organizations and industry. It is assumed that ESRI will guarantee support and further development of ArcView in the future.

ArcView offers a number of data interfaces, of which the most important are text files (read-only tables), dBase files (read/write tables) and an SQL interface. The text file interface lacks flexibility some-what, e.g. in the selection of column delimiters. Furthermore, ArcView has a command line interface via its built-in language Avenue. Both synchronous and asynchronous execution of external programs are possible.

Chapter 3

Realization of GREAT-ER

The intention of this chapter is to give a detailed insight into the tool's technical functionality. However, it should not be viewed as a user's manual.

The realization of GREAT-ER considers principles of good modeling practice (Trapp & Matthies, 1998) to support quality assurance in terms of transparent and comprehensive simulations.

All source codes are available on request. The contents are summarized in a number of project progress reports to ECETOC and to some extent published in ECETOC Special Report No. 16 (ECETOC, 1999).

3.1 Technical Overview

3.1.1 Software requirements

The end-user prototype version of GREAT-ER requires a Windows NT 4.0 platform and the corresponding version of ArcView 3.0 (or 3.1). All other programs used within the prototype are installed together with the GREAT-ER system.

For the GREAT-ER development a number of tools are required or are at least helpful: A number of small standard Unix command-line tools (e.g. make, awk, sed, join) are used within the process of data generation (see also Chap-

ter 5). The initial data processing, which is performed on a Unix platform for the GREAT-ER prototype, might also require a full GIS (ArcInfo is used for the prototype data sets). A dynamic link library (DLL) including all special dialogs is created with Microsoft Visual C++. Some additional and more specific tools (txt2dbf and the simulation software itself) were developed in the programming language 'C'. Any standard compiler should be able to re-compile the sources. For GREAT-ER, the GNU C Compiler (gcc) is used.

3.1.2 The desktop GIS and its built-in language

The core of GREAT-ER is a number of Avenue scripts which are executed within ArcView. They are not incorporated into the GREAT-ER ArcView project file. This project file only has the ability to load a startup Avenue script (a text file) from the filesystem which will then direct any further initialisation. All other scripts are loaded within this process and the user interface is customized for GREAT-ER.

A special run-time element is established by a script ('sendEvents.ave') which regularly executes itself and each time executes the update scripts which are attached to *views*. *Views* are windows representing a scenario and displaying the catchment. Update-routines will mainly check for results from tools outside the ArcView system, e.g. whether a simulation has finished. This technology supports the user by concurrently running simulations and automatically updating views.

Several global variables are used to share entry points to run-time information. All these variables are initialized in one script (initialize.ave).

Avenue is described as an object-oriented language but does not allow the creation of new classes. Only pre-defined classes can be used. One of these is intensively used to build up the GREAT-ER data structures: the dictionary class. Dictionaries allow to collect data of any type associated with a key. Strings are used as keys. For example, a collection of model parameters is a dictionary which the key is the name of the parameter. Associated with each key is another dictionary with some pre-defined entries to be used. These are 'value', 'DistriTyp' and 'Comment' among others. Associated with 'value' is a string which is to be interpreted as a number if it is not equal to the text 'unknown'. 'DistriTyp' identifies the type of distribution which is coded as an integer. 'Comment' is associated with a text object.

3.2 GIS-model coupling

For the prototype development of GREAT-ER, *direct coupling* appeared to be the most adequate methodology:

- It allows concurrent and (almost) independent development of the GIS and the model part.
- Scientific development takes advantage of the ability to perform rapid tests of modifications.
- Independent performance optimization of the model is possible.
- GIS and model implementation remain exchangable.
- Compared to the expected usual simulation effort (Monte-Carlo approach) the performance loss by converting input and output data is surmounted.
- It is feasible with the given selection of GIS and the operating system.

Direct coupling simply means that the GIS 'knows' the data requirements and structures of the model and also 'knows' how to execute it. This knowledge has been coded in Avenue scripts. However, improvement towards a decision support system would benefit from a tighter coupling in a more compact tool.

For the execution of a simulation, GREAT-ER performs the following steps:

1. Check completeness of the data. The user is given a hint if any parameter has not yet been specified.
2. Check if this scenario is already running as a background process. Two processes working on the same scenario data set would interfere with one another.
3. All required data are written onto the file system. The corresponding Avenue routine creates almost all model input files in the correct format. However, for a better performance large data tables are first created applying a quick internal procedure of ArcView. Then awk is executed for the final reformatting.
4. A lock is set for this scenario to avoid double execution.

5. Results of previous simulations are removed.

6. A batch file is executed as a background process which first starts the simulation software, waits until it is finished, checks for errors that may have occurred and, if such were not found, a reformatting routine prepares the results for ArcView (awk, `txt2dbf`).

7. The lock is removed.

All running scenarios are checked regularly to see if the computation has finished. When a simulation finishes, the results are loaded into the system for manual and automatic access.

Any change to the input/output structure of the simulation software requires an adaption of the Avenue routines.

3.3 User interface development

3.3.1 Review of ArcView

For confirmation purposes the adequacy of the selected GIS tool for the planned purpose is tested in detail. The results of the review also facilitate an improved estimate on several aspects of the user interface design and development, ultimately uncovering semantic limitations or opportunities.

- ArcView offers a graphical user interface with many built-in features for the visualization and analysis of maps, geographic data and corresponding attributes.

- A built-in macro language (Avenue) supports the customizing and extension of the desktop GIS's user interface. Furthermore, access to most elements of the GIS is offered.

- Basic relational database management features are offered.

- Connectivity within a Windows NT environment is offerd via several communication modules (e.g. DLL, DDE, SQL, ODB, etc.).

- It is not possible to define submenus. This increases the number of main menus and menu items if a lot of functions and commands are to be accessible via the menu. This disadvantage was considered in planning the GREAT-ER menu structure.
- It is not possible to create complex custom dialogs. The macro language Avenue offers the class 'MsgBox' which contains some basic methods such as 'OK' and 'Yes / No' dialogs. Combinations of i.e. text entries and so-called radio buttons are possible with the ESRI Dialog Designer Extension for ArcView, but this extension has also limitations. This is a major disadvantage, because the parameter sets of the model or the substance properties necessitate user-friendly and more complex dialogs. Alternative dialog techniques had to be investigated.
- Some parts of GREAT-ER can not be implemented based upon Avenue (e.g. the model, some tools and complex dialogs). Two connectivity technologies were chosen, Dynamic Link Libraries (DLLs) and direct command execution. The latter reveals some problems, one of which is the missing return value (success / failure indicator).
- The design of Avenue reveals a major weakness concerning compatibility to other versions of ArcView: Though the version number can be tested in Avenue and conditional command blocks can be followed in order to always use the correct class methods, the concept of enumeration types already inhibits the compilation of scripts.

3.3.2 User interface modes

Since the desktop GIS ArcView forms the basis of GREAT-ER, numerous GIS and data management functions are provided. To make the GREAT-ER part transparent and straightforward to use without losing the powerful functions mentioned above, it was decided to offer two user interface modes: the expert mode and the easy-to-use mode.

'Expert' in this context means that users are familiar with GIS in general and with ArcView in detail. The expert mode gives access to all ArcView operations and commands. This includes the possibility of undocumented changes or to delete basic GREAT-ER data sets. Obviously, this feature contains additional functions but its use can also be harmful. Users should decide carefully whether to perform operations in the expert mode or not. An expert user should be able to decide which kind of operations are not able to harm the whole system and its data sets.

'Easy-to-use' basically means that it is impossible for users to harm the system and/or the underlying GREAT-ER data sets. If users are only interested in performing simulations and viewing results, this mode is perfectly sufficient.

3.3.3 Scenarios

An adequate concept is developed for the management of GREAT-ER data. A 'scenario' is defined as the collection of all data needed to perform a simulation.

Scenarios can be stored and loaded and thus offer the quick retrieval of a certain situation. A unique title for each scenario assists the search of a specific one in a long list.

Additionally, scenarios form the basis for the exchange of complete simulation data sets (export/import).

The most important information on a scenario is summarized in the title bar of the corresponding view window. This makes the title, catchment ID, substance ID, modification state of scenario/substance and the simulation state permanently visible and thus prevents users from becoming confused when concurrently managing several open scenarios.

The results of a simulation are also actively part of a scenario. This means they are stored and retrieved when saving and loading scenario data sets. The results of simulations that take a long time to perform can be stored and loaded quickly.

All data related to a scenario are stored in one directory of the file system (as a subdirectory of the global scenario directory specified by the environment variable 'GSCENARIOS').

3.3.4 Substance data

Substance properties are a central data set of GREAT-ER. The physico-chemical properties of a substance determine its behaviour and fate in the environment. Many of the model's processes require these substance parameters.

GREAT-ER offers a complete substance database management. Each database is stored in one file and it is possible to change the database via a menu command. The substance data of a database can be managed with standard functions (Save, Delete, Open, etc.). The default substance database delivered with GREAT-ER is write-protected.

It is not possible to decide which substance properties should be included in the data set, taking into account the flexibility of having a sufficient data set even for other (as yet unknown) models that might be integrated into GREAT-ER. It was decided to include only the parameters that are requested by the current version of the simulation software. The Object Data Base (ODB) format for substance data management is used in a flexible way: All value entries are objects which, besides a numerical value, have an ID and further information (e.g. comment). It is possible to add further properties to any value (e.g. unit) and to add further substance properties (e.g. certain partition coefficients). This can be done for just one entry in the database without injuring the database consistency or losing compatibility with older database management routines (i.e. dialogs). However, it is of course necessary to perform source code changes in order to add any new structure item that is to be considered by GREAT-ER itself.

Connectivity to other exposure assessment tools and databases (i.e. EUSES, IUCLID) can be added. IUCLID offers access via an SQL server, and ArcView also offers SQL connectivity. This ensures adequate interconnection of both products in principle.

3.3.5 Alternative dialogs

Two main possibilities to integrate special dialogs were identified. The first method is to implement the dialogs with a high programming language and to create a Windows Dynamic Link Library (DLL) to be started from Avenue scripts. All usual software development systems offer an easy method to create dialogs but internal actions (e.g. range-check functions) have to be programmed as well as the data transfer from ArcView to the DLL and back. This data exchange is based on a datapool within the DLL which is loaded and read by Avenue routines. Structures for the datapool parameters are analogous to the dictionaries used in the Avenue scripts. The chosen method is a Windows-only solution. DLL connection is performed with the integrated software development system 'Microsoft Visual C++'.

3.3.6 Map projections

The displayed maps must have different projections depending on the size of the displayed areas. Small regions can be projected assuming a 'flat earth' (tangential approach). For bigger regions this will produce increasing distortion the larger the displayed map is. Universal Transverse Mercator (UTM) is one of the optimal projection methods for small areas. Europe is divided into numerous UTM zones of 6 degrees range in longitude. Areas covering more than 6 degrees in longitude are better displayed using, e.g. Lambert-Azimuthal projection to consider the global shape of the earth. However, ArcView offers many projection methods but for conversion of a projection it is necessary to have the data in geographic coordinates. This means that for both visualizations it is necessary to transform the conversion to either UTM or Lambert-Azimuthal projection. This is a time-consuming process for ArcView ranging from seconds to a minute, and can be irritating when zooming in and out and shifting the map in any direction. The only acceleration can be obtained by making the corresponding (background) coverages invisible by default and switching them on for a print-out, demo, etc.

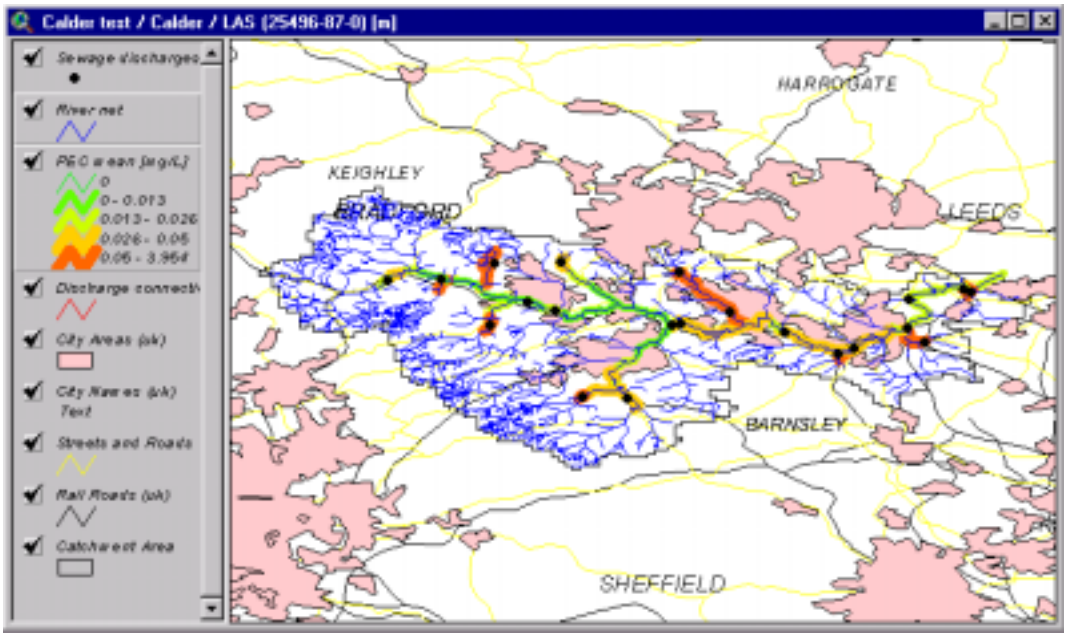


Figure 3.1: *River Calder: Geographic projection*

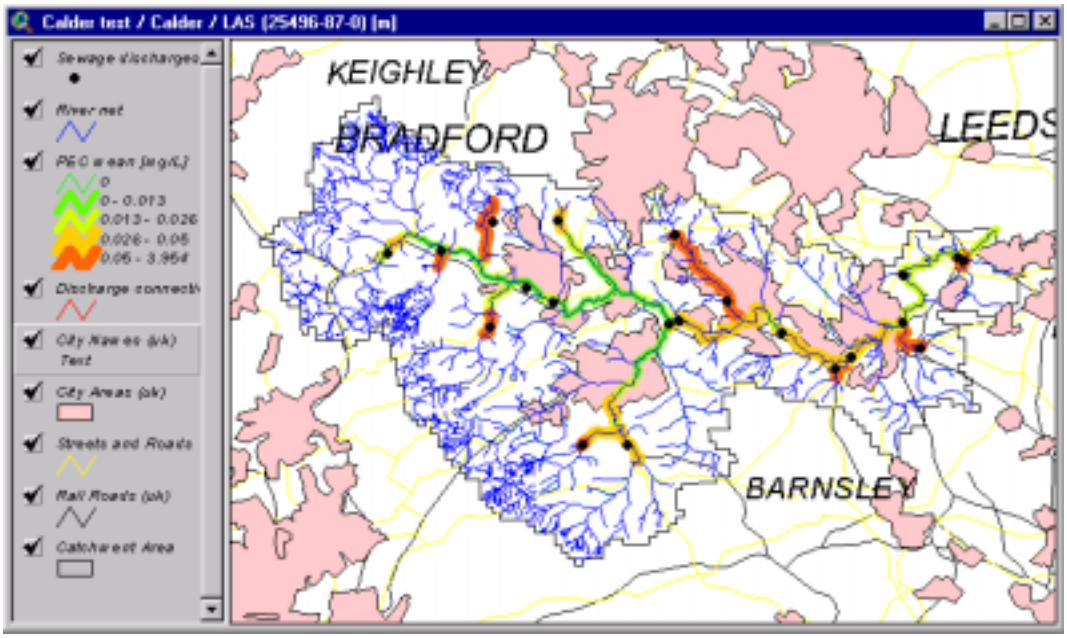


Figure 3.2: *River Calder: UTM Zone 30 projection*

3.3.7 Market data

Market data are related to areas (only catchments, not political or other areas) or points (discharge sites). These data describe consumption (mass per capita and time) for points or areas, and input (mass per time) for point discharges only.

Market information is substance-specific and therefore is stored as a substance property. This is carried out in two lists, one containing the catchment-related market information, and the other containing discharge-specific consumption / input.

The data for discharge sites are tupels: first the consumption, which will override the value taken from the catchment default for the specific discharge, and second the input, which is additional to the consumption. The user can specify either none, one or both values.

This is realised by simply clicking on a discharge and editing the values in the dialog. The expert mode also offers methods to edit the underlying tables directly.

3.3.8 Menu structure

The menu structure of the GREAT-ER user interface is designed in view of a typical simulation session based on the scenario concept: The first step in a session should be to create a new *scenario* or select an existing one. The first object under investigation is the *substance* which can be modelled within several *catchments*. After selecting and editing a catchment and its substance the *model* has to be set up with different parameters and finally started. The simulation results can be analyzed with some *analysis tools* (e.g. PEC calculations) and visualized with several *display* options (e.g. additional background data). Finally, the user interface provides access to the *help system*.

3.4 Incorporation of expert knowledge

Software tools are used for the prediction of environmental concentrations applying mathematical models which ultimately may support ecological and economic decisions. Besides a sound scientific base for the model itself, implementation of the user interface must offer the highest quality assurance adequate for the corresponding task.

The importance of this quality aspects lies in the fact that decision support tools are no longer applied by the model developers themselves nor by other experts. Hence users do not have the expert knowledge that is important for correct application of the tool. In the following, some features of quality assurance as realised in GREAT-ER are described. First approaches towards quality assurance were discussed and realised for the tool CemoS (Baumgarten et al., 1998 and Trapp & Matthies, 1998).

3.4.1 Error and warning ranges

All entered values are compared against two ranges: a warning range and an error range. The warning range specifies the usual range of a parameter. If an entry exceeds the parameter warning range, a dialog will point out this possible mistake. Nevertheless, it is possible to run a simulation with such settings. In contrast, the error range defines the physical range of a parameter. Values outside this range are impossible and will lead to errors. If an entry exceeds the error range it is not possible to leave the dialog until a value within

the warning range has been entered. A dialog will give an error message and hints on the logical range of the parameter.

3.4.2 Default values

The warning range gives non-expert users a hint to avoid senseless values. Initially, a user might not be able to insert a reasonable value. In this case he can use the default value defined by the model developers.

3.4.3 Parameter requirement indicator

The underlying simulation model offers several switches for the selection of complexity modes. Higher modes consider additional processes which require the input of further parameters. All parameter dialogs offer an interface for editing values of all modes. This may lead to a large number of entries which may confuse a non-expert user trying to perform a simulation in the simplest mode. The model documentation provides information on the actually required parameters. However, this information can also be incorporated into the user interface, as is the case with GREAT-ER:

In any dialog, parameters color-coded in green are definitely not used by the current model selection. Nevertheless, it is possible to edit these parameters.

Black parameters *might* be used within the current model selection. Parameters definitly *not* required can not all be identified for the underlying simulation system of GREAT-ER, because it offers to specify different model complexities for each geographic object (i.e. river reach). An exact prediction for all parameters would necessitate scanning all corresponding datafiles, which might exceed acceptable feed-back time for users.

3.4.4 Parameter commenting

Simulation results can only be trusted for official purposes if the performed activities are completely transparent. For assessing environmental exposure, several parameters occur which can not be measured in laboratory conditions or which reveal an unclear definition.

Almost any parameter in GREAT-ER can be provided with a user comment. This comment should mention the data source and might contain additional information on laboratory practices, etc.

3.5 Visualization of results

A new level of complexity is reached by the interconnection of geographic reference and stochastic simulation. This creates the need for comprehensive visualization methods to make simulation results interpretable and comparable.

Three basic methods are realised for GREAT-ER. These form an entry platform for screening and in-depth analysis.

3.5.1 2-D percentiles

The spatial aspects of the results are based on the corresponding maps. The visual impression of single geographic objects can be extended to display certain information. For river networks, geographic objects are lines representing river reaches. The concentrations can be displayed as text next to each object, but this is feasible for only a few geographic objects displayed at the same time. Besides showing the actual value as text, barplots may also indicate the concentration visually compared to a specific value or to the highest occurring concentration.

However, the simulation results for single geographic objects are distributions based on the natural variation in time. Displaying the entire distribution curve will most likely overload the map with information. The criteria of the curves need to be selected to display only one value per object. The mean value is used as a default. The user may then select any percentile of the distribution.

3.5.2 2-D classification

A very usable integration of information associated with geographic objects is the modification of the object's appearance itself. In GREAT-ER, each river

reach is color-coded with a concentration selected from the distribution curve. Color-coding only makes sense with a small number of different colors which leads to the need of (spatial) classification.

The applied classification is based on the frequency of concentration occuring in geographic objects (river reaches). Since the reaches are not equally long, the accumulated length of the color-coded classification does not reflect the spatial fraction of concentrations based on the total river network length: Consider a river with two reaches, one with a length 1 km (class: low concentration, colored in green) and 99 km (class: high concentration, colored in red), respectively. Though the two classes are of equal size (each has one element), the visual impression is a ratio of 1:100.

A free selection of the classes is offered where the classification, which is a percentile value of the distribution curve, can also be chosen. The resulting spatio-temporal color-coded classification allows simple identification of sites exceeding certain criteria, cf. hot-spots.

3.5.3 1-D stream profiles

A more classical visualization method is the downstream profile for concentrations. Several percentiles can be plotted together in one graph. This type of result visualization is the most adequate one for comparing simulation results and monitoring data, because the latter are usually only available for a small number of sites rendering 2-D representation unfeasible.

In GREAT-ER any river reach can be selected for a profile plot. The river attributes for the graph need to be specified and are then collected by an algorithm travelling downstream.

Chapter 4

Data incorporated in GREAT-ER

All data described in this chapter are incorporated in the GREAT-ER 1.0.1 distribution (ECETOC, 1999).

4.1 Substances

GREAT-ER is delivered with two substances in the chemical data bank, boron as an inert and LAS as a readily bio-degradable chemical. Both are used as detergents and hence have a widely dispersive use pattern. Boron is additionally used in the photo-chemical, glass/ceramics, metal and paper industry as well as in agriculture. It also appears at a geological background level, but the amount used for detergents is assumed to surmount all other types of exposure. LAS is almost exclusively used as a household chemical (about 90%). Table 4.1 summarizes the properties of boron and LAS required for model complexity mode 1.

4.2 River networks

The GREAT-ER system contains several catchments of which the largest and at the same time most detailed catchments are located in the Yorkshire area of the United Kingdom. This area was also selected for numerous third-party environmental pilot studies from which comprehensive additional data are available. Expert knowledge is offered by local authorities.

Table 4.1: Properties of boron and LAS

Property	Boron	LAS	Unit
Consumption*	0.8 (0.219)	1.2 (3.29)	$kg/cap/a$ ($g/cap/d$)
Removal in sewers**	0%	0%	% of total mass
Removal in primary settler**	0%	0%	% of total mass
Removal in trickling filter plants	0%	94-98***%	% of total mass
Removal in activated sludge plants	0%	98-99.5***%	% of total mass
In-stream removal	0.0	0.06	$1/h$

* This corresponds to the Yorkshire area
** Removal aggregated in plant type-specific (therefore overall) removal
*** Uniform distribution

Four Yorkshire catchments are incorporated in GREAT-ER: the River Aire and its subcatchment River Calder, River Don and River Went (figure 4.1).

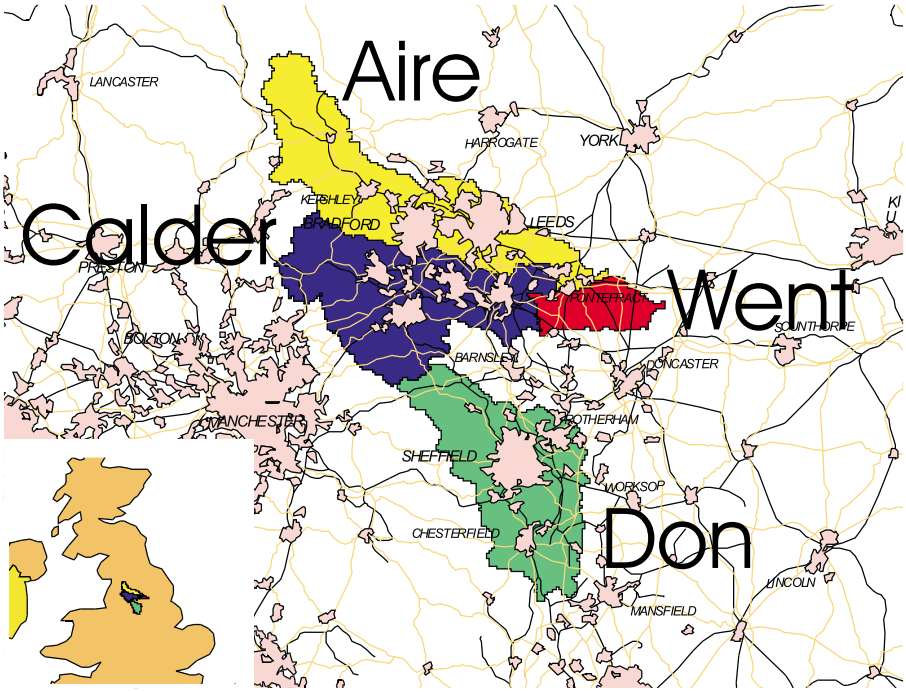


Figure 4.1: *UK catchments Aire, Calder, Don and Went*

The digital river network and associated flow statistics are taken from the database of MicroLowFlows (Gustard et al., 1992). The river networks are split

into stretches with coordinates at a resolution of 1:50,000. The lengths of the stretches which are important for travel times are provided on the basis of a 1:20,000 resolution and in the actual GREAT-ER data set range from about 10 m to 4 km with an average of about 780 m.

The flow statistics are provided as log-normal distributions with mean and 5th percentile (=95th percentile low flow) for all stretches (estimated for the ungauged sites) and also consider artificial influences (e.g. abstractions), which is a recent feature of MicroLowFlows to better fit the requirements of GREAT-ER.

The flow estimation in MicroLowFlows is based on dry weather conditions and for sewage treatment works considers the consented / design dry weather flow.

4.2.1 Calder

The River Calder rises in the Pennine Moors and is a predominantly urban catchment of 955 km^2. A population of 798,000 is served by 21 sewage treatment works. The main stream has a length of about 86 km and confluences with the River Aire. The mean flow directly above the confluence is 17.8 m^3/s. The river network is split into 1,545 stretches.

4.2.2 Aire

The headwaters of the Aire rise on carboniferous limestone moorlands. Up to its confluence with the Ouse the main stream has a length of about 148 km of which about 115 km are covered by the present Aire catchment with a final mean flow of 34.74 m^3/s. The catchment includes the River Calder watershed and drains a total area of 1940 km^2 (985 km^2 without Calder). 34 (13) treatment plants serve a population of about 1,927,000 (1,125,000). Above the Calder confluence almost all of the treatment plants are of the trickling filter type. The total digital river network consists of 3,022 stretches.

4.2.3 Don/Rother

The Don catchment includes the River Rother and contains a large city, Sheffield, in its center. The present catchment contains the upper part of River Don beginning upstream from Doncaster. The catchment has an area of 877 km^2. A population of 815,000 is served by 9 treatment plants, one of which serves Sheffield with 422,000 inhabitants. All other plants are connected to the River Rother which lies to the east of the catchment flowing north. The River Rother confluences with the River Don after about 53 km of the main stream. Above the confluence with the River Don, the Rother carries a mean flow of 4.44 m^3/s. The catchment ends with a mean flow of 11.12 m^3/s. The total digital river network consists of 1,011 stretches. Although River Don is the major stream in this catchment, the focus in this thesis will always be on the River Rother.

4.2.4 Went

The Went is a small, predominantly rural catchment draining about 195 km^2. Parts of the catchment consist of steep valley woodland and limestone grassland. 7 treatment plants serve 27,000 people. The main stream has a length of about 28 km and then confluences with the River Don. The mean flow above the confluence is 1.24 m^3/s. The digital river network consists of 127 stretches.

4.3 Discharge sites

Generally all discharges in the UK are treated. In Yorkshire about 2/3 of the treatment plants are trickling filter (TF) and 1/3 are activated sludge (AS) plants.

For the release computation the per capita consumption and the population numbers connected to the treatment works are of big importance. The given population equivalents are based on planning/construction purposes for treatment works and reflect water consumption and/or water quality. They do not necessarily present equivalents in correlation to the consumption of any substances. Population equivalents can indicate the degree of industrial influence but the type of industry can not be identified. In order to serve as an industry indicator, they are included in tables 4.2 to 4.5, although real population

figures are more suitable and are used for studies within this thesis.

Some plants are described as combined AS/TF. In these cases, parallel treatment is assumed for the simulations and they are regarded as trickling filters. This choice is based on the fact that the less efficient discharge unit dominates the more efficient ones concerning the final effluent concentration. Nonetheless, the knowledge about variations of the technical concepts of treatment plants is taken into account for the comparison of simulation and monitoring. The risk assessment point of view remains on the conservative side, and hence must also assume TF plants.

All data (including coordinates) were obtained from the Yorkshire Environment Agency and other local authorities.

4.3.1 Calder

Figure 4.2 and table 4.2 give an overview on the discharge sites of the Calder catchment.

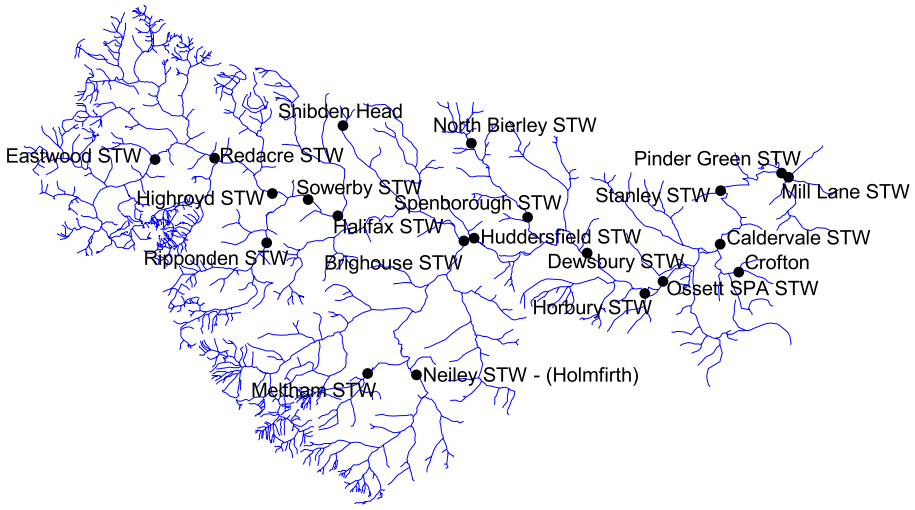


Figure 4.2: *STPs in the Calder Catchment*

A plausibility problem is identified with the STP *Osset SPA*. The receiving tributary has a mean flow which is almost the same as the treatment plants' final effluent. Consequently, this will result in the unlikely situation that in-stream

concentrations are as high as in the final effluent. Two explanations are conceivable: Either the tributary is actually a carrier canal to the main stream or the STP effluent flow was not considered as an artificial influence for the flow estimation routines. Regardless of this, simulation results for this tributary (consisting of stretches with the IDs 25711, 126815 and 25712) should not be considered for standard analysis procedures, because the unusually high values might suggest a situation that is not comparable with the rest of the data.

Table 4.2: Sewage treatment works in the Calder catchment

	Type	Population	Population Equivalents	ADF m^3/d
Eastwood STW	TF	14244	21367	14919
Redacre STW	TF	7953	10396	4759
Highroyd STW	TF	10346	11000	5825
Ripponden STW	TF	4700	4847	1699
Sowerby Bridge STW	AS	12419	16148	4775
Halifax STW	AS	99213	226221	52367
Shibden Head STW	TF	9408	9408	3478
Brighouse STW	TF	49892	71157	24424
Meltham STW	TF	7902	7967	4274
Neiley STW	TF	18043	36924	8560
Huddersfield STW	AS/TF	169946	671304	101810
North Bierley STW	TF	39440	86698	18160
Spenborough STW	TF	37357	64045	14055
Dewsbury STW	AS/TF	127050	309012	59600
Horbury STW	TF	14837	15000	4605
Ossett SPA STW	TF	17153	33752	6052
Crofton	AS	8926	9276	1834
Caldervale STW	AS	93726	132208	44095
Stanley STW	AS	17636	17729	4762
Mill Lane STW	TF	37617	41231	12818
Pinder Green STW	TF	650	650	385

TF = Trickling Filter, AS = Activated Sludge, AS/TF = Combined plant of both types
ADF = Actual Daily Flow
Gray rows: STP final effluent is covered by the monitoring campaign

4.3.2 Aire

Figure 4.3 and table 4.3 give an overview on the discharge sites of the Aire catchment.

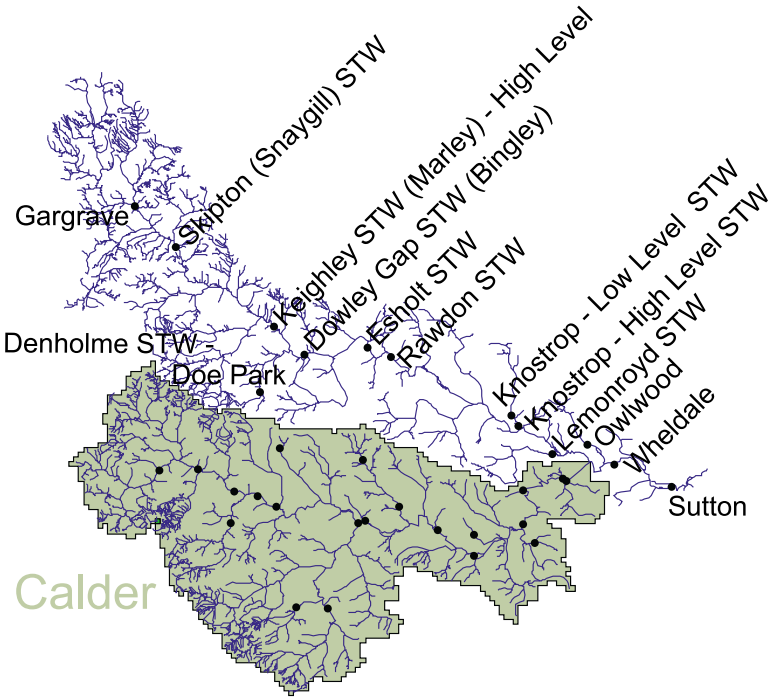


Figure 4.3: *STPs in the Aire Catchment*

Table 4.3: Sewage treatment works in the Aire catchment

	Type	Population	Population Equivalents	ADF m^3/d
Rawdon STW	TF	6642	6681	1979
Denholme STW - Doe Park	TF	2885	2885	716
Knostrop - High Level STW	TF	323519	410443	129219
Knostrop - Low Level STW	AS/TF	218714	439712	143520
Lemonroyd STW	TF	27260	27260	8840
Wheldale	TF	23539	61810	9074
Owlwood	TF	36572	40810	7805
Sutton	TF	45116	56377	12372
Gargrave	TF	1588	2297	523
Skipton (Snaygill) STW	TF	17278	19414	6829
Marley STW - High Level	TF	79077	120351	35161
Dowley Gap STW (Bingley)	TF	33717	50460	13621
Esholt STW	TF	312826	736884	133444

TF = Trickling Filter, AS = Activated Sludge, AS/TF = Combined plant of both types
ADF = Actual Daily Flow
Gray rows: STP final effluent is covered by the monitoring campaign

4.3.3 Don

Figure 4.4 and table 4.4 give an overview on the discharge sites of the Don catchment.

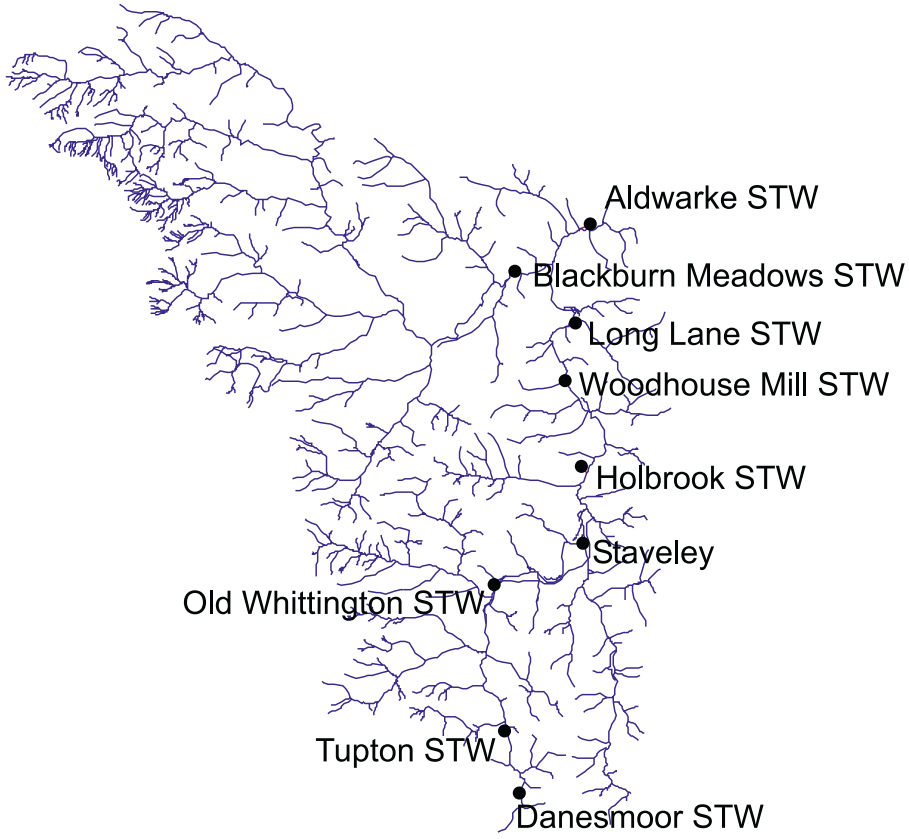


Figure 4.4: *STPs in the Don Catchment*

Table 4.4: Sewage treatment works in the Don catchment

	Type	Population	Population Equivalents	ADF m^3/d
Old Whittington STW	AS	96856	149627	34060
Holbrook STW	TF	26500	25668	6030
Woodhouse Mill STW	AS	96764	114469	26260
Long Lane STW	TF	22842	23130	5343
Staveley	AS	26378	22224	7046
Blackburn Meadows / Sheffield	AS	422677	497020	180000
Danesmoor STW	TF	6020	6300	1479
Aldwarke STW	AS	107203	139388	32410
Tupton STW	TF	10105	10105	2995

TF = Trickling Filter, AS = Activated Sludge

ADF = Actual Daily Flow

Gray rows: STP final effluent is covered by the monitoring campaign

4.3.4 Went

Figure 4.5 and table 4.5 give an overview on the discharge sites of the Went catchment.

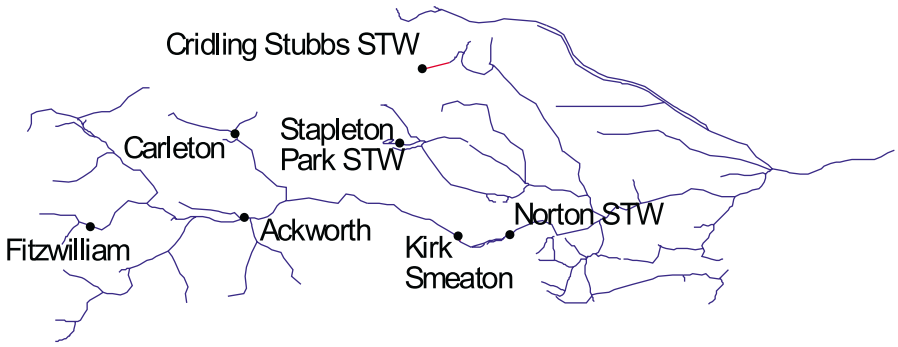


Figure 4.5: *STPs in the Went Catchment*

Table 4.5: Sewage treatment works in the Went catchment

	Type	Population	Population Equivalents	ADF m^3/d
Fitzwilliam	TF	845	845	364
Ackworth	AS	7737	7803	2591
Carleton	TF	6619	6619	1368
Norton STW	TF	10334	10334	3553
Stapleton Park STW	TF	1333	1788	498
Cridling Stubbs STW	TF	185	185	32
Kirk Smeaton	TF	664	664	90

TF = Trickling Filter, AS = Activated Sludge
ADF = Actual Daily Flow
Gray rows: STP final effluent is covered by the monitoring campaign

Chapter 5

Data: Composition and processing

Besides the environmental model itself, the availability and quality of environmental data mainly determine the validity of simulation results. This chapter introduces general aspects and problems of spatial data for environmental exposure assessment. A proposed procedure is described for how to deal with the collection, processing and quality assessment of data considering also the origin and purpose of readily available data. The technical aspects will be focused on but it must also be stressed that obtaining spatial data in the EU reveals serious coordination and authorization problems.

In this chapter an example of the proposed procedure is given in which the digital river network is the backbone of the spatial data set for ”down-the-drain” exposure assessment. A prototype realization proves that quite simple data formats are sufficient for automatically generating input data for simulation and visualization. The example is given for GREAT-ER but the methodology is equally applicable for similar systems.

All source codes are available from the author on request. These represent the most precise description of formats and algorithms.

5.1 Introduction

The data required for the regional geo-referenced exposure assessment is characterized by its complexity and inhomogeneity. This is partly due to the fact that there are large amounts of single data. But another fact plays an impor-

tant role: the composition of such data is not readily available. It is typically scattered among several competent authorities and other data owners. In the past those different parties had seldom experienced the necessity to harmonize their data and hence it is to be expected that all obtainable data are in several digital formats and / or geographical resolutions. Inconsistent ID reference systems are likely to be faced as well. The approach to regional geo-referenced exposure assessment has created a need for the coordination and composition of base data which are, though collected and applied for various topics, nothing more than *raw* data in respect to the new task.

The reason for this lies also in the way in which most geographic data are used: for mapping purposes where structural integrity is much less important than visual appearance.

The data composition for exposure assessment in river networks reveals many problem-specific aspects. When one would set up a very special recipe for certain data sources or data origins, this task would lack re-use capabilities and also would not be considered adequate for other comparable tasks. In the framework of a simulation system for a growing number of catchments, re-use capabilities require special attention to ensure the efficient creation of many digital river network data sets. The solution for data composition must be simple and open enough to offer quick understanding, general applicability and also to enable technical solutions to be adapted.

The usual steps of data composition and processing as a GIS task are typically single manual steps of applying geographic methods and routines. The required quality and structure of the final geographic data (i.e. maps) is approached in a straight-forward manner. The data sources and intermediate stages are often eliminated as soon as the latest version passes a quality control. It is unusual for GIS users to automate the processing steps because each new job reveals other aspects of processing which can not all be foreseen. To make the results transparent and to avoid later wrong use, all processing steps must be documented very carefully. This is especially important when results are not only used for reporting purposes but also for further processing and comparison with other data sets processed in a similiar way.

A well defined intermediate stage of the data is a general approach to separate the specific problems of raw data from the actual data preparation. Setting up this intermediate state on text-based formats increases transparency. A set of pre-defined files requests data at an adequate level of simplicity while at the same time being able to grant some aspects of quality and completeness. Furthermore, a high degree of automation can be created starting from pre-

defined text files.

As an alternative intermediate state a pre-defined database management system (DBMS) can also serve the task of separation. Such an approach creates further needs in software requirements and skills to handle them. Depending on the frequency of usage, the DBMS solution, once set up and tested, can be more efficient than the solution based on simple text files. The advantages arise from the DBMS environment which offers support for distributed systems (including Internet applications) and high performance in handling large data sets. With its higher level of sophistication such a solution loses self-explaining transparency which needs to be compensated by comprehensive usage documentation and / or by dedicating a human operator. Proving its principle operability with the text files as an intermediate stage is a recommended step before setting up a DBMS solution. The two-step approach is therefore choosen for the GREAT-ER data processing.

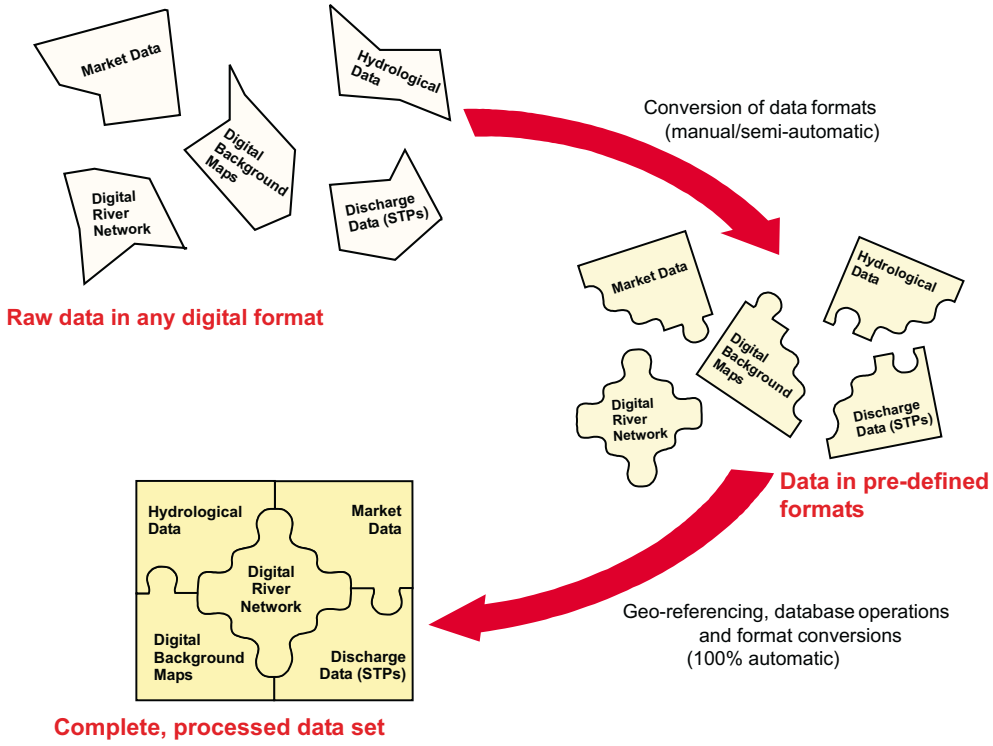


Figure 5.1: *Two steps of data processing*

Inventing an intermediate stage for data consequently results in two steps of data processing. In a first step, collected raw data are transformed into the

pre-defined intermediate form. A second and independent step then creates final ready-to-use data from the intermediate stage.

The intial data processing is very much determined by the gap between demanded data quality and formats and the present state of the raw data. The required files have to be produced manually from the original raw data or with special routines which probably have to be developed anew.

The final data processing is a well defined task, provided that all required data are collected and fit the demands of pre-defined formats. Of course, the final environment of the geo-referenced exposure assessment tool must also be defined precisely to make the data processing fully automatable. In this context, the simulation method determines the structural shape and order of the data while the user interface and data management modules determine the actual formats of files.

The two steps of data processing are illustrated in figure 5.1. An example is given for an exposure assessment tool for river networks. Three main states of the five data groups are shown. These data groups cover any information that is related to space in one way or another. For example, physico-chemical data of a substance are excluded but the substance's market data are included as regionally resolved information.

Another view of the finally produced data sets is shown in figure 5.2. The data groups form layers which build the visual interface available to the user.

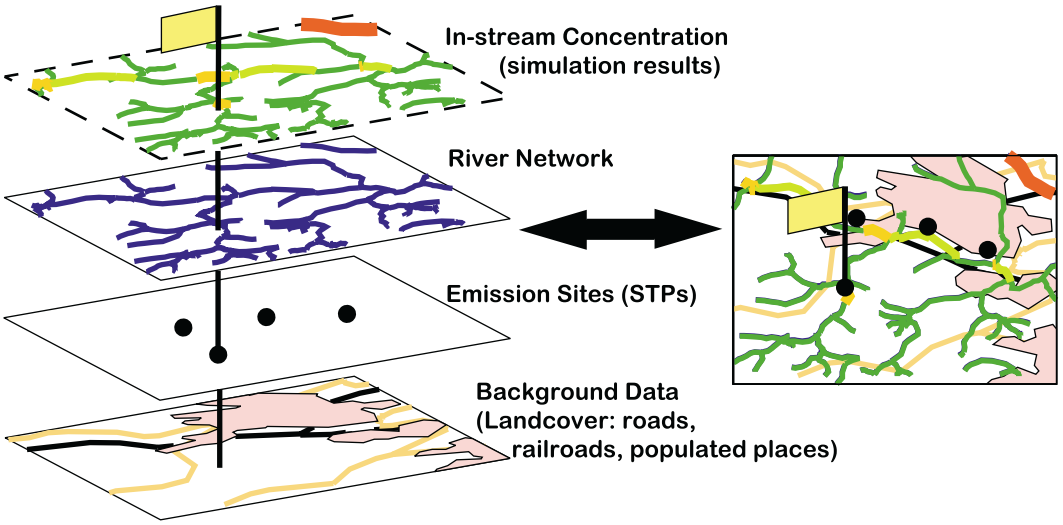


Figure 5.2: *Layered view of data groups*

5.2 Data requirements and composition

5.2.1 Digital river network

General overview

The digital river network describes the structure of the river network in question. At the very bottom it consists of lines (also referred to as 'polylines' or 'arcs'), each of which represents a river segment. Furthermore, it consists of nodes which represent the point of connection of two segments. Lines are expressed as a sequence of at least two coordinate pairs, nodes are expressed as exactly one coordinate pair (figure 5.3b).

Lines have to be considered as vectors; the first point of the sequence is the starting point and the last is the end point. The vector direction is used to represent the flow direction (figure 5.3c). The length of a segment can be computed from the coordinates.

Each transition from one segment to a subsequent one (sequence, confluence or bifurcation) is defined by a node which contains the information 'from' and 'to'. The ID's of the upstream ('from') and the downstream ('to') segments are stored. The geographic position of a node is uniquely defined with the last point of the upstream and the first point of the downstream segment. If these two points are not identical, the two stretches are not connected and no corresponding node exists.

Besides the persistent Arc-Node geometry model, a second, flexible segmentation can be set on top of it. The two-dimensional geometry model lacks the ability to precisely locate sites along the lines between two given coordinates. This can be solved by introducing a one-dimensional measuring system with only starting and ending information.

These one-dimensional geometric objects are called 'routes' . For digital river networks, each river is a 'route'; the whole network constitutes a 'route system'. Any site along the rivers in a route system can be referenced by the route name and the distance from the route start (e.g. River Rhine at kilometer 744 which is at Düsseldorf, Germany). The start and end of routes are a matter of definition. In Germany, the distance reference of rivers that lead into the ocean begin at the source. The distance reference of tributaries begins at the point of confluence in reverse flow direction. Standard GI Systems offer methods to

create route systems from an arc-node model. This must be a two-step routine where a raw route system is first created based only on the information contained in the arc-node representation (figure 5.3d). The raw route system requires two further adjustments. First, the digitized rivers do not necessarily start at the actual source and hence the starting points of the routes must be set to the correct value for later consistent referencing. Second, the length computed from the digitized lines depends on the resolution and hence is always lower than the actual river length. A remeasurement routine that is provided with the actual length of a route, is to be performed to gain revised coordinates for any event along the stream. The route system approach for exposure assessment was applied for intermediates in the River Rhine catchment (Koormann, 1997 and Koorman et al. 1998) and for a small tributary of the Rhine, the Itter catchment (Schulze, 1998 and Schulze et al., 1999).

Fundamental to the creation of a digital river network are maps of the corresponding area. Traditional printed maps can be used as a basis but usually such maps now exist in digital format.

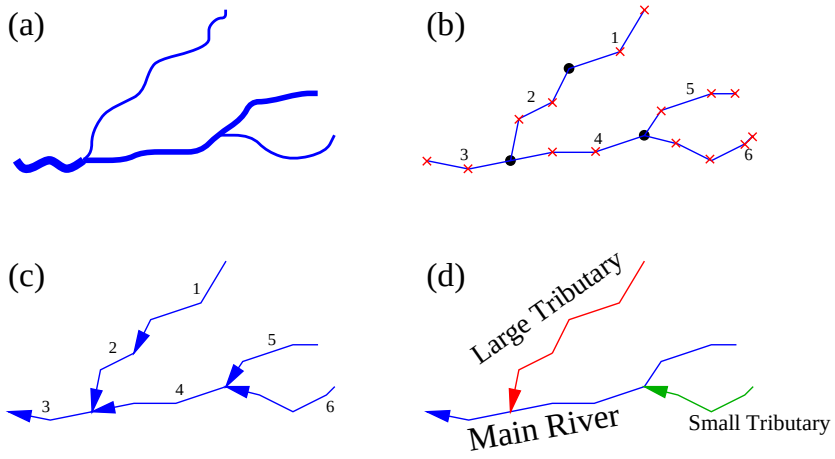


Figure 5.3: *Steps of creating a digital river network: (a) Reality (i.e. aerial photo or comparable basis). (b) Digitized network (crosses and dots = coordinates, dots = nodes, numbers = IDs of lines). (c) Direction of lines (vectors). (d) Representation as route system.*

Digital maps can be created in two ways, either by collecting coordinates directly with a Global Positioning System (GPS) or from aerial photographs. While GPS data are naturally already geographic objects in vector format, pictures taken from planes or satellites offer information in a raster format. I.e.,

surface waters consist of single points (squares) and hence do not contain information about e.g. flow direction, but on the other hand may offer information about areal aspects (e.g. width of rivers, shape and size of lakes) and possibly further properties (temperature, depth). The raster information must be converted to vector data to gain a digital river network as described above. This conversion can partially be automated but manual reworking by human experts is unavoidable for quality assurance.

Solution for GREAT-ER

As a requirement for data, the arc-node model is preferred rather than the route model. This is because, especially for more detailed river networks, an assignement of river names and remeasurement become problematic and time-consuming.

In summary, the following demands for a digital river network have been defined:

- Only the coordinate sequences for the river segments are to be given. No extra node specification is required.
- Should two segments be connected, it is necessary that the end point of the upstream and the starting point of the downstream segment are identical. This also means that for confluences or bifurcations the upper segment(s) always end at that point and one (or two) new segments start from the same point.
- All river segments must be directed downstream.
- All segments must have a unique ID.

Only confluences and bifurcations enforce the beginning of new stretches. Sequences like segments 1 and 2 in figure 5.3 have no topological meaning for the network. Splitting up segments into a sequence of two or more stretches becomes important for the semantic purposes the digital river network has to serve. Simulation models using digital river networks may need to simplify river segments by assuming stretches to be homogeneous in a hydrological sense. In this case flow data would be attributed to river stretches and hence long segments need to be split up according to the required resolution or available knowledge about hydrological changes. Additionally, for visualization

purposes the level of segmentation also determines the capabilities to graphically display stretch-specific information. Hence, when setting up the digital river network the later purpose of these data should already be considered at this point and segmentation should be performed. This task can also be undertaken automatically provided that the single coordinates of the lines are more or less equidistant.

5.2.2 River network attributes

Depending on the needs of an applied exposure simulation model, information about certain properties of the river network are required. Usually these are floating point numbers (e.g. flow in $[m^3/s]$ or flow velocity in $[m/s]$). Further attributes such as boolean (shipable [yes|no]) and others are conceivable. Most of the traditionally measured data are expressed as statistics due to natural fluctuations and seasonal variation. For example, flow data are usually expressed as log-normal distributions characterised by the mean and standard deviation (or mean and low, e.g. 5th, percentile).

However, all discrete attributes (in space) have in common that they correspond with a certain river segment. A set of discrete parameters for a river network can hence be given in tabular style. Continuous parameters (in space) are more difficult to attach to digital river networks, and applications using such data also reach a level of complexity which is seldom acceptable for actual environmental simulations. Usually distribution curves are derived from temporal data series which are then applied for a specific task.

Depending on the purpose for which the river attributes are needed, in some cases the variation of each river segment is of importance and for other parameters variations are only of interest for entire rivers or river networks. Furthermore, decisions regarding the stretch-wise resolution of parameters are influenced by data availability. Consequently, a task-specific separation of parameters is helpful and can be solved by defining river classes and associating each river segment to a river class.

Parameters that are semantically important and presumably different for each segment are directly associated to river segments. For many parameters the reasons for actual availability at high resolution represents the opportunity not only to measure but also to interpolate between measurements or to estimate them from other attributes. Hence, again depending on the semantic context, not all parameters are obligatory. A flexible (interpolating, estimating) system

of optional parameter specification is possible.

The methods of chemical exposure assessment are organized in a tiered approach in terms of physico-chemical processes. Lumped transport and elimination processes will be applied first and if it is not capable of explaining monitored behavior, refined substance-specific processes will be added (e.g. photodegradation). Any refinement increases data demands, uncertainties and, on a very practical side, expense and time.

The counterpart of the lumped parameters for chemicals (transport and elimination) are hydraulics and geometry of rivers. These basically determine the dilution and dispersion of chemical loads. Accompanied by length and name, these were chosen as direct attributes. All other parameters are assembled in river classes. This choice reflects the data requirement of the simulation software incorporated in GREAT-ER (Boeije, 1999). The concept of the intermediate data stage still allows us to consider data requirements of any comparable simulation model by providing corresponding automatic data processing routines for the second step (figure 5.1).

Flow, Velocity, Depth and Width

Overview of interrelation For a sufficient period of time, the flow passing the full extent of a river segment can be assumed as constant in space and time. In this case the continuity equation 5.1 is true.

$$Q = vA \tag{5.1}$$

where Q = flow, v = velocity and A = cross-sectional area.

Rivers have no rectangular cross-section and hence A can not simply be substituted by the product of depth and width. Depth and width have a correlation with the flow within certain ranges. A set of power equations 5.2 can express the interrelation of velocity (v), width (w), depth (d) and flow (Leopold & Maddock, 1953).

$$v = aQ^{\alpha} \quad ; \quad d = bQ^{\beta} \quad ; \quad w = cQ^{\gamma} \tag{5.2}$$

where both $\alpha + \beta + \gamma$ and $a \cdot b \cdot c$ must be equal to 1.

Compared to velocity and depth, width is generally less variable (in space) for most rivers (Chapra, 1997). Especially when dealing with one-dimensional models, width is of less semantic importance. In contrast, depth is important, i.e. for chemical volatilization processes. In this approach width is hence dropped from the list of required river properties. It should of course be taken note of when considering transversal dispersion (two-dimensional river models).

Site-specific measurements and site-specific regression functions are optimal to gain most adequate information on river properties. In the absense of such data, and this is true for most sites, literature or new field studies need to be undertaken.

Numerous studies were conducted to obtain values for parameters a, b, c and α, β, γ which vary with different climatic location and degree of artificial influences on the river networks.

It has also been shown that the hydraulic geometry can alternatively be obtained from the corresponding watershed size: Based on a strong correlation of bankfull flow and watershed area, a geometric relationship was shown and extrapolated for all river segments (Miller et al., 1997). Another method is based on the hortonian stream order for river networks. This ordering method attaches the order of 1 to all source stretches. The order of any other stretch is the maximum order of the directly connected upstream stretches plus 1. A good correlation of the hortonian order and the river geometry has been proved for a watershed in Arizona, U.S.A. (Miller et al., 1996).

Besides the statisticial approaches described above, mechanistic techniques can be applied. The best known method is the Manning equation (5.3) (Chapra, 1997).

$$v = \frac{R^{2/3} S^{1/2}}{n} \tag{5.3}$$

where n = Manning's roughness coefficient [-]
R = hydraulic radius (equal to A/P) [m]
P = wetted perimeter [m]
S = energy slope (here equal to geological slope) [-]

Substitution into the continuity equation results in the approach of Manning-Strickler (equation 5.4) (Dyck & Peschke, 1989).

$$Q = \frac{AR^{2/3}S^{1/2}}{n} \tag{5.4}$$

Besides flow and velocity, the depth can also be obtained based on the Manning equation. The depth is the root of equation 5.5 (Chapra, 1997).

$$f(y) = \frac{1}{n}\frac{[(w_0 + sy)y]^{5/3}}{(w_0 + 2y\sqrt{s^2+1})^{2/3}}S^{1/2} - Q \tag{5.5}$$

where s = side slope and w_0 = bottom width.

For the GREAT-ER project a new study has been undertaken to estimate the velocity of any river reach in the UK (Round et al., 1998). A statistical approach was chosen based on hydrological databases for the UK. The quality of the velocity estimation model (equation 5.6) was not significantly reduced in terms of performance by excluding the hydraulic radius. The error analysis showed that with 68% confidence the true velocity lies within a factor of 1.59.

$$v = 10^{-0.583}Q_{mean}^{0.283}\left(\frac{Q}{Q_{mean}}\right)^{0.495} \tag{5.6}$$

To obtain average depth values, variations in three dimensions have to be made. First, depth can vary in transversal and longitudinal extent of the river bed and next, it changes in time due to flow variation. For unknown geometry and known flow an application of the Regime Theory provides equations 5.7 and 5.8 (Simons & Albertson, 1960).

$$w = 6.175\sqrt{Q} + 0.305 \tag{5.7}$$

$$d = \begin{cases} 0.610 + 0.93r & : \quad r > 2.13 \\ 1.21r & : \quad r <= 2.13 \end{cases} \tag{5.8}$$

where $r = 0.57Q^{0.36}$.

Selection of values and methods Literature review has shown examples of the ability to estimate the interrelated river properties flow, velocity, depth and width from each other in several ways. Also area and mathematical topology can be included for estimations. The latter can even be applied when no stretch specific properties are known.

In general, higher efforts are required to measure the geometry of river segments while flows are easier to obtain: In most European countries a system of gauging stations is in operation, delivering flow statistics for some sites. Usually interpolation equations can easily be set up or are even readily available for ungauged sites between two gauging stations. Ungauged headwaters are more complicated to establish flow estimation. The flows can be computed from geo-morphology and other information, i.e. MicroLowFlows additionally applies soil type and precipitation information for this task (Gustard et al., 1992).

For the methodology introduced in this chapter, flow data are chosen as obligatory, and velocity and depth are optional parameters. Optional parameters will be estimated during the data processing. Consistency checks based on the literature can be integrated in the final processing.

Most estimation equations derived from statistical studies are associated with several limitations concerning the environmental framework of the investigated river network (natural / regulated, arid / humid, lowland / highland, etc.). If estimation equations are available which were set up for the same or a similar type of river these should be preferred and the resulting values should be used. Automatic selection of estimation routines can not decide adequately compared to human judgement. This is due to the fact that much general information on a river network is required and this information can not easily be formalized.

One important property of the discussed parameters is variation in time, which means that these parameters usually consist of three subparameters describing their statistical distribution (i.e. distribution type, mean, standard deviation). This must be kept in mind when dealing with corresponding data collection and estimation.

Length

In conjunction with velocity, the length basically determines the transport of a (chemical) load. The process of transport in terms of chemical exposure mod-

eling is a relatively fast process and hence the travel distance should be given as accurately (feasibly) as possible.

Obtaining the actual length of rivers should in general not pose a serious problem because it is relatively easy to measure. For some countries, canoeing guides offer this information in great detail, i.e. for Germany (Deutscher Kanu-Verband, 1985).

In theory, there are of course no problems in calculating the length of each segment based on coordinates. These values are naturally (sometimes much) lower than the real length of the river segments. These differences become obvious when taking into account that even detailed digitizing (e.g. one point every 300m) can not reflect the actual shape of a very curvy river. Additionally, the digitizing itself reveals uncertainties.

Name

Names for each river segment are not meant to identify stretches in terms of database operations. Obviously one river can be split up in very high numbers of stretches and therefore unique name specification does not make sense. The names are required rather for identification and orientation purposes in the final simulation tool. Names of rivers can hence be associated with other site names or any information that is felt to be helpful to users.

River Classes

The composition of the river class parameters is to be selected dependent on the semantic purpose of the simulation models. This way, the number of properties can be reduced to a minimum to retain low complexity and an easy overview.

The methodology presented in this chapter takes the river class composition as used for the simulation software in GREAT-ER (Boeije, 1999). Included are typical water quality parameters, i.e. suspended solids, dissolved oxygen, biological oxygen demand and pH. All these parameters are only used in higher model complexity modes, i.e. not applying lumped degradation rates. Among the biochemical properties, the type of segment is also described (river or lake).

5.2.3 Catchment boundary

The catchment of a river network is defined by the most downstream point, i.e. the most downstream segment. The catchment boundary represents the natural drainage area. Any point in that area will drain to that most downstream segment. The catchment boundary must always be a closed polygon.

The introduced data composition additionally enables several separated river networks to be joined. This helps users to deal with e.g. flat coastal regions, entire islands or even whole countries. In such cases the catchment boundary should be the union of all single subcatchments.

Information about the actual border of a catchment serves two purposes. First, it supports geographic methods, e.g. spatial selection routines. Second, within the final tool the user is given further support for orientation and interpretation.

In regions where water management has a long tradition, watershed boundaries should be readily available. Any new or re-calculation applying a GIS will require digital elevation data. This computation is often performed and several technologies are available, partly already built into a GIS.

5.2.4 Discharge data

The discharge data, in this context, are a set of points through which substances are released into the river networks, e.g. sewage treatment plants. The discharge information will not be able to express diffuse input (unless it is made discrete as a series of point inputs). Diffuse input is not focused in the presented methods because a different class of simulation models would be required.

Discharge points with geographical reference define an area where any release will pass this point travelling ”down-the-drain”. Such areas are scale-independent geographical units interconnected by a network topology (figure 5.3). This scale independency allows us to flexibly *refine* or *aggregate* (parts of) the geographic data where required. Refinement on the one hand offers dynamic, problem-oriented geographical resolution to avoid unnecessary data collection and helps to deal adequately with data gaps and to zoom into areas of special interest. Aggregation on the other hand enables us to reduce complexity, to compare regions by abstracting data and to calibrate actually

needed scales of data.

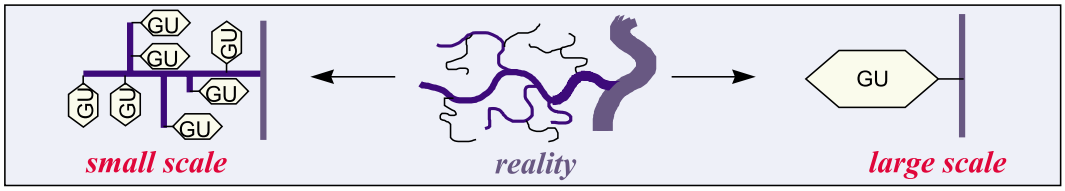


Figure 5.4: *Geographic units (GU): scale independent*

Coordinates The coordinates of a discharge site are obligatory. In the case of treatment works they describe the position of the actual plant and not the point at which the emission enters the river. The connection to the river will be established by explicit specification of the corresponding river segment. This enables provision of location information to the user who is supported in interpretation by offering the distance of the plant to the river and whether the emission takes place on the left or right bank.

Population For simulations of chemical ingredients in consumer products, the per capita consumption is usually available in mass per time. To properly estimate the release, the actual real population connected to a discharge is required. These numbers should be readily available from responsible environmental protection authorities, water authorities or water supply companies. For purely industrial discharges the population is zero.

It is not feasible to use population equivalents, because these numbers include industrial emissions based on a special selection and weighting of water quality parameters. For example, in Germany a per capita emission of 200 liters of waste water with an equivalent load of 60 g of biological oxygen demand (BOD) is assumed. Based on this, several population equivalent types are defined (DIN 4045 (12 / 1985); Römpp, 1997): Equivalents based on BOD (EGWB60), based on waster water emission (EGWW200), resulting sludge volume (EGWS2,0) and resulting dry sludge (EGWTS80). Across Europe several national standards are to be expected; for some countries variations between water authorities are even possible (e.g. in the U.K.). For most consumer substances no correlation with the standard water quality properties exists and hence population equivalents may reveal helpful background information, but are not required for chemical exposure assessment.

Flows Flows within the waste water and associated paths are measured and defined in several ways across Europe. For the data requirement definition a most common and practical selection is presented here. Above all the most important value is the flow leaving the plant and entering the river network. The actual daily flow (ADF) will mean the average (measured) emission of a plant in m^3/d.

In many cases consented maximum emissions are available. Such values are only variable with the consenting interval and do not necessarily represent the real emissions. For those sites for which no ADF is available the consented daily flow (CDF) is the best approximation, since the operators of waste water treatment facilities usually try to fill the consented flow.

The inflow of a treatment plant consists of three components: Runoff, non-domestic and domestic flow.

Discharge Type The type of a discharge is considered in the simulation with specific models to estimate removal efficiency. Usually the technological type of the corresponding waste water treatment plant is required (i.e. trickling filter plant, activated sludge plant), but the discharge type can be "direct discharge" (no chemical removal) as well.

Some sewage treatment plants are actually a combination of different technologies. A specific modeling of such situations is not feasible. A conservative simplification choosing the less efficient type is recommended, but for transparency the situation should be noted (see the *Name* below).

River connection Discharges are connected to the first point of the given segment which is to be specified with its ID. For an optimal representation of the actual river entry point, the river network digitizing or segementation should already consider these points.

In principle, GIS routines (distance measures) can be applied to find out the most reasonable stretches. These methods are automatic but always require human plausibility control.

Name The only purpose of the names is to support user orientation and interpretation as well as reporting. The names need not to be unique and should contain further information or hints when necessary.

5.2.5 Background maps

Though not used in the simulation models, background information is important for the final end user system. They allow visual orientation and plausibility control as well as aiding interpretation of simulation results, which is carried out by the user.

Different types of background information can be provided: raster landcover maps, vectorized information (geographic objects) on e.g. soil types and land use. Additionally, semantically related (temporal) information can be prepared such as monitoring sites with attached measured time series.

5.2.6 Pre-defined file formats

In this section the file formats for the data of a digital river network as used for GREAT-ER are described. This is done at a technical level to provide precise transparency.

All files are text based and use the character set ISO 8859-1 (ISO Latin 1). The format is line-oriented; each line starts with a '#' (comment line) and all empty lines have no meaning for the processing and can be inserted anywhere.

All coordinates are given in the same projection: "geographic projection" with unit "dd" (decimal degrees). A point, rather than a comma, is used as decimal point. Numbers are not allowed to contain any other delimiters. Exponential representation uses the letter 'E' or 'e'.

Digital River Network (.drn)

Each river segment is introduced with its ID `StretchID`. All following linewise comma-separated coordinates define this segment until a new ID appears. The first coordinate pair is the start, the last coordinate pair is the end of the segment. While the order of the segments does not matter, the order of the coordinates is, of course, of great importance.

River Network Attributes (.rna)

The river network attributes are stored in a simple tabular structure. Columns are separated by commas. The following example illustrates the order of the columns:

```
# Aire/Calder Catchment, Yorkshire
# StretchID,Qmean(mˆ3/s),Q5(mˆ3/s),vmean(m/s),v5(m/s),
# ... RealLength(m),depthmean(m),depth5(m),Name
26954,34.751000,12.280656,,,2491.655000,,,Unnamed
22954,0.009000,0.002359,,,1648.222000,,,Unnamed
7333,34.739000,12.276415,,,1629.885000,,,Unnamed
22932,0.005000,0.001088,,,540.190000,,,Unnamed
26950,34.731000,12.274977,,,2336.740000,,,Unnamed
22923,0.017000,0.002167,,,2447.520000,,,Unnamed
...
```

In this example no information is available for flow velocity and depth and hence the corresponding columns are left empty.

Discharge Sites Data (.dsd)

Each line defines one discharge. The first column is its unique ID, followed by the coordinate pair and the attributes. The order of the columns is shown in this example:

```
# ID,X,Y,Pop,DWF(mˆ3/d),Flow(mˆ3/d),Type,StretchID,Name
41,-1.390494,53.422081,422677,130000,180000,AS,26842,Blackburn STW
42,-1.349155,53.369724,95750,21800,24506,AS,26843,Woodhouse Mill STW
43,-1.326922,53.445114,107203,31250,32410,AS,26841,Aldwarke STW
```

Catchment Boundary Polygon (.cbp)

The polygon is described linewise by single coordinate pairs (x,y). The following example illustrates the format:

```
-2.052912,53.683811
-2.060483,53.683807
-2.060477,53.679314
-2.068047,53.679310
...
-2.045346,53.688309
-2.045342,53.683804
-2.052912,53.683811
```

This example also shows that the polygon must be closed, cf. the last and the first point must be identical.

Background data (.bgd)

To support the user with a quick and easy access to geographic background information, a list of background data sets can be defined for each catchment.

In the ".bgd" file names of additional background data that are to be loaded by the menu entry "Display / Add Background Data" are specified. The data can be given in two ways. In a simple case, a background data set is only a geographic data set that ArcView can load as a theme, i.e. an ArcInfo Coverage, an image, a grid or a shapefile. This case is indicated by the value "yes" in the field "ctch_flg" (see below). In the more complex case a background data set consists of a set of geographic data, indicated by the value 'no' in the field 'ctch_flg'. In this case, instead of the name of the data set an ID for a script that knows how to load the geo-data sets is specified.

Name This is the name for the corresponding background data. It is used for user selections in the GREAT-ER system.

Mapname In the case of a simple background data set, this is the file name of the coverage, image or shapefile that contains the data. Otherwise this field contains the script which can load the background data sets.

Type In the case of a simple background data set, this is its feature type. Valid entries are "point", "line", "polygon", "shape", "image" and "grid". Otherwise the field is left empty.

Legendfile In the case of a simple background data set, this is the name of an optional ArcView legend file. Otherwise this field can contain (optional) parameters for the script specified in the field ”Mapname”.

ctch_flg This is a boolean field that contains either ”yes” or ”no”. ”yes” indicates that this background data set consists of one theme that does not need a special script to load into the active view of GREAT-ER. ”no” means that a special Avenue script has been developed by the user which runs within GREAT-ER and loads the backgroud data set into the active view.

There is one special case: The DCW data (Digital Chart of the World) are incorporated in the GREAT-ER project for the whole of Europe. For this incorporation specific scripts loading the DCW data in dependence of the viewed country are developed. If the DCW data are to be a background data set for a catchment, the following line needs to be inserted into the file:

```
DCW data,,,,
```

If a user installs the DCW data for countries that have not yet been included, e.g. France, Spain or Denmark, these data need to be installed analogous to the included countries.

Example ”.bgd” file This example file describes background information for the River Itter (Germany):

```
#River Itter background data
#Name,Mapname,Type,Legendfile,ctch_fl
Cities,itter\cities,polygon,itter\cities.avl,yes
Tk25,itter\tk25tiffz.tif,image,,yes
Rhein,itter\rhein,polygon,itter\rhein.avl,yes
DCW data,,,,no
```

5.3 Final automatic data processing

The final data processing requires syntactically and semantically correct source files of the files ”.drn”, ”.rna”, ”.dsd”, ”.cbp” and ”.bgd” as described above.

The whole processing with its subsequent steps is managed by a *makefile*.

This part of the data processing is, in principle, not dependent on the required simulation system, though it does simplify the simulation system and user environment developed within the framework of GREAT-ER. The simple tabular structures allow the easy extension of the data files for adaptation to any other comparable simulation system.

The data to be generated can be divided into two groups: geographic data and the corresponding attribute data. The geographic data form the basis for visualization and geo-referenced data processing. These elements (cf. points, lines and polygons) are associated with attribute data via unique ID's. Their interconnection is based on the same ID reference.

Processing of the background data (".bgd" file) is not related with the main processing due to its special task to support the user. It is not discussed further in this document.

5.3.1 Geographic data

The establishment of geographic data consists of two processing steps (figure 5.5). First, the geographic elements (their ID's and coordinates) are extracted from the files in the pre-defined formats and converted into another format that can be handled directly by a GIS, in this case aimed at ArcView which comfortably handles the Shapefile format.

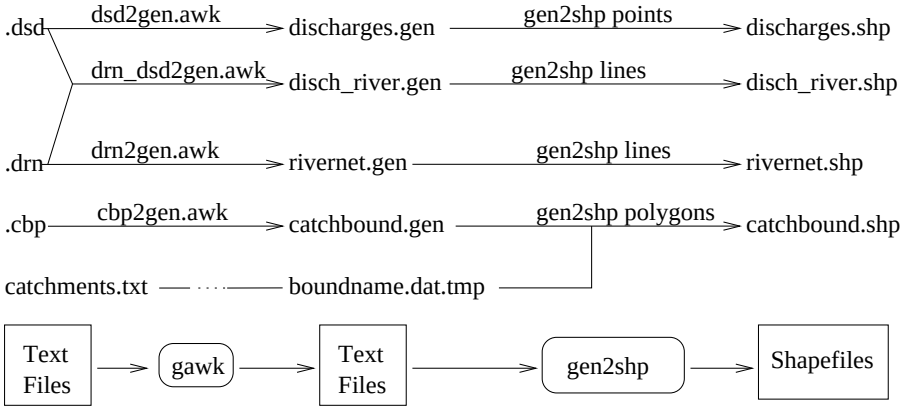


Figure 5.5: *Command sequence for automatic processing of geographic data*

A straightforward conversion is executed for the two digital maps: discharge sites (”discharges.shp”) and river network (”rivernet.shp”). For the boundary of the catchment (”catchbound.shp”) a further attribute is appended from the datafile ”catchments.txt”: the full name of the catchment.

The connection lines between discharges and corresponding river segments are computed using information from the ”.dsd” and ”.drn” files. A special shapefile is created (”disch_river.shp”), which provides visual information on where the discharge is connected to the river network.

5.3.2 Attribute data

While the attribute data are being processed (figure 5.6), the three main data files (”.dsd”, ”.rna” and ”.drn”) are first joined and then create the primary attribute files for river segments (”river.att”) and discharges (”disch.att”) as well as the secondary attribute files for treatment plant types (”wwtp.att”) and river classes (”rivclass.att”). The two secondary files are (currently) static and basically describe generic treatment plants and rivers.

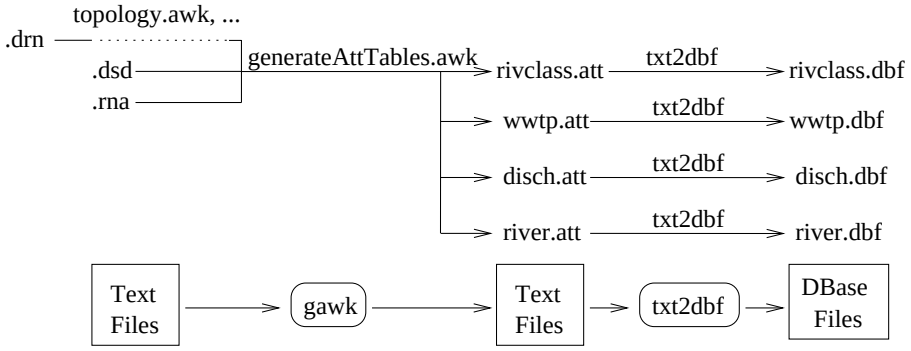


Figure 5.6: *Command sequence for automatic preparation of attribute data*

In the final processing step, text based tables are converted into DBase format. The purpose of this is to optimize data management within the visualization tool ArcView.

5.3.3 Core algorithms

Several data processing steps consist of simple routines (e.g. format conversions). They are not explicitly explained here. In this section the core problems of data processing are discussed.

Creation of River Stretch Sequence

The information about the logic sequence of single river segments plays an essential role for simulation models. A simulation model must be able to *walk* down from any source to the final mouth. Hence, each stretch must offer information (the ID) about its successor. This purpose is served by a simple two-column table `fromStretchID,toStretchID`.

The ".drn" files, if semantically and syntactically created as described above, contain the topology information, but coded in the given coordinates. The arc-node model is used to convert the topology to the coding based on StretchIDs. Building the arc-node model from a ".drn" file associates all segments with IDs of the corresponding "from-node" and "to-node" to describe their direction. Consequently, the two-column tables `fnode,StretchID` and `tnode, StretchID` can be extracted directly from the arc-node model ("topology.awk"). Joining these two tables by the node ID and then omitting the node ID results in the two-column table `fromStretchID,toStretchID`. In both columns IDs may now occur multiple times due to confluences and bifurcations. To create a table where the `fromStretchID` is a unique identifier, the `toStretchID`s are joined up into a (space character divided) list. The resulting two-column table `fromStretchID,toStretchIDlist` is a straightforward representation of the downstream topology. Analogously, an upstream topology table `toStretchID,fromStretchIDlist` can be created starting with the `fromStretchID,toStretchID` table and exchanging the position of the two columns.

Creation of Binary River Network Topology

Fast and straightforward access for any algorithm walking up or down the river network is realized using simple data tables. The minimum number of columns to describe any river network topology is four. Demanding one column being a unique segment ID and also allowing only exactly one ID in any

field, virtual segments need to be added to grant topologic consistency (e.g. confluence of three streams at one point). An additional column finally describes the segment type (listed in table 3.1). This is needed to correctly interpret the content of the other columns. Extending the network by virtual segments will not increase but decrease the total table size because a further column would rarely be filled for natural networks where a confluence of two streams is usually the maximum. For artificial networks (e.g. sewer systems) the situation can be different.

Table 5.1 *Segment Types*

	Number of segments		
Type	Upstream	Downstream	Description
0	0	0,1,2	Source
1	1	1,2	Normal segment
2	2	0,1	Confluence
-2	1*	0,1,2	Bifurcation

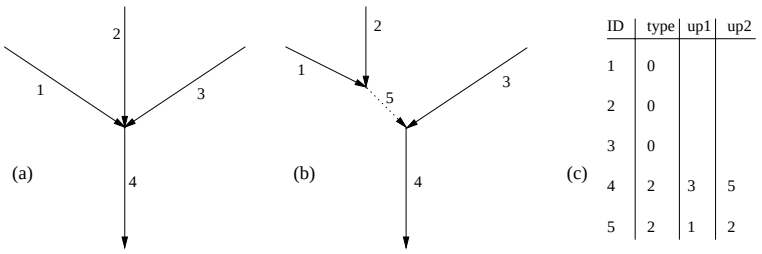
* which must be of type 1 with 2 downstream segments

Algorithms for walking on binary trees from the leaves to the root (cf. downstream) are typically implemented recursively starting with the root (cf. mouth). For exposure assessment in river networks the binary recursion has to be ”post-order” which means performing an action for a segment after all upstream segments have already undertaken this action. Hence, the upstream direction is required for the two-column topology table.

The above-described tables for the river stretch sequence are the basis for building the binary topology. During the data loading routine, the highest segment ID is looked up. This is important to invent further ID’s for virtual segments.

When all data are loaded, first bifurcation situations are searched for. The corresponding segments (type ”-2”) are associated with their neighboring segments. This is important in order to reduce the final topology processing to a one-pass computation (a single linear processing step). Next, the segment table is processed straightforwardly. Depending on the number of upstream and downstream segments, the type is then decided. Several special cases need further treatment i.e. insertion of virtual segments:

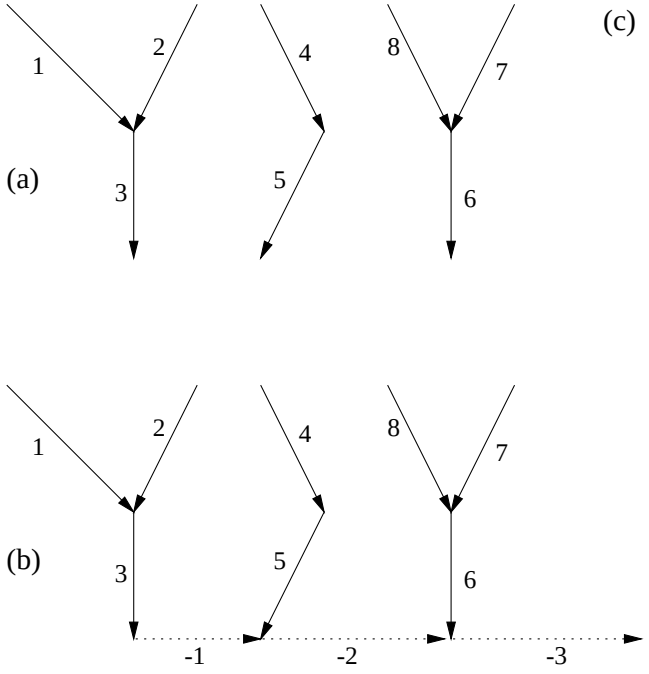
(I) Confluence of three segments at one point When exactly three segments are directed to the same point, a new (virtual) segment is inserted as shown in figure 5.7.



ID	type	up1	up2
1	0		
2	0		
3	0		
4	2	3	5
5	2	1	2

Figure 5.7: *Resolving confluence of three segments (a) into binary representation (b). Numbers are stretch IDs. The table of binary network topology of (b) is (c).*

The virtual segment obtains a length of 0. The ID does not appear in the geographic data of the river network and is neutral for the database operation ”join”. The simulation software must of course know how to treat segments with a length of 0: The results (e.g. chemical loads) must be transferred one-to-one to the next downstream segment.



ID	type	up1	up2
1	0		
2	0		
3	2	1	2
4	0		
5	1	4	
6	2	7	8
7	0		
8	0		
-1	1	3	
-2	2	-1	5
-3	2	-2	6

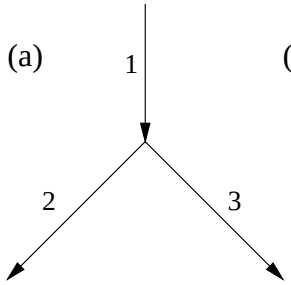
Figure 5.8: *Resolving multiple end segments (multiple catchments) (a) into one catchment (b). Numbers are stretch IDs. Table of binary network topology of (b) is (c).*

(II) Several end segments If several catchments are joined (e.g. to cover an entire island) the problem of multiple start segments (mouths) occurs and needs to be incorporated into the topology.

The solution is a linear interconnection with virtual segments. By definition all "ocean segments" are attached with negative ID's.

In the example shown in figure 5.8, the lowest ID is -3 and it is the new starting segment for the resulting catchment. The ocean segments influence neither visualization nor simulation.

(III) Bifurcation For bifurcation situations both neighbours are of the type "Bifurcation" (figure 5.9). Field `up1` will contain the ID of the upstream segment and field `up2` will contain the ID of the neighboring segment.



(b)

ID	type	up1	up2
1	0		
2	-2	1	3
3	-2	1	2

Figure 5.9: *Handling of bifurcations (a) in the binary presentation (b).*

(IV) Unconsidered situations Some situations may occur in the digitial river network, but are unlikely to happen in reality for natural catchments. In practice it is more likely that wrong digitizing or another error may cause the situation. Such situations are not considered in the presented methodology though it is possible to resolve them with further virtual segments or additional definitions.

(IV.a) Multiple confluence More than three segments are directed to the same point (figure 5.10a). This situation could be managed similar to the confluence of three segments using virtual segments, but from a hydrologic point of view this situation is most unlikely and the chance of using wrong base data is much more probable.

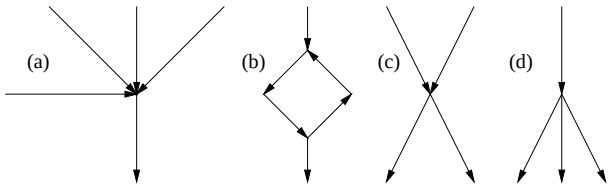


Figure 5.10: *Unconsidered situations: Multiple confluence (a), crossing (b), circle (c) and multiple bifurcation (d).*

(IV.b) Crossing streams Both, a confluence and a bifurcation take place at one point (figure 5.10c). Such a crossing is unlikely. Furthermore, a chemical fate model would need additional information to adequately deal with such a situation (what fraction of which upstream segment goes into which downstream segment?).

(IV.c) Circular flow The segment sequence forms a ring (figure 5.10b). Building a ring, a river network is in topological terms not a directed graph. The applied simulation system is not capable of dealing with this situation. Such circles may occur, for example, in estuaries where the flow direction changes over time following a certain probability distribution.

(IV.d) Multiple bifurcation Flow separation of one segment into more than two segments (figure 5.10d). This situation is again unlikely from a hydrological point of view. An error in the base data is to be supposed.

5.4 Raw data processing: Examples

It is not possible to develop a routine for the raw data processing because the starting point is not defined. This section gives examples of the initial data preparation illustrating typical cases.

The preparation of the pre-defined data files including fulfilment of the demands on internal logic can mean strenuous work when performed manually. In many cases the processing of raw data consists of text-based command sequences or of graphical interaction that can usually be substituted by regular

command sequences. Automatic (or semi-automatic) routines are much easier to repeat for the same raw data and if implemented flexibly can be applied to other data of the same raw format. Furthermore, processing steps are most precisely described by the actual command sequences; in descriptions of graphical interaction there remains potential for misinterpretation. This also means that automatic routines would also be easier to perform for a third person without direct support by the original developer and hence should be preferred whenever possible.

Any geographic data are more or less stored in comparable structures and formats. Translation into the required formats consists basically of syntax and unit conversions. In the following, two examples are introduced. The main difference compared to the aspired structure lies in the representation of the river network. In the first example raw data are available in an arc-node model and in the second example in a route system (see also page 61).

While the data sets for river networks represent a complete, enclosed unit, discharge data can be pre-processed to obtain different emission scenarios for comparison, e.g. aggregated discharges vs. real emission situation. Several areal selections from one long list of discharge sites is another example.

Raw Data from MicroLowFlows

General information about MicroLowFlows MicroLowFlows is a software for computing flow statistics at ungauged sites. For this, data from gauging stations, precipitation, information about soil types, elevation data for catchment size estimation and artificial influences (abstractions, discharges) are considered. From this geo-referenced water balance the flow statistics for all ungauged river segments are computed.

MicroLowFlows was developed at the Institute of Hydrology (IH) in Wallingford, U.K. In the framework of the GREAT-ER project this tool has been extended by a special consideration of artificial influences.

Except for detailed sewage treatment data, the database of MicroLowFlows contains all information about the catchment which is required.

The IH has provided small extraction tools for easy retrieval of data. These tools produce simple text files which then have to be further processed to fit the demands of the data composition.

MicroLowFlows export files An extraction process must be given a starting segment. This uniquely determines the catchment. The extraction tools will create simple text files which offer the data in tabular structure. Some demands are already fulfilled in these files, but a reformatting and a mathematical and geographical rework is still necessary. In the following the four extracted text files are described.

Coordinates and attributes This file contains the central data set with the river network. All river segments are included with their coordinates either in UTM or UK national grid projection. In MicroLowFlows a segment has a maximum of 9 coordinates. Only due to this fact of limitation was it possible to represent the river network as a table. Besides the coordinates, the real length and the natural flow (mean and 5th percentile) are provided. MicroLowFlows contains a river network in a lowered resolution. The ”length” information contains the computed length based on the high-resolution digital river network and hence gives a better value than the length that can be computed from the coordinates.

Segment names Pairs of stretch IDs and strings describe the river names. Some names occur frequently in this table.

Artifical influence flows Within the GREAT-ER project, MicroLowFlows has been extended by a special consideration of artificial influences. Examples of these artificial influences are abstractions for drinking water, cooling, irrigation and discharge via treatment plants. The corresponding file contains the improved flow data (mean and 5th percentile) for all segments.

Catchment boundary The computation of the natural flows in MicroLowFlows is based on digital elevation data in grid format. Each square of the area is associated with a height. These data lead to the watershed boundaries. The borderline is given in the file as a line-wise sequence of coordinate pairs. During computation the so-called ”figure of 8” problem may occur. This means that a part of the catchment area is connected to the rest of the catchment via exactly one point (just like the shape of the digit ”8”). In such cases it is recommended to increase or decrease the catchment size (i.e. pick the stretch downstream or upstream from the one which invokes a problem).

Conversion from raw to pre-defined formats Data conversion (see figure 5.11) covers three parts: The relatively simple routines of unit conversion (`awk` scripts) and the projection change from UTM (Universal Trans Mercator) to geographic coordinates (via ArcInfo).

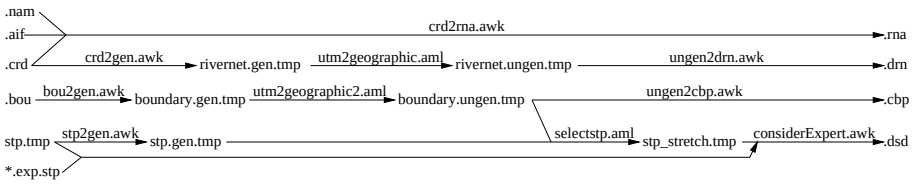


Figure 5.11: *Command sequence for MicroLowFlows data*

A more complicated part is the connection of sewage treatment plants with the river network, because in the present case no location information on STP-river connection is available. For each discharge site the closest river segment is computed and the connection is drawn from the discharge site to the starting point of the river segment.

This may reveal two problems: First, the plant may be connected wrongly, because due to some reason the actual pipeline (or carrier canal) does not follow the shortest distance to the river. Such errors can only be detected and corrected with local expert knowledge (which is given as a table in the files `*.exp.stp`, compare figure 5.11).

Second, the connection of the plant to the starting point of the closest segment is not usually optimal. A more convenient point on the river would be the one providing the shortest distance to the discharge site. The required simulation model only can treat segments as a whole and hence only can assume an emission into the starting point of a segment. Hence for optimization, segments are split up.

For this splitting, several (geo)graphical solutions are possible. A method is chosen that does not produce entirely new points, but which selects the best of the given points of the closest segment. The segment is split into two subsequent stretches; the downstream stretch is then connected with the discharge site (see figure 5.12).

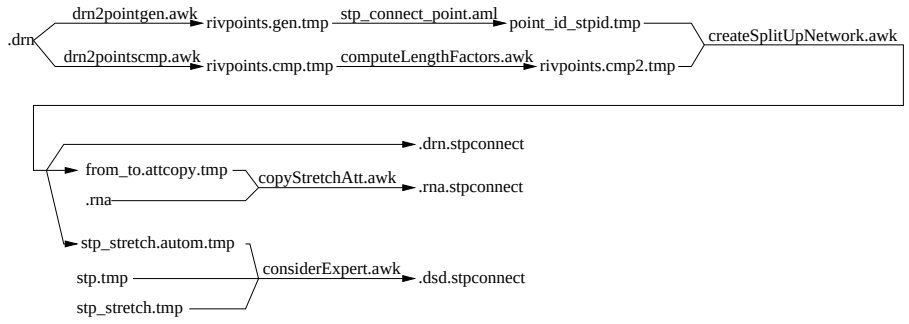


Figure 5.12: *Command sequence for improved discharge connection*

The downstream part is given a new ID. Both segments receive the attributes of the original stretch except, of course, the real length. For computation of the new lengths for the two segments it is assumed that the distances between the single coordinates of the segment is equal. Then factors can be calculated for the lengths. A more precise method than applying length factors is to compute distances from the coordinates.

In general, for geographic distance calculations the projection is of high importance. Only tangential projections (e.g. UTM, UKGRID) allow adequate distance representation within their valid extent.

Raw Data from RhineNet

General information about RhineNet RhineNet is an ArcView-based simulation system for the transport and elimination of chemical intermediates in the River Rhine and its tributaries Main, Mosel and Lippe (Koormann, 1997).

The river network in RhineNet is available as a route system (see page 61). Discharges are ”events” and are associated with certain points on the route system. Also all other spatial changes in attributes (e.g. hydrological data) are stored as events of the route system.

Data conversion The coverage containing the river network routes and events is processed with the ArcInfo command `eventarc` to produce the desired segments (arcs). Next, it is necessary to switch the direction of the segments of the tributaries according to the German system of river measurement (see page 61). The change of direction is done manually in ”ArcEdit”, a sub-module of

ArcInfo. Finally, the geographic data are exported using the command `ungenerate`.

The actual coordinates of the discharge sites were computed with the ArcInfo command `eventpoint` based on the route system and then also exported with the command `ungenerate`. The final format adaptation is done manually and for several parts facilitated by the abilities of advanced text manipulation tools.

5.5 Data processing and quality check: Example

In this section the full procedure of data processing is described for a number of catchments located in the Yorkshire area, UK. Special emphasis lies on the identification and treatment of errors and other problems.

For the first phase of the GREAT-ER project, pilot study areas were chosen to cover situations along the north-south axis of Europe. The selection considered areas where comprehensive data and experience are readily available. For the northern situation, catchments within the Ouse basin in North England were selected: the River Aire including the River Calder as a large, predominantely urban area, the River Went as a rural area and the River Rother (part of the River Don) as a mixed or average situation (compare figure 4.1).

5.5.1 Data parentage

Data availability is quite good for the selected area. The Institute of Hydrology, Wallingford, has developed and maintains a software tool for the estimation of hydrological data at ungauged sites: MicroLowFlows (Gustard et al., 1992, see also previous section). This tool has incorporated a graphical user interface with visualization of the corresponding river networks. Hence, the internal database of MicroLowFlows can deliver the flow data as well as the digital river network. The association of flow attributes to river segments has already been done. Besides, MicroLowFlows also delivers the coordinate sequence of the catchment boundary.

Locating and establishing the properties of the sewage treatment plants in the corresponding area is a complicated task because the existence of such data at several different authorities leads to a number of data sets which differ from

each other in both values and extent. As a starting point, data from the UK Environment Agency (EA) were chosen.

As background data, 1 : 50,000 raster maps of the Ordnance Survey were internally used for plausibility checks. The actual and final background data (in vector format) were downloaded from Pennsylvania State University and are based on data from the U.S. Geological Survey (USGS). Their resolution is about 1 : 1,000,000.

5.5.2 Data review

Review of the delivered data shows that the MicroLowFlows database does not offer any information on flow velocity or estimation method to compute it from other data. Even interpolating methods are not possible because flow velocity measurements are not available for this area in a similar quality as flow measurements. An explicit study for velocity estimation for this area was conducted (Round et al., 1998, equation 3.6).

MicroLowFlows limits the number of river stretch coordinates to 9. The digitial river network in MicroLowFlows (1 : 50,000) is a reduced version of an original data set at a 1 : 20,000 scale. Information on length is copied from the higher resolved data sets and hence only the visual representation of the river network gives an impression of lower resolution.

Discharge information and background data are assumed sufficient from a technical point of view.

5.5.3 Data processing

Besides format conversions, the main task of the initial processing for the Yorkshire catchments is to build an interconnection of sewage treatment plants and river networks. MicroLowFlows includes those STPs which were relevant for hydrological modeling. For the purpose of chemical ”down-the-drain” exposure assessment, all plants must be considered.

Since the discharge sites data are available without information on the actual emission point at the receiving stream, a simple connection method is applied: Each STP is connected to the nearest point of a river stretch (see page 86). It

turns out that this interconnection does not harmonize with the flow data from MicroLowFlows. A cross-check routine based on flows is implemented as part of the final processing to test the corresponding plausibility.

In this test the total effluent flow of the emission point is compared with the mean flow in the connected river segment, its upstream segment(s) and its downstream segment. If the upstream / downstream segment has the same hydrological properties, the segment farther down- / up-stream is taken for comparison. Differences in the flow statistics between discharges and streams usually occur. The reason for this is different time ranges of the underlying statistics.

In general two classes of errors occur: First, the default connection routine selects a stretch upstream or downstream from the one selected for the hydrologic modeling. For the upstream case and a large effluent this would mean a low chemical dilution and hence a simulated but *incorrect peak concentration*.

A second error class is the connection to a wrong stream. For large effluents and small streams this is easily identified due to a high deviation in the flow comparison. The opposite case passes the plausibiliy test. As a consequence *missing peak concentrations* occur due to high dilution. This underlines the need for further tests incorporating, e.g. comparison of simulation results with water quality data.

5.5.4 Data quality

The final data processing stage can not identify all errors. Initial system applications are part of the final quality checks. For the Calder catchment a typical site-specific calibration gives insight into potential problems:

Sewage treatment works are usually improved or extended within a few years. Especially older plants have evolved to a collection of (parallel or sequential) units with different removal efficiencies. The updating of plant type information at a central point is done at a longer interval leading to potentially out-of-date information. For initial applications these potential errors need to be considered and carefully checked.

For the identification of wrong plant types, monitoring data for a readily degradable substance are required. Monitoring effluent concentrations is certainly sufficient for identification, but is rather expensive. Significant errors

can also be identified with in-stream measurements.

This identification method is illustrated for the River Aire where the detergent chemical LAS takes the role of the readily degradable substance. Simulation results for the mean and 90th percentile concentration are compared with the data of a three-year in-stream monitoring study based on monthly measurements (figure 5.13). Simulated and measured concentrations show good accordance for the first 60 kilometers downstream and then start to differ seriously. At this point a waste water treatment plant is located which is simulated as a trickling filter plant (TF plant). TF plants are technologies performing less efficiently than modern ones such as activated sludge plants. As a result a high emission is computed.

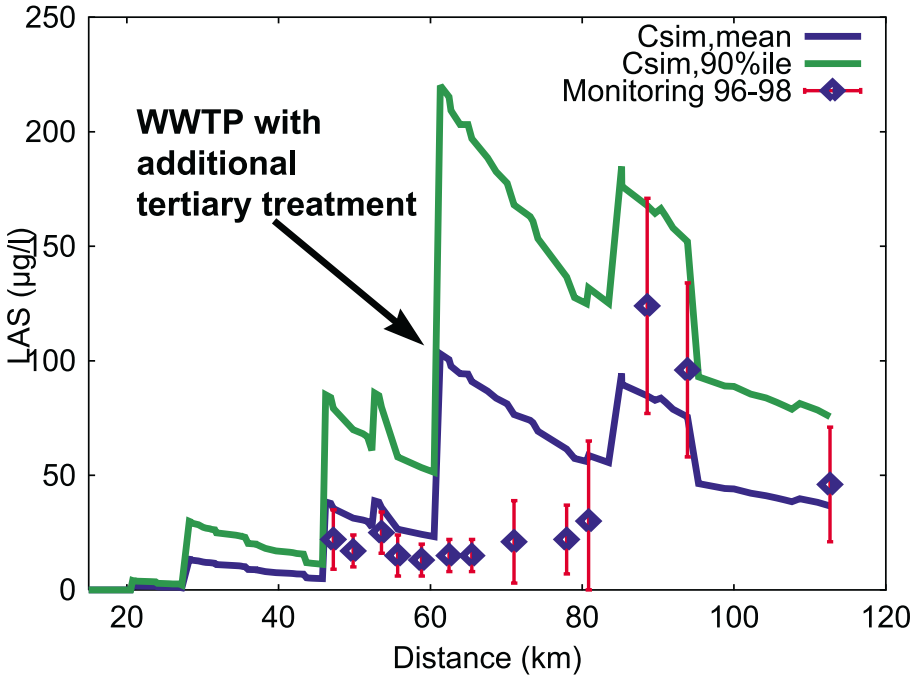


Figure 5.13: *Wrong plant type at the River Aire: Simulation used a simple Trickling Filter plant (TF)*

The monitoring results reflect the real situation where the corresponding plant was recently extended with a tertiary treatment unit providing a higher overall removal efficiency. On-site monitoring of influent and effluent concentrations shows a removal efficiency for LAS of 99.8%. Applying this value for the simulation resolves this site-specific problem and accordance of simulations and monitoring is gained for the whole river Aire (figure 5.14).

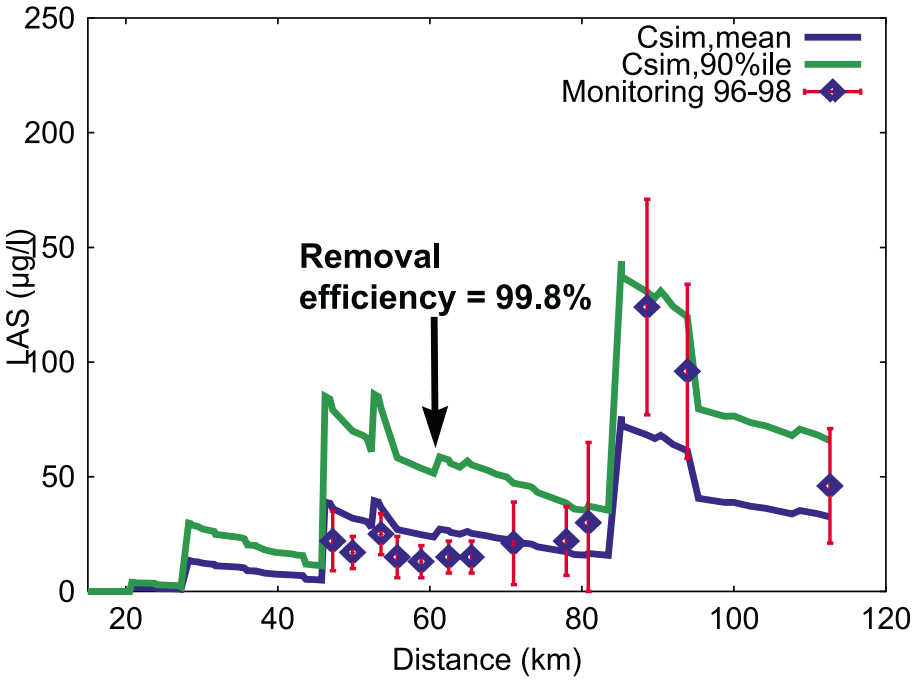


Figure 5.14: *Site-specific calibration*

From this example some conclusions can be derived:

1. Site-specific calibration of a catchment may be required before actual applications can start.

2. Outdated information on plant types results in higher in-stream concentrations. These errors can be significant for plant improvements where chemical upstream load is low. Such information can either be made more reliable through review by local experts. Another method is an adequate monitoring for identification of serious deviations.

3. The planning of monitoring campaigns should consider covering the task of identifying wrong plant types. This regards the location of the sampling site on the one hand and the selection of chemicals on the other hand. Sampling sites should be selected downstream of any treatment plant. The monitoring should include a readily degradable substance that has a wide dispersive use pattern.

Chapter 6

Application and analysis

The simulation tool GREAT-ER consists of a number of single analytically solved models (no numerical solution necessary). These single models are well-known formulations that have been applied to several tasks. Their frequent use has improved the model validity. For GREAT-ER, the models were again discussed with regard to their adequacy within the new task (Boeije, 1999). These studies include generic tests (hypothetical case studies), but are not based on real-world data.

When the GREAT-ER project was officially completed in March 1999, an initial application / calibration study was presented. The intention of this study was to apply the best real-world data available. The initial application for the surfactant ingredients LAS and boron showed simulation results that were well within a factor of 5 compared to the monitoring data. This range was achieved for all monitoring sites of both types: in-stream and final effluent concentrations. An initial calibration further improved the simulation for single catchments. It was achieved by adapting the overall removal distributions of the sewage treatment plants. Calibration was made mainly by visual feedback to parameter shift. It was accompanied by expert knowledge / requests concerning site-specific data.

The GREAT-ER system has integrated several well-reviewed aspects and methods, but a more comprehensive application and analysis study needs to be carried out on it as an entity. The test of the integrated system can be split up into several aspects, some of which are reflected by these general questions:

- Which are the most sensitive parameters?

- How many Monte-Carlo samples need to be applied?
- What is the spatial uncertainty?
- To what extent is the uncertainty accumulated?
- How can the quality of simulation results be judged?

Most of these questions cannot be answered for the GREAT-ER system in general. They must be answered for each simulation task anew. Thus, to support its practical application and analysis, hints, examples, detailed recipes and especially sound methods need to be set up. In this chapter a first application is presented that elaborates general methods and hints on how to approach a GREAT-ER application. It is intended to tackle several aspects. A complete sound method that is generally applicable cannot be developed due to the lack of detailed studies performed to date. More information and experience covering the wide range of application types is required to identify a common basis of adequate methodologies. The intention is to initiate and contribute to a profound basis for further studies that apply the GREAT-ER system or comparable geo-referenced simulation approaches.

Sequence of steps for application and analysis:

First, an assessment of the uncertainties implicated by the GREAT-ER approach is required to judge the reliability of the simulation results (first section). The Monte-Carlo approach involves stochastic uncertainty, which decreases with increasing numbers of Monte-Carlo samples. A method is developed to identify the number of Monte-Carlo samples that are sufficient to assure simulation results within a (tolerable) range are obtained.

To compare the simulation results with measurements with the given complexity, a set of methods for assessing the model exactness is discussed and selected in the second section. Standard absolute and relative measures are selected, but with reference to spatial rather than temporal distribution.

A summary on available monitoring data is given in the third section. The data originate from a monitoring campaign that accompanied the development of GREAT-ER. The presented data are essential for the review of simulation results and are used to assess the model's exactness.

Finally, simulation results are obtained considering the measures for uncertainties, and are then compared with the given monitoring data considering

the measures for model exactness (fourth section). This section also discusses the outcome of the simulated results and of the comparison with measured data. The comparison is carried out spatially explicit (using specific sites) as well as spatially summarized (catchment performance).

6.1 Measures for stochastic uncertainty

In the application of GREAT-ER, it must be considered that the Monte-Carlo approach in itself contributes to the overall uncertainty by introducing a stochastic uncertainty. The lower the number of MC-samples, the higher this stochastic uncertainty is.

Some basic provisions need to be made to deal with this type of uncertainty:

1. Identification of a criterion x to tolerate stochastic uncertainty.

2. Definition of an appropriate measure for x based on the mathematical characteristics of distribution curves obtained from simulations with the same number of MC-samples (e.g. the span of means, as in figure 6.1).

3. Selection of a limiting value of x that may not be exceeded (e.g. an absolute value or relative fraction) based on the above measure.

4. Setting a maximum number of sites (again absolute or relative); limit x must not be exceeded.

As a result, a number s of Monte-Carlo samples can be estimated as sufficient to reach the maximum tolerable stochastic uncertainty.

Steps 1 and 2 are of a more general nature, referring to the simulation method itself, and are treated in this section following the general overview on uncertainties. The latter two steps should be carried out for actual application studies in which explicit requirements on the reduction of stochastic uncertainty are known. Thus, these are discussed in the section dealing with simulation results.

6.1.1 Overview of uncertainties

Uncertainty in the present context is used as a description of the difference between observed (O) and simulated (P) properties, such as single values or distribution curves. It should be kept in mind that O is nothing other than an approximation of the real value (usually better than P). Uncertainties are based on selected task-oriented measures (e.g. $|O - P|$ for $O, P \in \mathcal{R}$ or $|\mu(O) - \mu(P)|$ where O and P are distributions and $\mu(X) = \text{mean}$).

It is differentiated between variability and uncertainty. The variability of properties describes frequencies and ranges of values occuring over time or space respectively. In the following, variabilities are always expressed as distribution curves.

Due to its high complexity, the GREAT-ER system implies several different sources of uncertainty, some of which potentially outnumber others. The contribution of each source of uncertainty has to be assessed independently, if possible. Thus, different sources can be compared in order to identify sources of major and minor importance. However, it is not feasible to predict an overall uncertainty, because no information is available on how the single uncertainties sum up. The overall uncertainty of the model should only be described when based on comparisons with measurements.

Basic uncertainties can be associated

- with the model itself (”model uncertainty”),
- with input parameters (”parameter uncertainty”),
- with measured data (”monitoring uncertainty”) and
- with the Monte-Carlo method (”stochastic uncertainty”).

Model uncertainty

A model is a hypothesis for a real situation considering known aspects. The unconsidered aspects contribute to the uncertainty of the model. For example, consider a simple model for the degradation of a substance: $m(t) = m(0)e^{-kt}$. This model predicts accurately unless the aspects under which the degradation constant k has been obtained are not the same as for the situation to which the

model is applied. The unconsidered aspect could be that, in reality, degradation is not constant over time, as assumed by the model. Of course, the model could be refined to reflect this: $m(t) = m(0)e^{-k(t)t}$. However, it might fail for another situation, because it is still a simplified hypothesis for the real process. This also means that experiments are unavoidable for the detection and reduction of a model's uncertainty in a given situation. However, it is impossible to eliminate this uncertainty entirely.

Parameter uncertainty

Models use one or more input variables which need to be specified for a simulation. The uncertainty or natural variation of such parameters will have a definite influence on the uncertainty of the model's output variables. The effect depends on the sensitivity of the specific parameter (e.g. a small variation of an input parameter might cause a large variation in an output parameter or vice versa). Thus, it is desirable to identify highly sensitive input parameters and to ensure that their uncertainty is low or at least known. In the example given above, two input parameters are used, the initial mass $m(0)$ and the first-order degradation constant k. The parameter $m(0)$ can be given with high precision if the mass originates from a single source (e.g. in a laboratory experiment). However, the consideration of multiple sources in the environment (as in GREAT-ER) might require the estimation of $m(0)$, thus leading to an increase in the uncertainty of this parameter.

Monitoring uncertainty

Concentration measurements underly technical limitations of the applied methods (precision, accuracy, detection limit). Furthermore, each measurement reflects a specific situation in time and space rather than being generally representative. This basically means that both the model output variable and the corresponding measured value used for comparison have a specific uncertainty. In laboratory experiments, the uncertainty of measured values is quite low. Complex field measurements are usually less accurate, because the influence of different environmental parameters can hardly be determined with the same accuracy as in the laboratory. In the example given above the simulated time series $m(t), t = 0..24h$ might correspond well with environmental measurements made on a Tuesday, but due to some reason may not fit for an observation made on a Friday. In such a case, the measurements would need to be repeated several times for all weekdays to judge which is an outlier and

to search for the reason. In the context of GREAT-ER, this could be, e.g. that Monday is a washing day in a particular area and surfactants pass treatment plants with one day later. It may also be that the measurements have identified an unconsidered aspect for the model (e.g. that degradation is variable over time).

Stochastic uncertainty

Many of the input variables required in the GREAT-ER simulations show a natural variability over time and space. A Monte-Carlo approach is used to reflect the temporal variabilities (Boeije, 1999; compare also page 17). The most important parameter determining the precision of the model results is the number of (discrete) Monte-Carlo samples applied. In order to obtain a desired precision of the output distribution curves, a minimum number is required.

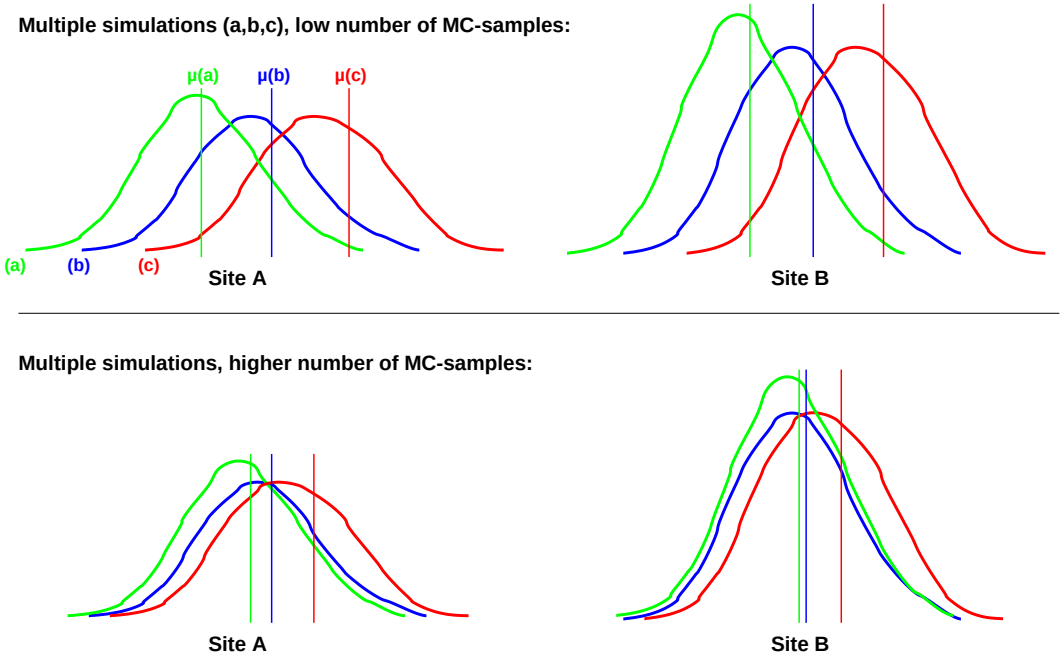


Figure 6.1: *With increasing MC-sample numbers, resulting distribution curves converge (in other words the uncertainty decreases). Vertical lines indicate means $\mu()$ of the corresponding distribution curves.*

Usually a compromise between computational effort and a tolerable uncertainty level is sought, although with current technologies the computational

effort becomes negligible for many applications. With the complexity of GREAT-ER, it remains relevant to find a practical number of MC-samples that suffices for a chosen maximum uncertainty.

Figure 6.1 shows the resulting distribution curves for two sites from multiple simulations illustrating the uncertainty decrease with increasing numbers of MC-samples.

A procedure of how to select an adequate number of MC-samples is descibed in the following section.

6.1.2 Defining a measure for stochastic uncertainty

The error for a Monte Carlo measurement is inversely proportional to the square root of the number of test cases. Thus it is sufficient to find out the uncertainty $x(s_{initial})$ for a reasonably high number of MC-samples $s_{initial}$ and then to deduce the required number of MC-shots s for the aspired tolerable uncertainty x:

$$s = \left(\frac{x(s_{initial})}{x} \right)^2 \cdot s_{initial} \tag{6.1}$$

Identification of criterion x

As a criterion for the tolerable stochastic uncertainty x of a simulation with s Monte-Carlo samples, the maximum deviation from the mean concentration for m locations is chosen:

$$\mu(O_i) - x \quad \leq \quad \overline{P_{i,s}} \quad \leq \quad \mu(O_i) + x \qquad \forall i \in L_m \tag{6.2}$$

where

$P_{i,s}$ = distribution of concentration at site i obtained from s MC-samples
O_i = $P_{i,\infty} \mathrel{\widehat{=}}$ population distribution of concentration at site i
$\overline{P_{i,s}}$ = mean of $P_{i,s}$, (unbiased) estimator for the population mean μ_i
L_m = subset of m locations from set of all locations l in L

Definition of appropriate measures $\hat{x}$ for x

The tolerated maximum stochastic uncertainty x is a spatial aggregation of single x_i for locations i in L_m. Sensible definitions for the measure $\hat{x}$ are based on standard characteristics of the distribution of $\{\widehat{x_i}\}$, such as minimum, mean or maximum, here the extremes and average over all sites in L_m:

$$\widehat{x_{min}} = min(\widehat{x_i}) \quad \forall i \in L_m \tag{6.3}$$
$$\widehat{x_\mu} = \mu(\widehat{x_i}) \quad \forall i \in L_m \tag{6.4}$$
$$\widehat{x_{max}} = max(\widehat{x_i}) \quad \forall i \in L_m \tag{6.5}$$

Hence, it is essential to find a measure $\widehat{x_i}$ for the stochastic uncertainties x_i of locations i to gain measures $\widehat{x_{min}}, \widehat{x_\mu}$ or $\widehat{x_{max}}$.

The distribution of means from a number of distributions of the same type is normal distributed, independent of the type of underlying distributions (Moivre-Laplace limit theorem). Thus, a series of n independent simulations each applying the same number of s MC-samples is performed and the sample standard deviation $\hat{\sigma}$ of the distribution of the corresponding means is taken as a measure for the absolute deviation at location i:

$$\widehat{x_i} = \hat{\sigma}(M_{n,s,i}) \tag{6.6}$$

where $M_{n,s,i}$ = distribution of $\{\overline{(P_{i,s})_j} : j = 1,..,n\}$ (normal distributed independent of the distribution of $P_{i,s}$).

The full definition of the measures $\widehat{x_{min}}, \widehat{x_\mu}$ and $\widehat{x_{max}}$ is therefore:

$$\widehat{x_{min}} = min(\{\hat{\sigma}(\{\overline{(P_{i,s})_j} : j = 1,..,n\}) : i \in L_m\}) \tag{6.7}$$
$$\widehat{x_\mu} = \mu(\{\hat{\sigma}(\{\overline{(P_{i,s})_j} : j = 1,..,n\}) : i \in L_m\}) \tag{6.8}$$
$$\widehat{x_{max}} = max(\{\hat{\sigma}(\{\overline{(P_{i,s})_j} : j = 1,..,n\}) : i \in L_m\}) \tag{6.9}$$

x is a choosen maximum value to tolerate (see also page 113), x with a "ˆ" is a measure. All of the three defined values in 6.7, 6.8 amd 6.9 could be used for comparison against x (to decide whether tolerated value has been exceeded).

Example for measures $\hat{x}$ for stochastic uncertainty

In ECETOC (1999) a rule of thumb is given for GREAT-ER simulations proposing that 500 Monte-Carlo samples should be satisfactory to obtain stable means for simulated concentrations in small / simple catchments, whereas large / complex ones would require 1000 samples. For a stable 90th percentile of the concentration, the numbers 2500 and 5000, respectively, are given.

This proposal is an estimate from previous experiences and is not based on mathematical formulations. Furthermore, it lacks a definition of the term "stable means".

The following example quantifies the stochastic uncertainty for the simple catchment of the River Went and the more complex catchment of the River Calder using the number of Monte-Carlo samples as proposed. For the spatial aggregation all loaded sites are included in L_m.

Table 6.1: Example of stochastic uncertainty ($n = 100$)

	m	s	$\widehat{x_{min}}$	$\widehat{x_{\mu}}$	$\widehat{x_{max}}$
				$[\mu g/L]$	
In-stream sites:					
Went / LAS	37	500	< 0.001	0.012	0.387
Went / LAS	37	1000	< 0.001	0.008	0.262
Calder / LAS	89	500	< 0.001	0.001	0.013
Calder / LAS	89	1000	< 0.001	< 0.001	0.006
Went / Boron	37	500	< 0.001	0.057	1.324
Went / Boron	37	1000	< 0.001	0.033	0.781
Calder / Boron	89	500	0.025	0.812	2.455
Calder / Boron	89	1000	0.014	0.467	1.409
Final effluents:					
Went / LAS	7	500	0.004	0.062	0.178
Went / LAS	7	1000	0.002	0.038	0.108
Calder / LAS	21	500	0.002	0.011	0.026
Calder / LAS	21	1000	0.001	0.006	0.015

m = number of sites
s = number of MC-samples
n = number of simulations
(One simulation is one application of the model performed with s number of MC-samples)

Table 6.1 lists the minimum, average and maximum of the spatially aggregated stochastic uncertanties $\widehat{x_{min}}$, $\widehat{x_{\mu}}$ and $\widehat{x_{max}}$ as defined in equations 6.7, 6.8 and 6.9 respectively. The strongest measure is $\widehat{x_{max}}$, which appears to be most adequate for the comparison of stochastic uncertainties from different scenarios.

Two observations show that it is not feasible to give a general rule for the number of Monte-Carlo samples independent of the actual scenarios:

- The variation in $\widehat{x_{max}}$ between LAS scenarios for different groups of sites (in-stream and final effluents) is approximately a factor of 2 for the River Went and a factor of 0.5 for the River Calder. Thus the absolute stochastic uncertainty differs by a factor of 4 for two different catchments with the same substance.

- The scenarios Went / Boron / 500 MC-samples and Calder / Boron / 1000 MC-samples have almost the same stochastic uncertainty $\widehat{x_{max}}$, which corresponds to the rule given by (ECETOC, 1999). In contrast, the scenarios Went / LAS / 500 MC-samples and Calder / LAS / 1000 MC-samples show a variation in $\widehat{x_{max}}$ of a factor of around 60. This shows that the rule does not apply for LAS and thus does not apply in general.

Note that the stochastic uncertainties can be significantly higher if incorrect data are applied: The inclusion of 3 problematic stretches of the River Calder, as described on page 51, would lead to a value $\widehat{x_{max}}$ = 162.347 for Calder / Boron using s = 500 MC-samples. Thus, the stochastic uncertainty can also be an indicator for the integrity of the underlying dataset.

6.2 Measures for model exactness

Comparison of simulation results and reliable monitoring data allows the exactness of the applied model to be described. For the Calder catchment, statistically sufficient numbers of measured concentrations have been collected in a comprehensive monitoring campaign for several sites in the catchment (in-stream and final effluents). To quantify their exactness, or at least to provide indicators for their exactness, measures for the statistical discrepancy of model and monitoring results can be used. As described in Gayler (1998), the general discrepancy can be expressed as:

$$D = D\{P(t_i), O(t_i)\} \qquad i = 1, ..., n \tag{6.10}$$

where t_i are the points in time for the measurements $O(t_i)$ and the model prediction $P(t_i)$. Based on measures from the type of equation 6.11, the discrepancy measure can also be expressed as an absolute discrepancy (equation 6.12).

$$M_\tau = \left(\sum_{i=1}^{n} |x_i - y_i|^\tau \right)^{\frac{1}{\tau}} \qquad \tau = 1, ..., \infty \tag{6.11}$$

$$D = \frac{1}{n} \left(\sum_{i=1}^{n} |P(t_i) - O(t_i)|^\tau \right)^{\frac{1}{\tau}} \qquad \tau = 1, ..., \infty \tag{6.12}$$

where $\frac{1}{n}$ normalizes sample sets of different lengths. The value of τ is a weighting for deviation, $\tau = 1$ gives all pairs $\{P(t_i), O(t_i)\}$ an equal weight while $\tau = \infty$ will consider only the one most extreme outlier.

For the GREAT-ER model and for monitoring, the vectors of predicted and observed values are not given as a sequence for different points over time. In GREAT-ER, time is already incorporated in the variability mechanism. A direct comparison of a site in time would require a comparison of the distribution curves of the predicted concentration and the monitoring results. While the distribution for the predicted concentration could be obtained at a stable level (by issuing a sufficiently high number of Monte-Carlo samples), the monitoring data are limited in their extent (for this study between 7 and 38 samples). Sampling distributions offer higher stability around the mean than for high / low percentiles. This means that the model exactness for specific situations over time (normal situation, e.g. average weather conditions) can be better assessed than others (extreme or rare situations, e.g. storms). The site-specific temporal model exactness is not further addressed here. Instead of time, the focus is on spatial aspects. Hence, t_i is replaced by s_i, which describes samples at different points in space. The medians of the concentration distribution curves $\widetilde{C_{\mathrm{sim}_i}}$ and $\widetilde{C_{\mathrm{mon}_i}}$ are used as the most stable values for $P(s_i)$ and $O(s_i)$. The discrepancy measure then describes a spatially aggregated model exactness which could be applied to different sets of objects (final effluents or in-stream sites). It would be conceivable to include any object of a catchment into a measure, but it will provide fewer options to interpret the exactness of single sub-models.

Based on (Gayler, 1998), four measures are selected (n is now the number of locations):

Parameter D_a (equation 6.13) describes the average deviation. This measure supports the identification of systematic errors. As all points are of equal weight, only a general over- or underprediction would result in extreme values of D_a. Good agreement would lead to D_a approaching 0, but this could occur with equal fractions of over- and underprediction as well. Hence, if no extreme outliers are present, this method can verify systematic overprediction ($D_a >> 0$) or systematic underprediction ($D_a << 0$). D_a is accompanied by $D_{|a|}$ (equation 6.14), which is the absolute average deviation. $D_{|a|}$ describes the average distance from the ideal situation ($\widetilde{C_{\mathrm{sim}_i}} = \widetilde{C_{\mathrm{mon}_i}} \quad \forall i$) and if no extreme outliers are present it will approach 0 the better the agreement is. Of course this only holds true if the distribution curves for monitoring and simulation are of the same type.

$$D_a = \frac{1}{n} \sum_{i=1}^{n} \left(\widetilde{C_{\mathrm{sim}_i}} - \widetilde{C_{\mathrm{mon}_i}} \right) \tag{6.13}$$

$$D_{|a|} = \frac{1}{n} \sum_{i=1}^{n} \left| \widetilde{C_{\mathrm{sim}_i}} - \widetilde{C_{\mathrm{mon}_i}} \right| \tag{6.14}$$

Another commonly used absolute measure is the average squared deviation D_{rms} (equation 6.15, RMS = Root Mean Square) which results from equation 6.12 with $\tau = 2$ giving outliers a higher weight than $D_{|a|}$.

$$D_{rms} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} \left(\widetilde{C_{\mathrm{sim}_i}} - \widetilde{C_{\mathrm{mon}_i}} \right)^2} \tag{6.15}$$

Applying absolute deviations across inhomogenously high, site-specific values would neglect exactness at sites of lower values. For example, with the treatment situation in Yorkshire, two types of sewage treatment plants (activated sludge, trickling filter) result in two different ranges of concentrations (for both substances, LAS and boron, the typical final effluent concentration differs between AS and TF-plants by a factor of 5). To deal with this and to compare different scenarios, a relative measure is required.

Equation 6.16 describes the average relative deviation, equally weighting over- and underpredictions and equally weighting all sites ($\tau = 1$).

$$D_{|a|,r} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| \widetilde{C_{\mathrm{sim}_i}} - \widetilde{C_{\mathrm{mon}_i}} \right|}{\min(\widetilde{C_{\mathrm{sim}_i}}, \widetilde{C_{\mathrm{mon}_i}})} \tag{6.16}$$

6.3 Available data from monitoring campaign

Associated with the development of the GREAT-ER system, an extensive monitoring program was performed in four Yorkshire catchments (Rivers Aire, Calder, Don / Rother and Went) and in the Lambro catchment (Italy).

Monitoring in Yorkshire covered boron and LAS in the final effluent of sewage treatment plants and at selected in-stream sites. Samples were taken monthly from 1996 to 1998. All data used in the following discussion can be found on the official ECETOC GREAT-ER 1.0.1 software distribution. Parts of the trickling filter monitoring program are published in (Holt et al., 1998).

Similar to the sewage treatment plants, (x,y)-coordinates for the sampling sites were given and the connection of these sites to the digital river network was partly done automatically, because for some cases the locations were not directly at river segments. This implicates potential spatial uncertainty for compared monitored and simulated concentrations.

6.3.1 Calder

In the Calder catchment, final effluent (FE) concentrations from 14 plants were measured (table 6.2). The measurements for one plant were omitted due to non-representative data.

Additionally, in-stream samples were taken at 15 sites in the main stream and at 3 sites in the tributaries Spen, Colne and Red Beck. The chosen downstream distance from discharges makes the assumption of (almost) full mixing reasonable (see table 6.3 and figure 6.2).

The in-stream monitoring site *Allerton Bywater* shown in the GREAT-ER software does not actually belong to the Calder stream. The samples were taken from the River Aire close to the confluence with the River Calder. Thus, in this study *Allerton Bywater* will only appear for the River Aire.

Table 6.2: Monitoring results for STP's in the Calder catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Eastwood	204.0	132.0	35	218.5	72.3	35
Redacre	514.0	374.0	23	276.6	100.3	23
Highroyd	96.0	37.0	7	239.1	95.7	7
Sowerby Bridge[1]	11.0	12.0	24	262.2	146.4	24
Halifax	69.0	67.0	41	598.8	253.4	40
Shibden Head	488.0	245.0	27	458.9	244.0	27
Brighouse	601.0	443.0	11	416.8	113.1	11
Neiley	344.0	230.0	17	413.5	105.9	17
Huddersfield	108.0	80.0	28	393.1	170.3	28
North Bierley	203.0	146.0	22	484.0	157.0	22
Dewsbury[2]	539.0	157.0	28	1116.3	375.1	28
Ossett Spa	358.0	314.0	16	691.5	155.9	16
Caldervale	22.0	15.0	17	682.4	231.0	17

N = number of samples

[1] called Milner Royd STW in the original monitoring data set

[2] called Mitchell Laithes STW in the original monitoring data set

Table 6.3: In-stream monitoring results in the Calder catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Calder Portsmouth	3.0	2.0	25	26.1	5.2	25
Calder Heptonstall	35.0	24.0	23	56.6	30.4	24
Calder Hebden Bridge	21.0	24.0	25	33.7	11.4	25
Calder Brearley Weir	28.0	20.0	25	40.9	19.4	25
Calder Sowerby Bridge	20.0	8.0	21	48.9	19.6	21
Calder Copley Bridge	27.0	20.0	23	54.4	19.9	23
Calder North Dean	35.0	22.0	24	144.1	73.0	24
Red Beck at Brookfoot	17.0	21.0	26	154.1	60.9	27
Calder Rastrick Bridge	23.0	19.0	18	150.8	107.8	18
Calder Cooper Bridge	26.0	18.0	22	128.6	61.2	22
Colne at Colne Bridge	21.0	19.0	25	97.9	135.5	25
Calder Battyeford	64.0	31.0	24	151.9	69.8	24
Spen A644	80.0	42.0	24	416.8	197.1	24
Calder Dewsbury	61.0	37.0	22	168.1	82.9	22
Calder Horbury Bridge	78.0	34.0	22	206.8	109.8	22
Calder Kirkgate	68.0	29.0	23	213.9	108.1	23
Calder Stanley Ferry	59.0	26.0	22	221.3	125.1	22
Calder Methley Bridge	39.0	25.0	23	244.7	109.9	23

N = number of samples

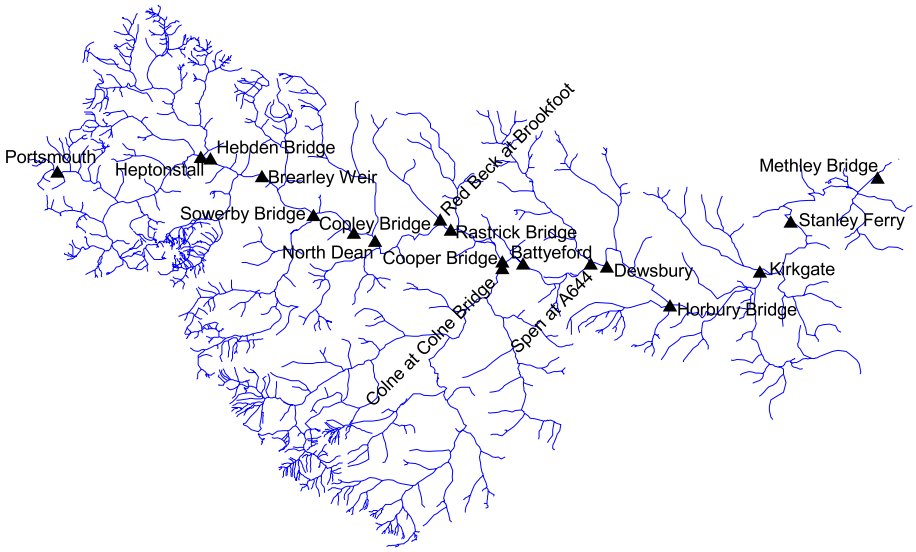


Figure 6.2: *In-stream monitoring sites in the Calder Catchment*

6.3.2 Aire

In the Aire catchment, 8 plant effluents were monitored (table 6.4) in addition to those in the Calder subcatchment. No data exist for 5 plants in the monitoring program.

Table 6.4: Monitoring results for STP's in the Aire catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Gargrave	223.0	87.0	14	455.1	176.3	14
Skipton (Snaygill)	218.0	93.0	30	450.5	167.4	27
Marley	183.0	64.0	34	504.4	201.6	34
Dowley Gap	504.0	154.0	30	427.0	150.3	30
Esholt	14.0	13.0	36	459.7	148.0	37
Knostrop	354.0	96.0	28	495.4	181.2	29
Wheldale	173.0	88.0	9	592.2	156.8	9
Sutton	361.0	117.0	22	837.1	185.2	22

N = number of samples

All in-stream monitoring sites are located along the main stream (table 6.5 and figure 6.3) of the River Aire. No data on the chemical load of the tributaries to the River Aire have been collected, except for the *Methley Bridge* site on its major tributary, the River Calder.

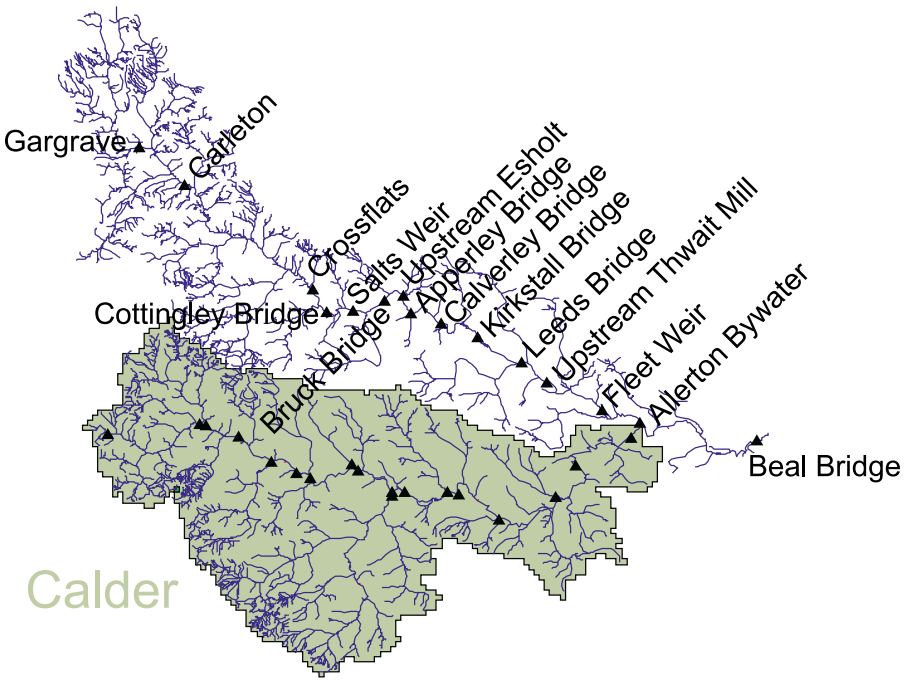


Figure 6.3: *In-stream monitoring sites in the Aire Catchment*

Table 6.5: In-stream monitoring results in the Aire catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Aire Gargrave	n.d.		9	20.0	0.0	9
Aire Carleton	n.d.		11	30.8	9.7	11
Aire Crossflatts	22.0	13.0	14	99.1	51.3	14
Aire Cottingley Bridge	17.0	7.0	16	109.6	61.4	16
Aire Salts Weir	25.0	9.0	15	107.7	55.4	15
Aire Buck Bridge	15.0	9.0	16	112.8	56.1	16
Aire above Esholt	13.0	7.0	13	113.2	60.6	13
Aire Apperley Bridge	15.0	7.0	19	177.6	93.6	19
Aire Calverley Bridge	15.0	7.0	14	157.1	89.1	14
Aire Kirkstall Bridge	21.0	18.0	15	164.1	74.6	15
Aire Leeds Bridge	22.0	15.0	17	163.8	70.6	17
Aire US Thwaite Mill	30.0	35.0	23	167.6	78.7	23
Aire Fleet Weir	124.0	47.0	28	246.9	117.5	28
Aire Allerton Bywater	96.0	38.0	38	279.6	98.2	38
Aire Beal Bridge	46.0	25.0	25	259.0	118.0	25

n = Number of samples
n.d. = not detected

6.3.3 Don/Rother

No data exist for two plants emitting into the River Don and one at the Rother in the monitoring program. The remaining 6 plants cover 4 TF and 2 AS types (table 6.6).

In-stream monitoring primarily addressed the Rother and also covers one tributary, the Drone (table 6.7 and figure 6.4).

Table 6.6: Monitoring results for STP's in the Don / Rother catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Holbrook	218.0	82.0	25	831.8	223.7	26
Woodhouse Mill	n.d.		25	987.0	277.1	26
Long Lane	577.0	201.0	26	987.5	222.3	26
Danesmoor	697.0	353.0	26	986.3	329.6	26
Tupton	101.0	138.0	26	1057.1	272.4	26
Old Whittington	n.d.		26	704.0	151.4	26

N = number of samples
n.d. = not determined

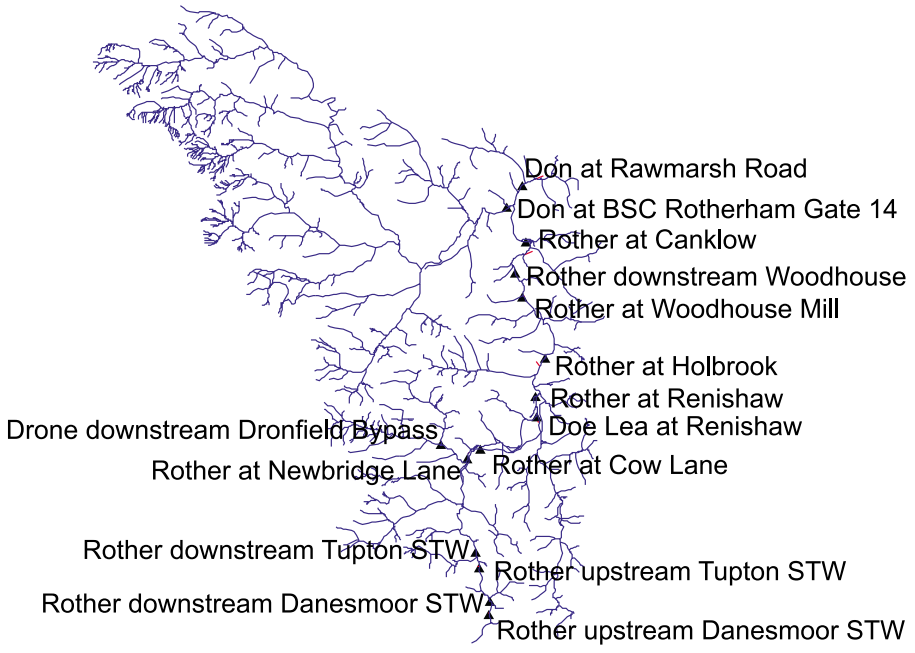


Figure 6.4: *In-stream monitoring sites in the Don/Rother Catchment*

Table 6.7: In-stream monitoring results in the Don / Rother catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Rother US Tupton STW	32.0	27.0	20	308.2	147.9	20
Rother DS Tupton STW	46.0	32.0	20	311.4	159.4	20
Drone DS Dronfield Bypass	86.0	48.0	20	266.9	99.6	20
Rother at Newbridge Lane	25.0	18.0	21	180.8	74.5	21
Rother at Cow Lane	14.0	7.0	19	317.5	141.1	20
Doe Lea at Renishaw	21.0	40.0	19	408.9	155.2	19
Rother at Renishaw	25.0	38.0	21	295.0	133.2	21
Rother at Holbrook	30.0	29.0	20	328.9	182.5	20
Rother at Woodhouse Mill	26.0	20.0	20	299.2	118.5	20
Rother DS Woodhouse Mill STW	59.0	70.0	21	334.1	142.5	21
Rother at Canklow	50.0	54.0	20	353.6	138.4	20
Don at BSC Rotherham Gate 14	15.0	20.0	18	287.6	118.4	18
Don at Rawmarsh Road	21.0	15.0	19	441.8	311.4	19
Rother US Danesmoor STW	3.0	3.0	20	106.0	30.7	21
Rother DS Danesmoor STW	239.0	176.0	21	512.2	259.6	21

N = number of samples
US = upstream, DS = downstream

6.3.4 Went

The final effluent monitoring for the Went catchment has considered all of the present treatment plants (table 6.8).

In-stream monitoring in the Went catchment covers 6 sites at the main stream and one tributary, the Little Went (table 6.9 and figure 6.5).

Table 6.8: Monitoring results for sewage treatment works in the Went catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Fitzwilliam	322.0	353.0	18	891.2	372.4	18
Ackworth	6.0	7.0	26	785.4	210.9	26
Carleton	413.0	291.0	19	747.5	193.0	19
Kirk Smeaton	153.0	129.0	20	1092.4	253.7	20
Askern Norton	199.0	156.0	18	720.4	128.8	18
Cridling Stubbs	197.0	296.0	19	1156.5	241.9	19
Stapleton Park	196.0	59.0	19	795.2	146.0	20

N = number of samples

Table 6.9: In-stream monitoring results in the Went catchment

	LAS			Boron		
	Mean	St.Dev.	N	Mean	St.Dev.	N
	$\mu g/l$			$\mu g/l$		
Went upstream Hardwick Beck	31.0	71.0	23	179.1	51.1	23
Went at Ackworth	6.0	7.0	22	202.6	65.5	22
Went at Standing Flats Bridge	10.0	21.0	23	315.7	152.5	23
Little Went at Hardwick Road	137.0	74.0	19	530.0	209.8	19
Went at Went Bridge	12.0	18.0	22	311.6	136.8	22
Went at Walden Stubbs	8.0	10.0	22	269.0	95.2	22
Went at Sykehouse	10.0	17.0	24	244.1	81.5	24

N = number of samples

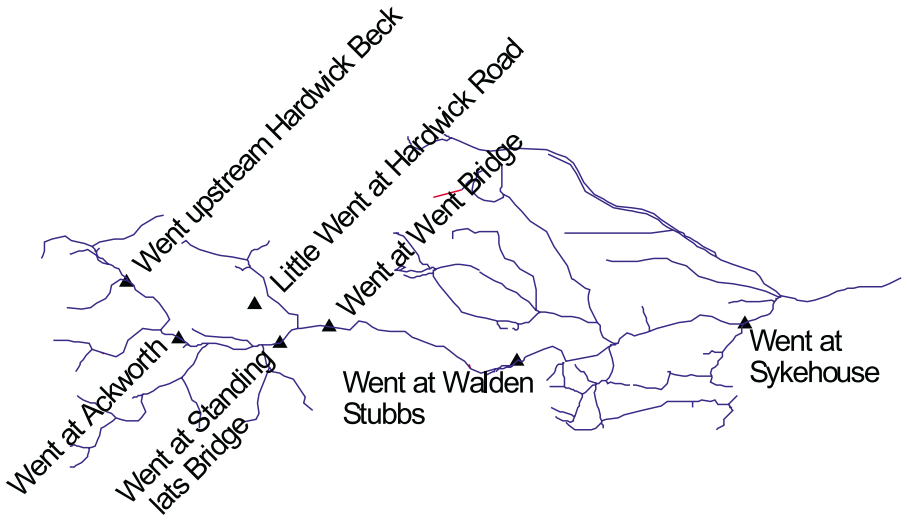


Figure 6.5: *In-stream monitoring sites in the Went Catchment*

6.4 Simulation results and discussion

In this section GREAT-ER simulation results are compared with the monitoring data. Both datasets consist of a number of site-specific log-normal concentration distribution curves, given by the mean and standard deviation of the underlying single values. In general, a comparison can be made based on several properties of the distributions: e.g. mean, median, mode and even more precise methods that consider the actual curves. The selection of appropriate parameters mainly has to take into account two aspects:

1. The uncertainty of the data:

   While for the simulation results the stability of the distribution curve parameters (i.e. mean) can be increased by increasing the number of Monte-Carlo samples, the monitoring data are based on a rather limited number of single values. Usually, the median is the most stable value since it gives less weight to extreme outliers.

2. The objective of the comparison:

   If the focus lies on 'usual' or 'typical' concentrations, the mode (the most frequently occurring value) would be an appropriate value to incorporate for the meaning of presented results while tolerating a slightly higher uncertainty. For risk assessment purposes, intervals including a certain percentage of all values (e.g. 90th percentile) are of interest rather than the median.

The present analysis is intended to examine the average (=mean) concentrations in order to learn more about the principle adequacy of the applied methods and to find out systematic or site-specific errors. The monitoring data have undergone a multiple outlier test ("mot"; Fox, 1998) which eliminated extreme values. Hence, the mean's stability is considered acceptable and is therefore taken as the basis for comparison.

6.4.1 Selection of numbers of Monte-Carlo samples

Depending on the intention for which a GREAT-ER simulation is performed, the number of required Monte-Carlo samples (MC-samples) ranges from a relatively low number which is sufficient for identification of a rough hot spot concentration level, to medium numbers used e.g. for the determination of

whether a PNEC is potentially exceeded, and to high numbers for a direct comparison with high-quality measurements.

In section 6.1.2 (page 99), a criterion x for stochastic uncertainty is identified and appropriate measures $\hat{x}$ for x are defined, of which $\widehat{x_{max}}$ (equation 6.9) was argued to be the most adequate for the present study. Besides this predetermination concerning the handling of stochastic uncertainty, two further selections need to be made:

Selection of the limiting value x for the maximum tolerated stochastic uncertainty and setting the maximum number of sites for which x must not be exceeded.

Selection of the maximum tolerated stochastic uncertainty

This study compares simulations with measurements and thus quite low tolerated x are selected to minimize the stochastic uncertainty:

$$x(LAS) = 0.05\mu g/L \quad (6.17)$$

$$x(Boron) = 0.25\mu g/L \quad (6.18)$$

The value for boron is chosen 5 times higher than the value for LAS because the observed in-stream concentrations for boron are much (about ten times) higher than for LAS. Note that in final effluents the stochastic uncertainty is 0 for boron in the absence of elimination processes.

Setting the number of sites to be within the tolerated stochastic uncertainty

For the present analysis a criterion is chosen that refers to the stretch with the highest stochastic uncertainty. In other words, the number of Monte-Carlo samples will be determined according to the worst performing site in terms of stochastic uncertainty. Most sites of the scenario would require far less MC-samples to achieve the same quality. Thus a possible alternative could be a percentage of sites considered sufficient (e.g. tolerate 5% of the sites to exceed tolerated limits).

According to equation 6.9, this means that any loaded site is included in L_m.

Initial test for stochastic uncertainty

In order to find a sufficient number of MC-samples achieving the chosen tolerated stochastic uncertainty, initial test simulations are performed for each scenario. 2000 MC-samples and 100 simulations are selected as a reasonable basis.

Table 6.10: $\widehat{x_{max}}$ for initial test with $s_{initial}$ = 2000 and n = 100

Scenario	Sites	m	$\widehat{x_{max}}$ $[\mu g/L]$
Went / LAS	in-stream	37	0.128
	final effluent	7	0.038
Don / LAS	in-stream	60	0.073
	final effluent	9	0.017
Calder / LAS	in-stream	89	0.002
	final effluent	21	0.007
Aire / LAS	in-stream	201	0.101
	final effluent	34	0.013
Went / Boron	in-stream	37	0.260
Don / Boron	in-stream	60	0.316
Calder / Boron	in-stream	89	0.580
Aire / Boron	in-stream	201	0.781

m = number of sites
$\widehat{x_{max}}$: see equation 6.9, used for $x(s_{start})$ (see below)

The initial test described in table 6.10 shows that for all scenarios, except for Calder / LAS, in-stream sites dominate final effluents in terms of contribution to stochastic uncertainty. With the exception of Calder / LAS, all scenarios need a higher number of Monte-Carlo samples to achieve the aspired quality.

Comparing different sites (and different scenarios) in terms of their required Monte-Carlo samples is also a useful source of information for general uncertainty analysis (uncertainty linked to catchment of linked to substance?, which model or dataset includes higher uncertainty (e.g. in-stream vs. final effluent)?, what are the relative differences between scenarios, sites in terms of uncertainty?).

Estimation of the required number of MC-samples

With the initial MC-samples s_{start} = 2000 and the maximum tolerated stochastic uncertainty x = 0.05 for LAS and x = 0.25 for boron, respectively, equation 6.1 was used to deduce the required number of MC-samples:

Table 6.11: Estimated number of MC-samples s for all scenarios

Scenario	$x(s_{start})$ $[\mu g/L]$	s
Went / LAS	0.128	13,107
Don / LAS	0.073	4,264
Calder / LAS	0.007	39
Aire / LAS	0.101	8,161
Went / Boron	0.260	2,163
Don / Boron	0.316	3,195
Calder / Boron	0.580	10,765
Aire / Boron	0.781	19,518

$x(s_{start})$ = max($\widehat{x_{max}}$(in-stream), $\widehat{x_{max}}$(final effluent)) (table 6.10)
$s = \left(\frac{x(s_{start})}{x}\right)^2 \cdot s_{start}$ (equation 6.1)

The values of s shown in table 6.11 are more or less another representation of the measured stochastic uncertainties shown in table 6.10, but they geometrically stress higher values. However, both tables show that great differences in the stochastic uncertainty occur due to characteristic details of the scenario rather than the complexity or size of the catchment (e.g. for the small Went catchment a much larger number of Monte-Carlo samples is required to stay below the same low stochastic uncertainty for LAS than for the larger Calder catchment).

All simulations in this study apply sufficiently high numbers of MC-samples chosen individually for each scenario according to table 6.11.

6.4.2 Final effluent concentrations

Mean simulated and measured concentrations for final effluents are directly compared in figures 6.6 to 6.13. The two different STP types are distinguished by different symbols. Those plants not covered by the monitoring program are indicated with their simulated value to the right of the right vertical axis. For simulation results in tabular form, see Appendix A.1 (page 152).

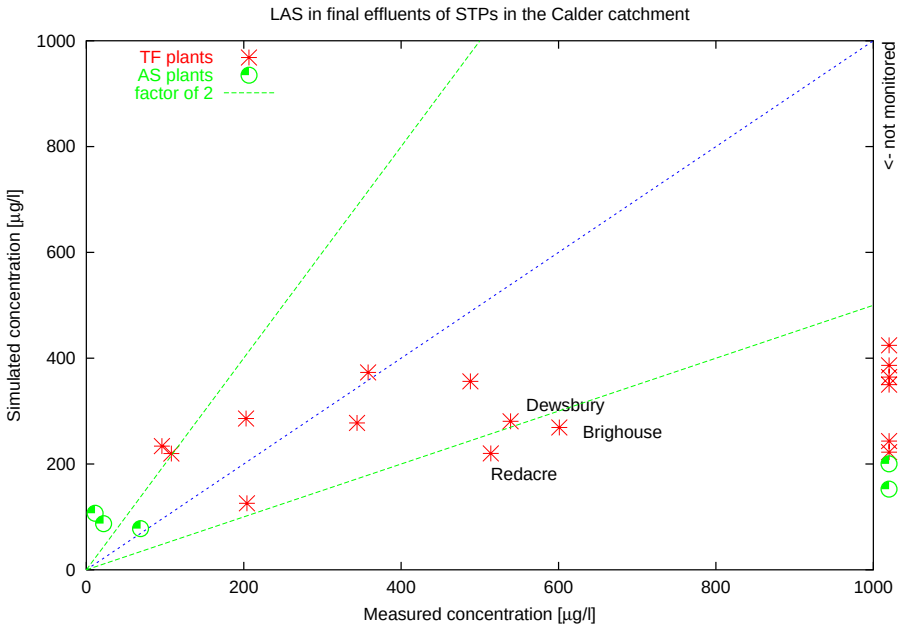


Figure 6.6: *LAS in the Calder: Mean FE Concentration*

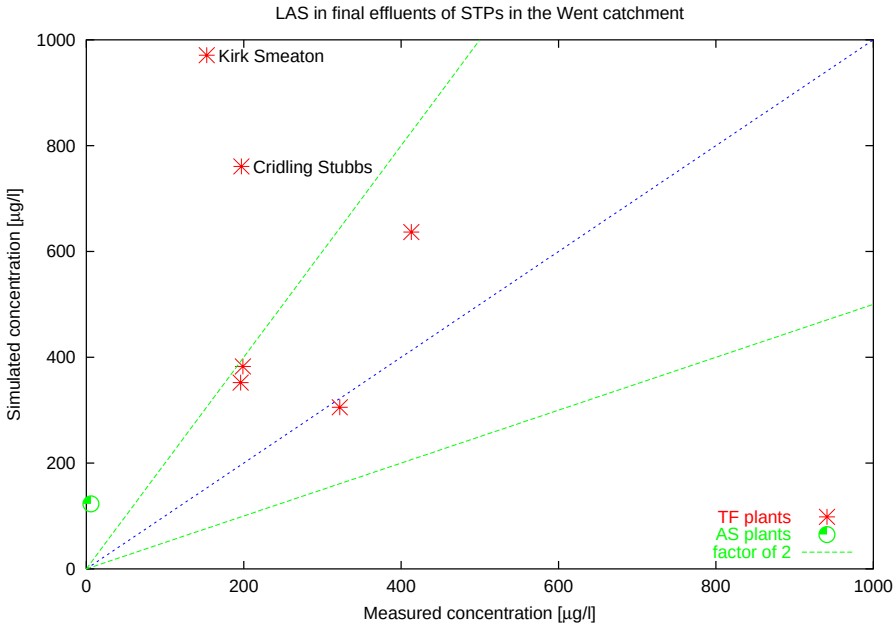


Figure 6.7: *LAS in the Went: Mean FE Concentration*

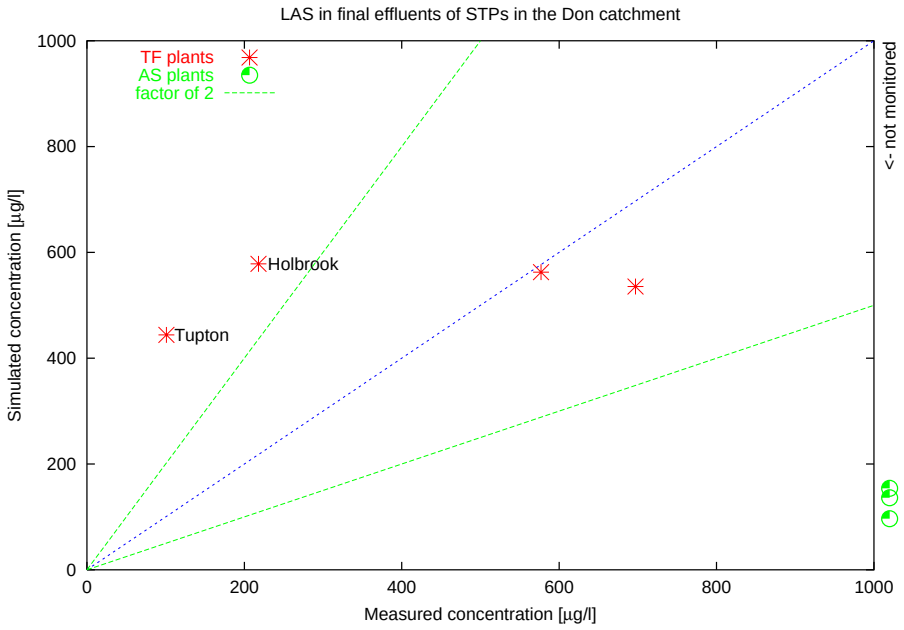


Figure 6.8: *LAS in the Don/Rother: Mean FE Concentration*

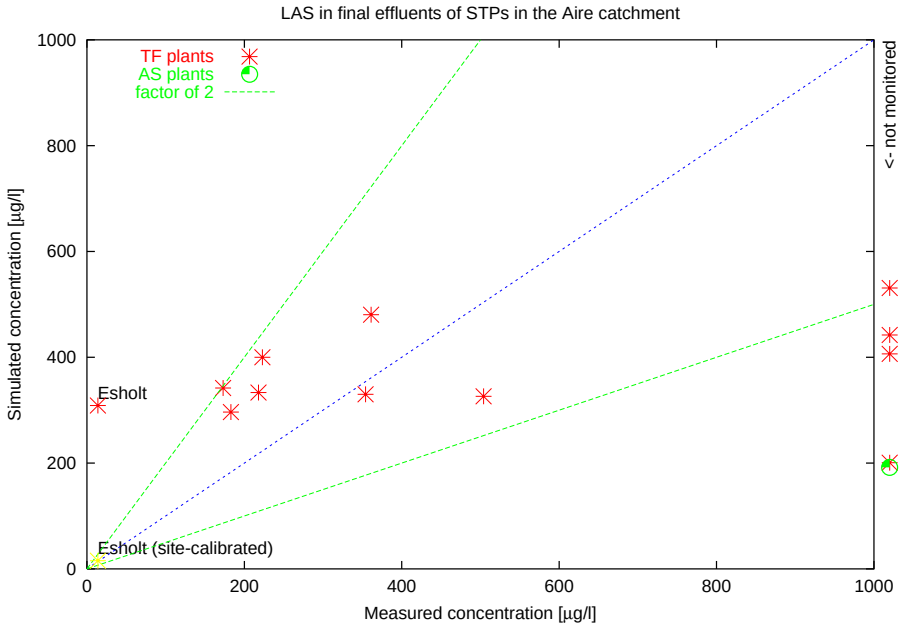


Figure 6.9: *LAS in the Aire: Mean FE Concentration*

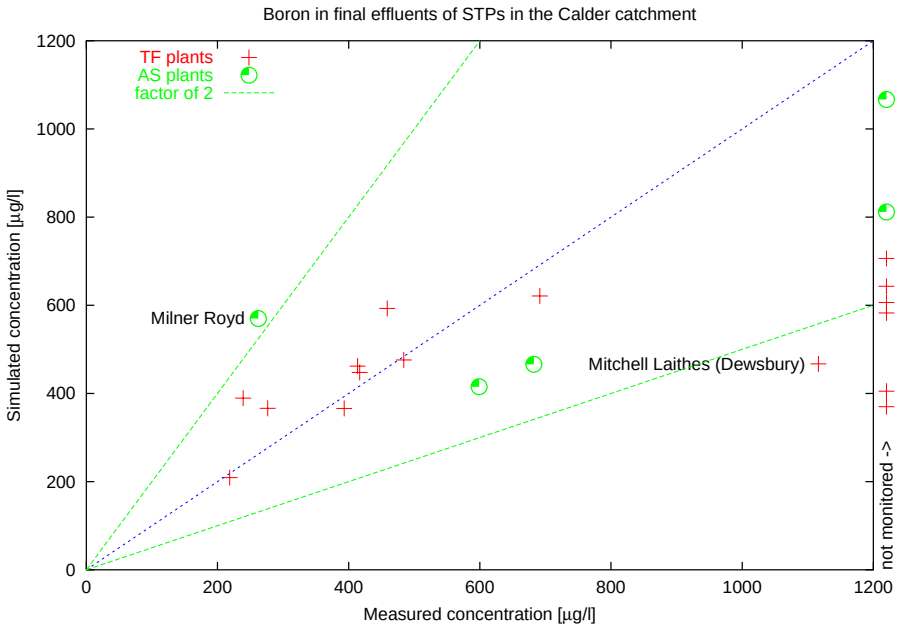


Figure 6.10: *Boron in the Calder: Mean FE Concentration*

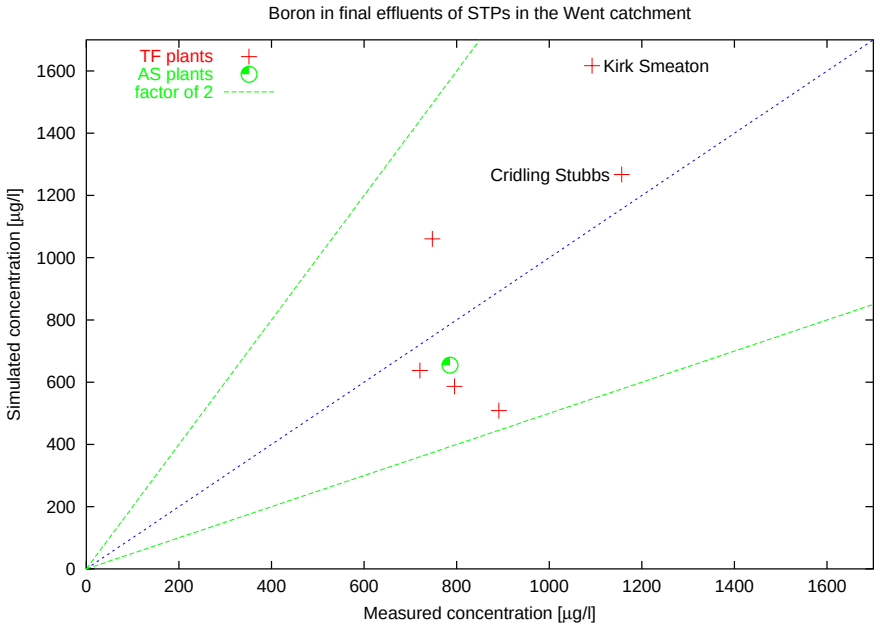


Figure 6.11: *Boron in the Went: Mean FE Concentration*

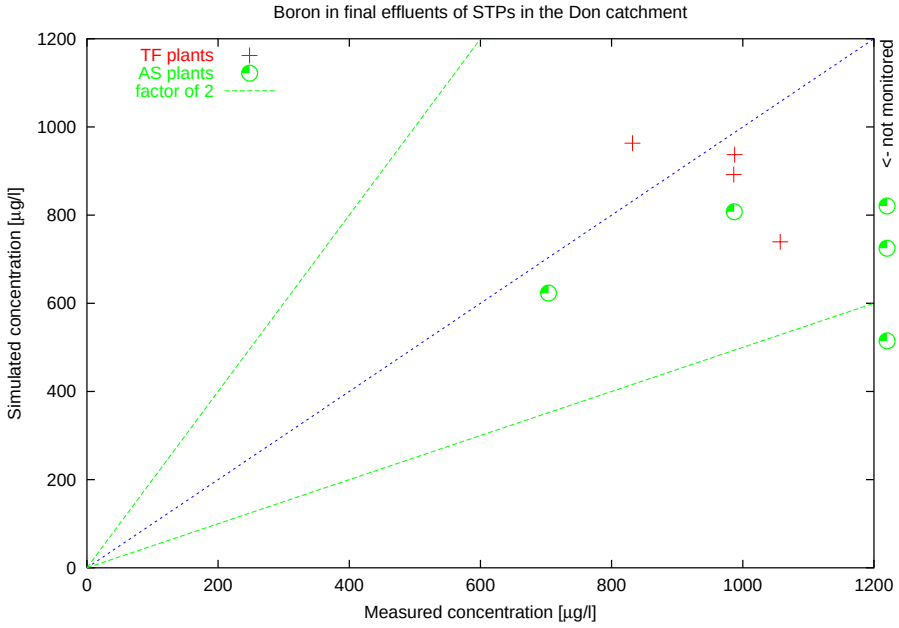


Figure 6.12: *Boron in the Don: Mean FE Concentration*

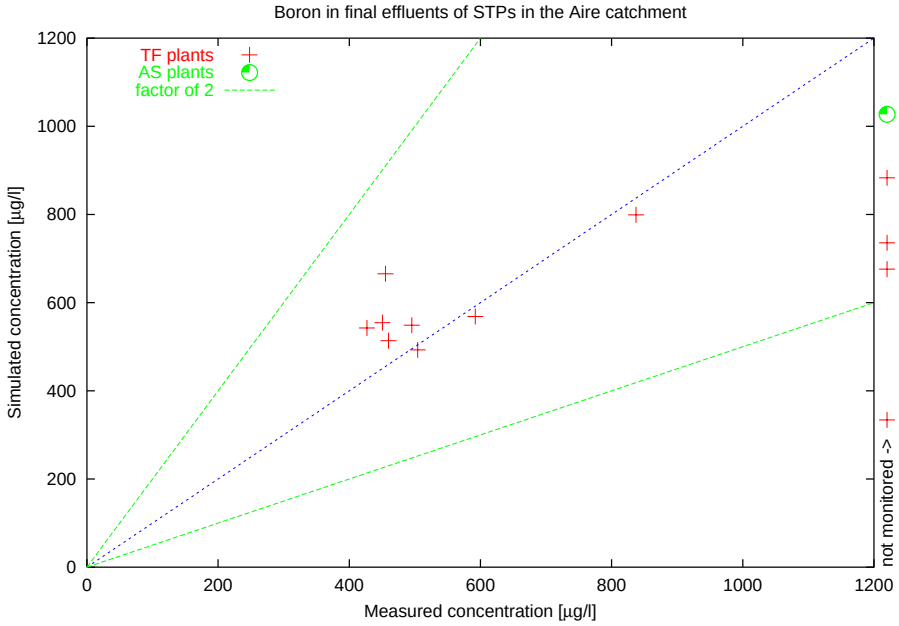


Figure 6.13: *Boron in the Aire: Mean FE Concentration*

Calculated measures for the model exactness based on the simulation results and available monitoring data are shown in table 6.12. This table summarizes the absolute measures D_a, $D_{|a|}$ and D_{rms} (equations 6.13, 6.14 and 6.15, page 104) and the relative measure $D_{|a|,r}$ (equations 6.16, page 104).

Table 6.12: Assessment of model exactness for median FE concentrations

	D_a	$D_{\|a\|}$	D_{rms}	$D_{\|a\|,r}$
Calder / LAS	-15.288	114.113	37.152	1.857
Went / LAS	323.269	323.269	157.384	6.384
Don / LAS	151.838	207.539	129.886	2.023
Aire / LAS	28.319	126.777	31.817	2.673
Aire / LAS calibrated	15.625	113.474	28.768	1.367
Calder / Boron	-2.241	145.605	59.593	0.423
Went / Boron	48.593	244.208	109.716	0.332
Don / Boron	-66.969	120.287	60.901	0.157
Aire / Boron	30.776	124.393	40.098	0.341

Discussion

Most of the monitored sites are simulated with an error that is less than a factor of 2, and all are within a factor of 10, except for one site in Went where a rather low measured LAS concentration is overestimated (Ackworth, figure 6.7 and table A1.2). However, for this case the absolute error is less than for many other sites, with an error within a factor of 2.

In general, the relative error for boron simulations is lower than for LAS in all of the catchments (table 6.12). The reason for this is probably that only one source of uncertainty exists for boron (release estimation), while for LAS two sources exist: release estimation and treatment removal.

Outliers Besides Ackworth, the most obvious outliers are the sites at Cridling Stubbs and Kirk Smeaton for LAS in the Went catchment (figure 6.7 and table A1.2). Both are simulated as trickling filter plants and overestimate the effluent concentration. Since these sites do not overestimate the boron concentration (though they are quite high), it is likely that the plants perform better than average trickling filter plants. One explanation could be that they own improved treatment facilities. However, these two plants serve less than 1000 people (with a short sewer system and low effluent flows), which might induce special flow and release patterns in terms of special days (washing days), for which the monitoring was not designed.

For the overestimation of LAS at Tupton and Holbrook in the Don / Rother catchment (figure 6.8 and table A1.3), similar circumstances may hold as for Cridling Stubbs and Kirk Smeaton.

Another outlier for LAS was the Esholt treatment plant in the Aire catchment (figure 6.9 and tables A1.4, A1.5). It was confirmed by the local authorities that this plant owns a tertiary treatment facility, so that it actually performs far better than usual trickling filter plants. This site was calibrated with a site-specific LAS elimination of 99.8%, which was derived from an influent / effluent monitoring comparison. Here, it is compared against the same effluent monitoring and hence fits quite well.

At the Dewsbury site in the Calder catchment (figure 6.10 and table A1.6) the measured mean concentration of boron was more than a factor of 2 higher than the simulated mean concentration. Dewsbury serves quite a large number of people, which means that the simulated mass flow for boron is even more significantly underestimated. LAS at Dewsbury is underestimated by a factor of almost 2. This points out the possibility of a wrongly estimated consumption for this site. Another explanation could be additional boron releases from sources other than surfactants. Dewsbury is a highly industrialized city, where glass or ceramics are manufactured and where other boron-releasing industries maybe placed.

Dewsbury is a city with high industrial activities where some might be glass and ceramics manufacturing and other boron releasing industries.

Tendencies A tendency of general over- or underestimation was not identified for LAS and boron.

Table 6.12 shows that LAS for the Went catchment is an exception, since it was overestimated (in terms of median concentration) for all sites ($D_a = D_{|a|}$). Also, the relative deviation is significantly higher than for all other scenarios. Since the relative deviation for Went / Boron is relatively low, the reason for the bad LAS modeling might be based on wrong treatment removal efficiencies.

Monitoring The selection of monitored sites did not consider the results of an initial simulation in order to identify locations of very high or very low predicted concentrations.

For example, in the Calder / Boron scenario (figure 6.10), the simulation shows

two sites with over 800 $\mu g/L$ that have not been included in the monitoring program. These sites significantly contribute to the main stream load and should be checked in order to avoid aftereffects for the subsequent in-stream simulation. For the LAS scenario in the Don/Rother none of the sites below 400 $\mu g/L$, which essentially are all activated sludge plants, have been measured (figure 6.8). Thus, this scenario lacks a basic check for sites with this type of treatment plant. As a third example, the Aire/Boron scenario (figure 6.13) shows that most monitored sites are from a group of sites with predicted concentrations of between 400 and 600 $\mu g/L$, although the predicted concentrations are almost equally distributed from 300 to 1000 $\mu g/L$.

These three examples show that initial simulations prior to monitoring campaigns should be used to identify sites that should be monitored.

The substance emission into the river network forms the link to the next model in sequence. Since the monitoring campaigns delivered only concentrations, the according mass flux could only be estimated using average effluent flows. An improved comparison of in-stream simulation with in-stream monitoring based on measured emission substance flux is not possible. Instead, the errors of the first models in the chain are carried forward to the next models and increase the overall uncertainty of the results.

For a separate evaluation of the river model, it is recommended to measure effluent flows accompanied to effluent concentrations.

In order to cover most of the catchment's mass flux it is recommended to select monitoring sites based on expected emission, rather than on release. For the Yorkshire situation, this basically means that trickling filter plants weigh more than activated sludge plants.

6.4.3 In-stream concentrations

Mean simulated and measured concentrations for in-stream sites are directly compared in figures 6.14 to 6.31. While the scattered plots include all monitoring sites, the profiles include only the main stream of the corresponding catchment. For boron, additional profiles with a constant background level, taken from a measurement upstream of any emission site (close to the source), are included. For simulation results in tabular form, see Appendix A.2 (page 157).

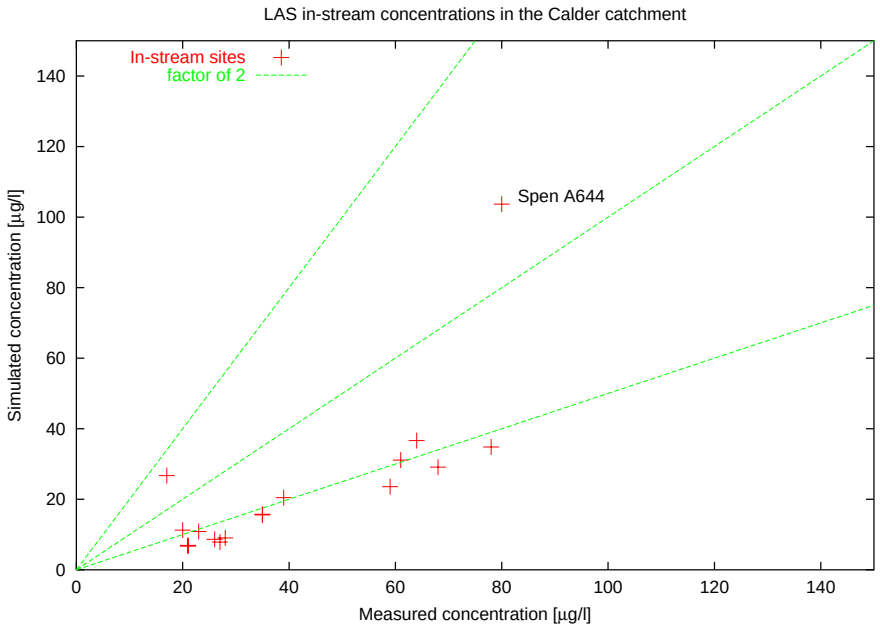


Figure 6.14: *LAS in the Calder: Mean in-stream concentration*

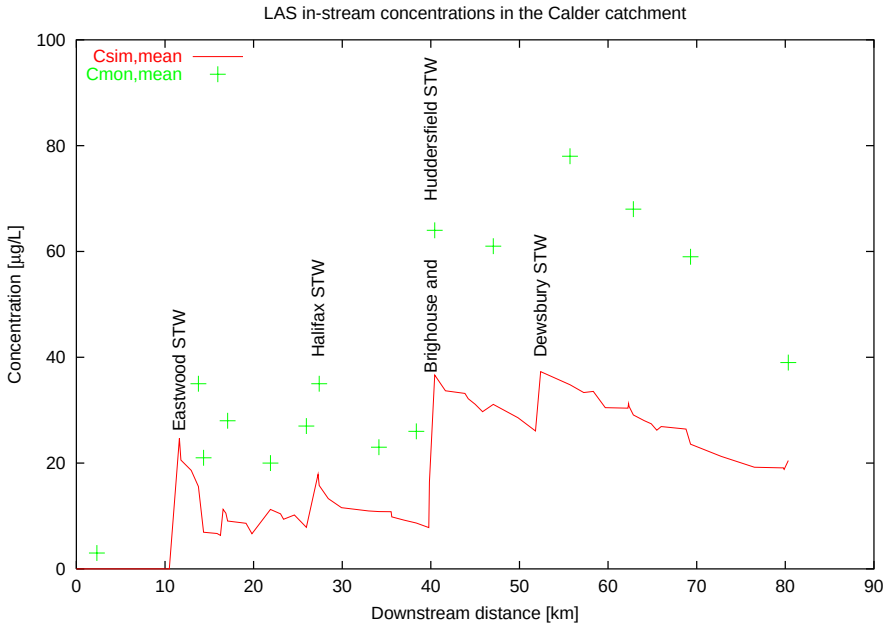


Figure 6.15: *LAS in the Calder: Mainstream profile*

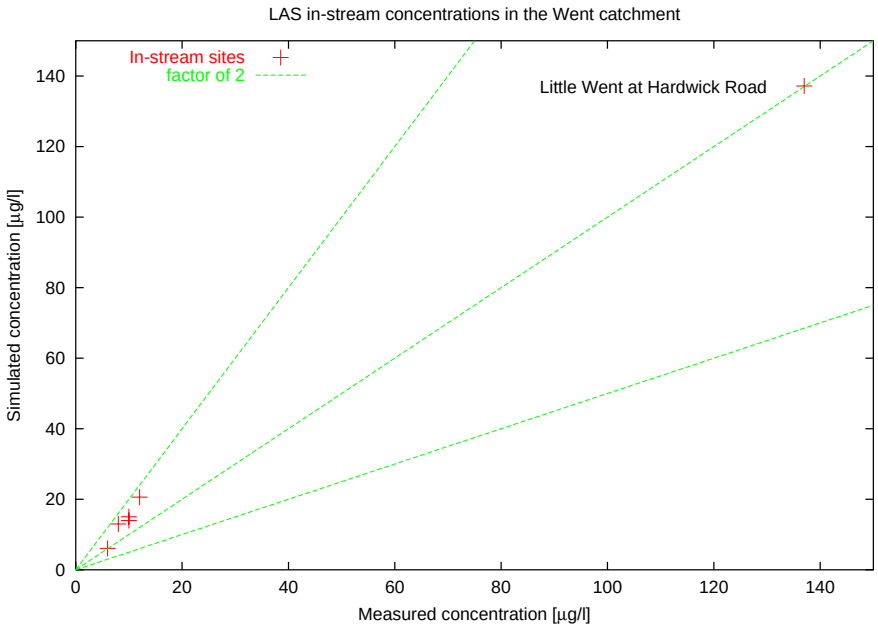


Figure 6.16: *LAS in the Went: Mean in-stream concentration*

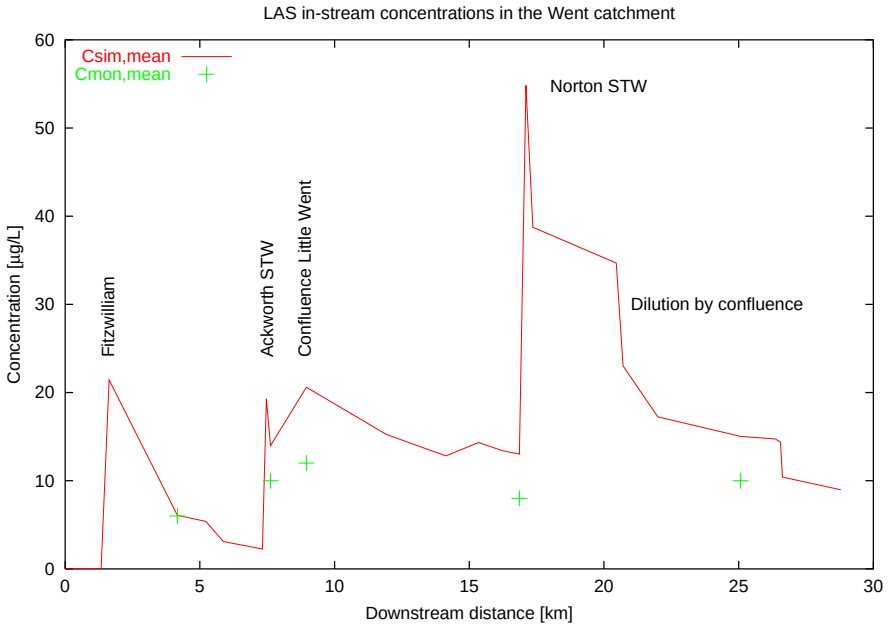


Figure 6.17: *LAS in the Went: Mainstream profile*

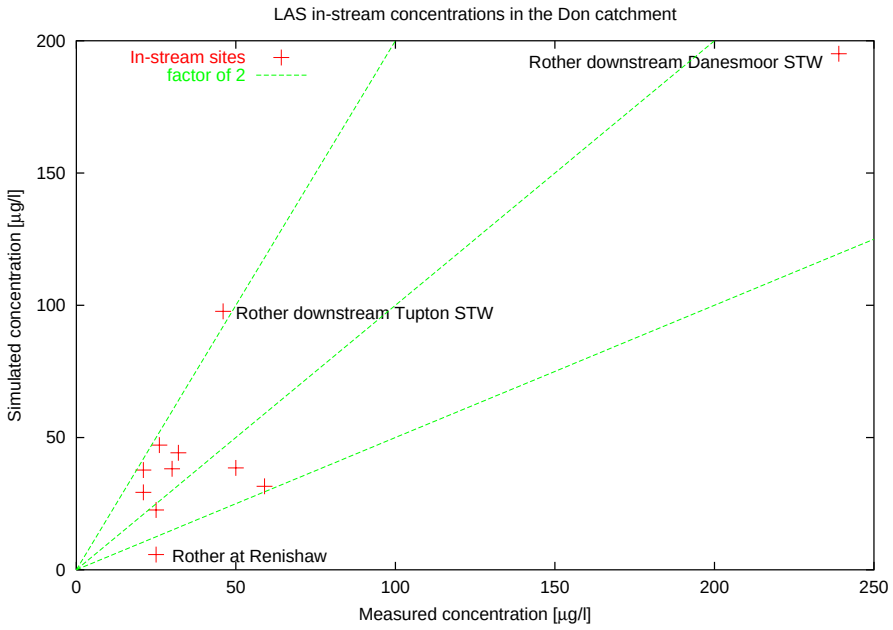


Figure 6.18: *LAS in the Don/Rother: Mean in-stream concentration*

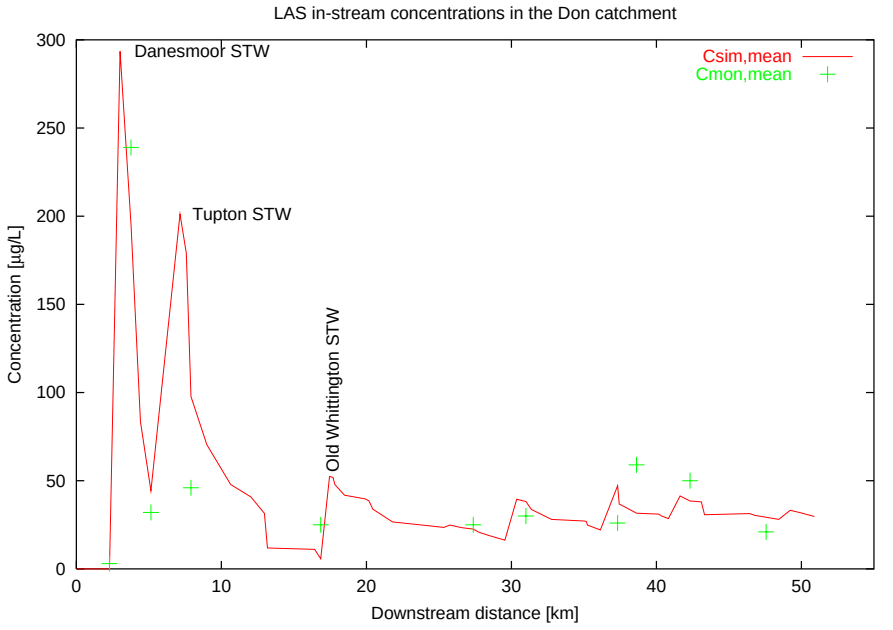


Figure 6.19: *LAS in the Don/Rother: Mainstream profile of River Rother*

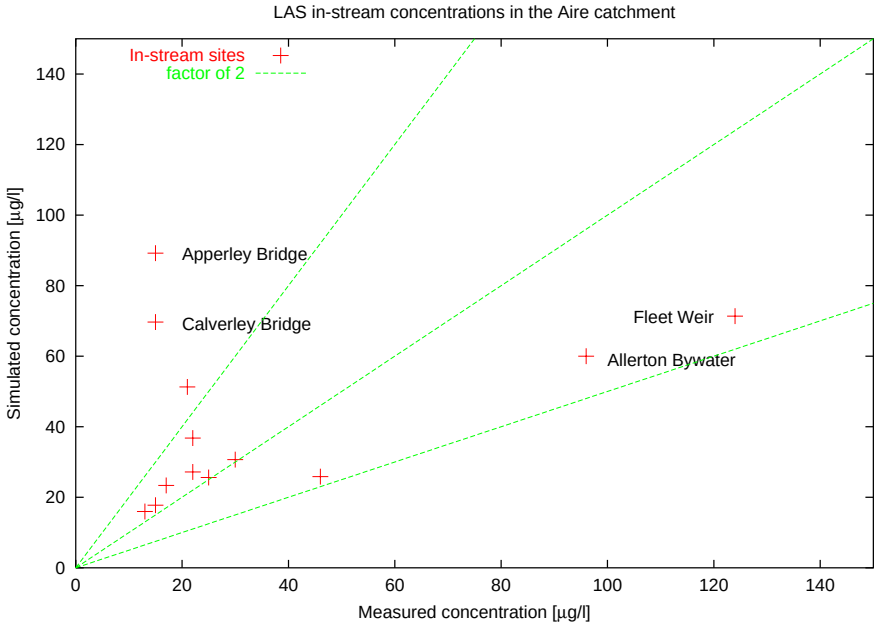


Figure 6.20: *LAS in the Aire: Mean in-stream concentration*

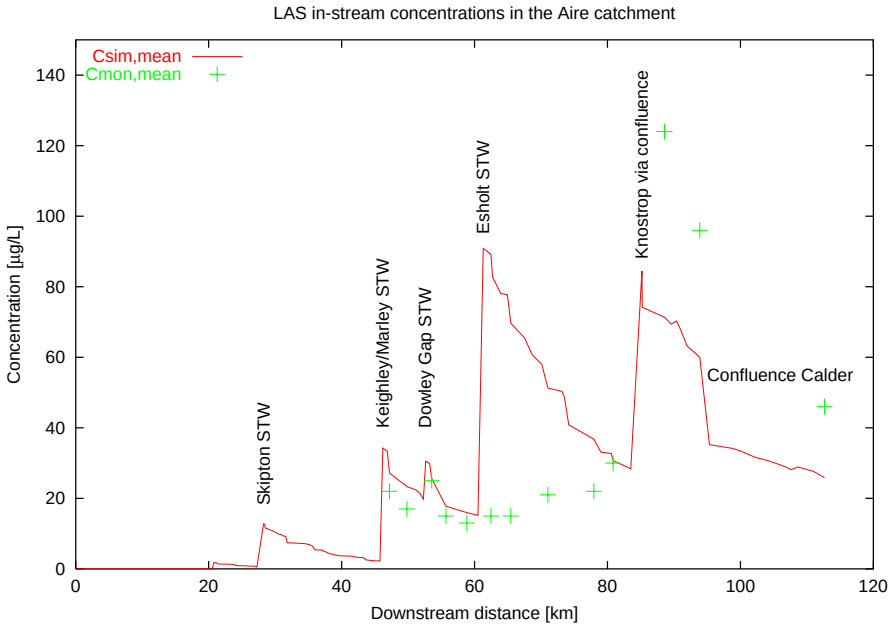


Figure 6.21: *LAS in the Aire: Mainstream profile*

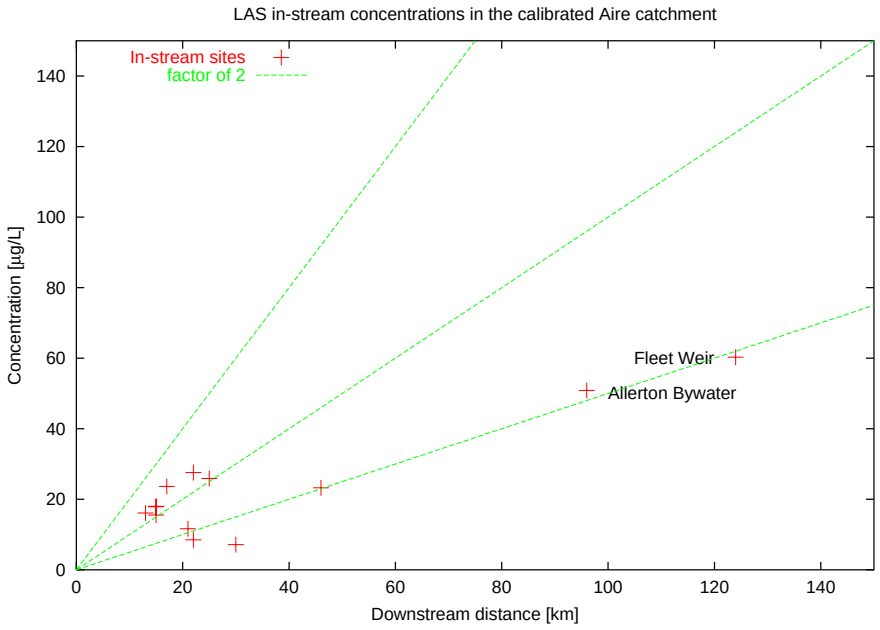


Figure 6.22: *LAS in the Aire (calibrated): Mean in-stream concentration*

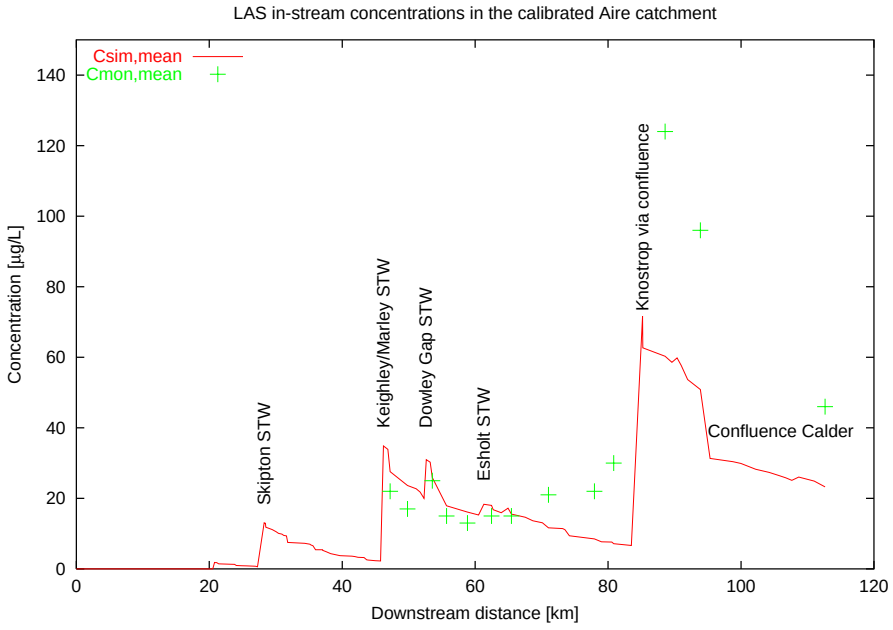


Figure 6.23: *LAS in the Aire (calibrated): Mainstream profile*

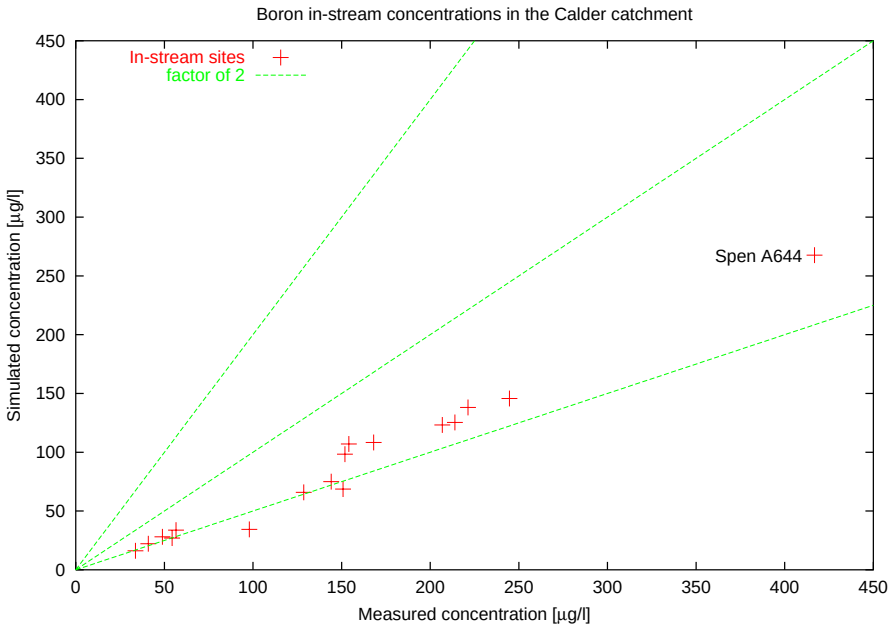


Figure 6.24: *Boron in the Calder: Mean in-stream concentration*

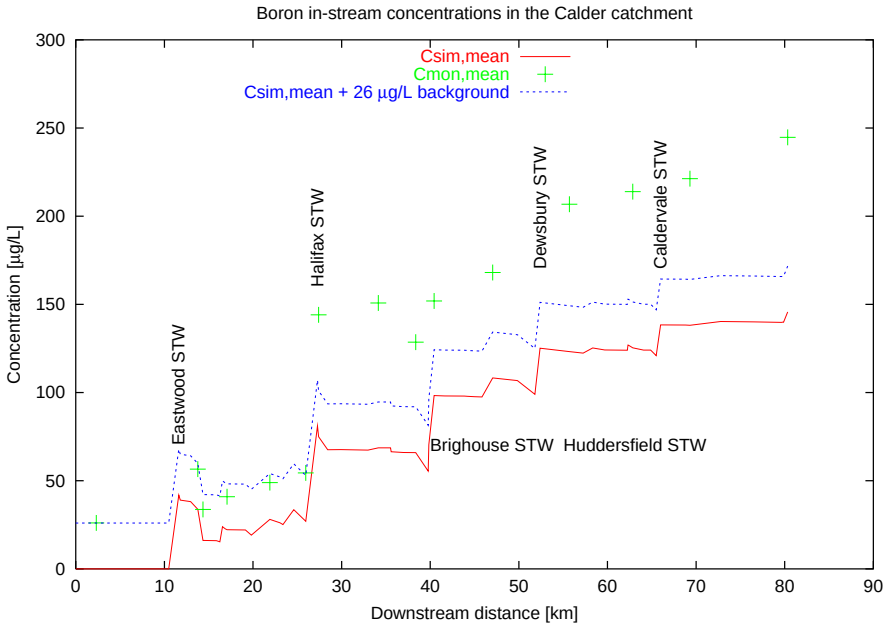


Figure 6.25: *Boron in the Calder: Mainstream profile*

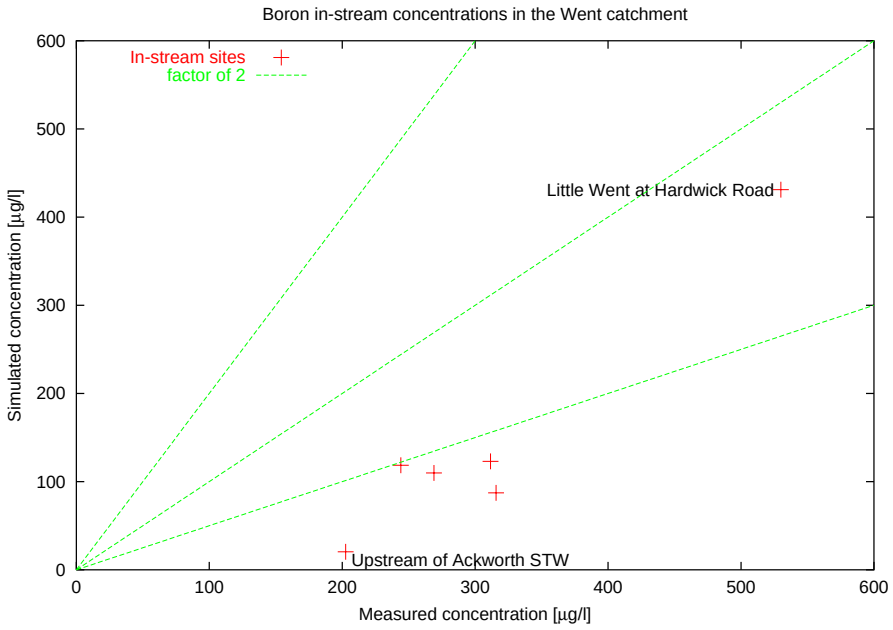


Figure 6.26: *Boron in the Went: Mean in-stream concentration*

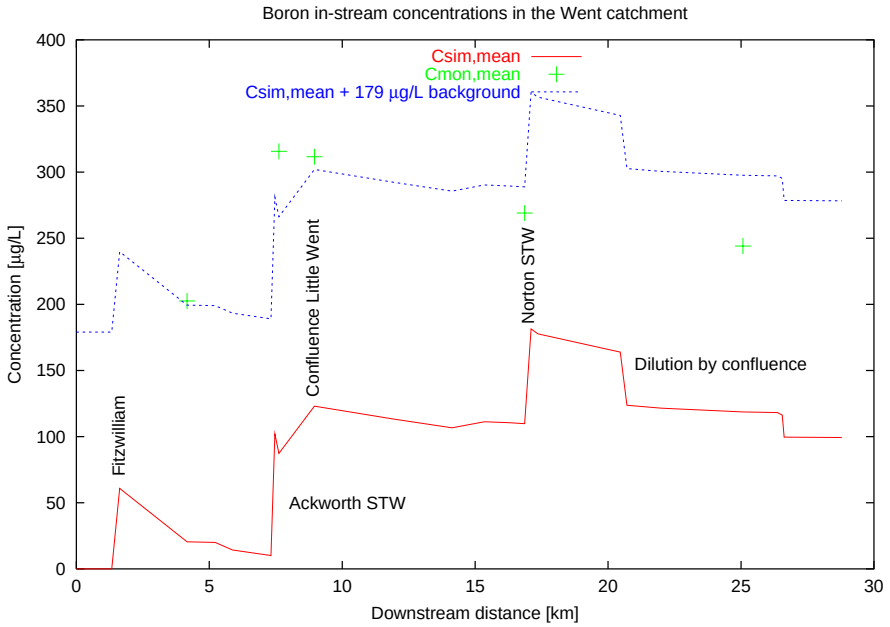


Figure 6.27: *Boron in the Went: Mainstream profile*

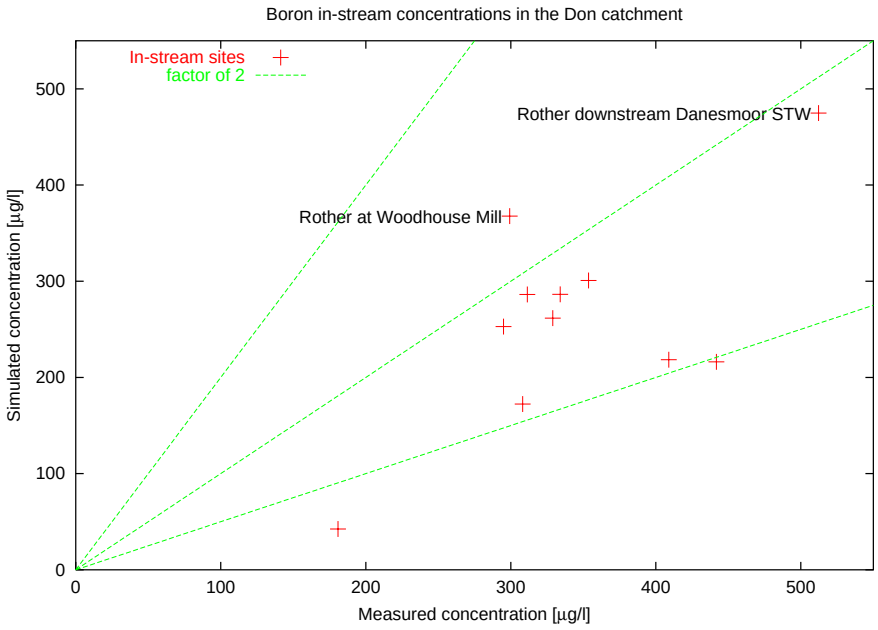


Figure 6.28: *Boron in the Don/Rother: Mean in-stream concentration*

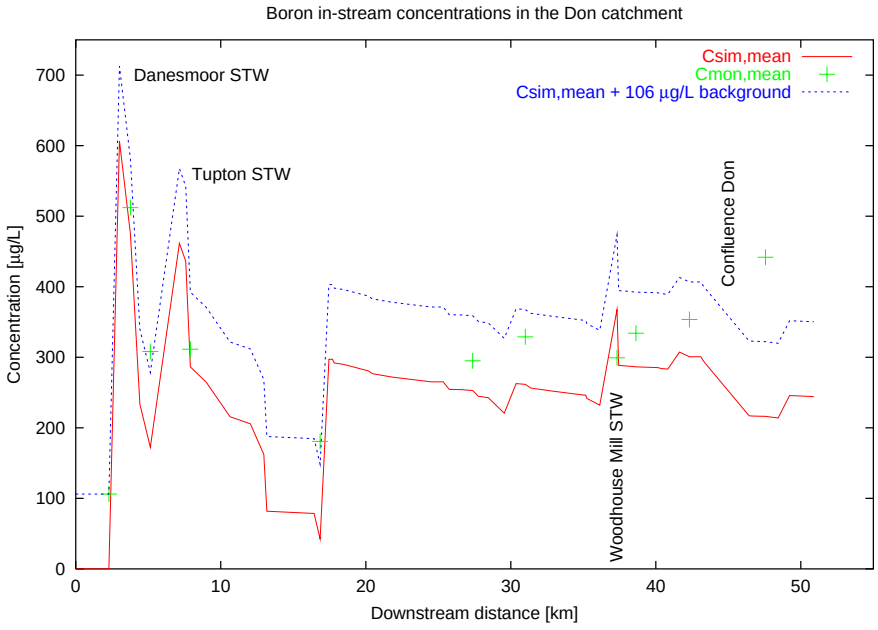


Figure 6.29: *Boron in the Don/Rother: Mainstream profile of River Rother*

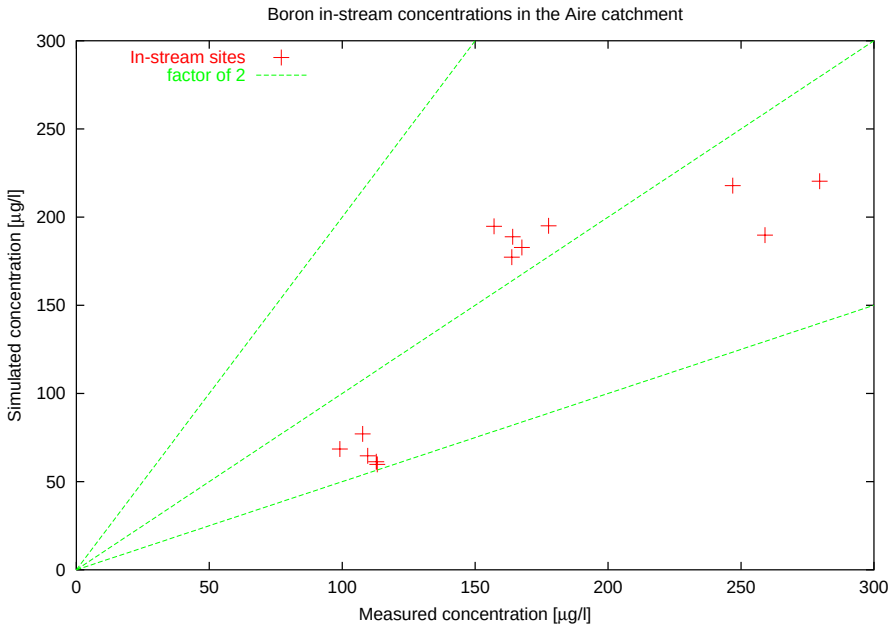


Figure 6.30: *Boron in the Aire: Mean in-stream concentration*

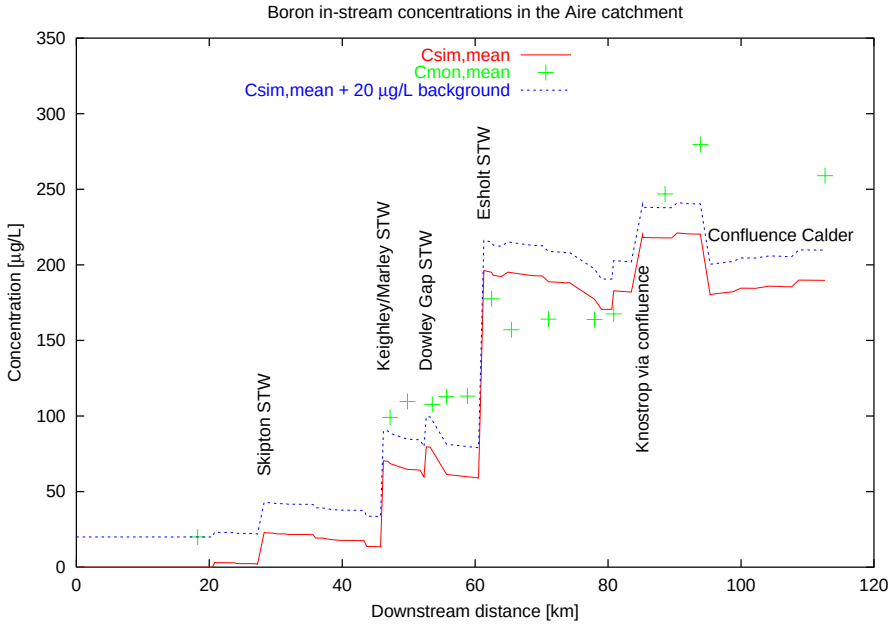


Figure 6.31: *Boron in the Aire: Mainstream profile*

Calculated measures for the model exactness based on the simulation results and available monitoring data are shown in table 6.13. This table summarizes the absolute measures D_a, $D_{|a|}$ and D_{rms} (equations 6.13, 6.14 and 6.15, page 104) and the relative measure $D_{|a|,r}$ (equations 6.16, page 104).

Table 6.13: Assessment of model exactness for median in-stream concentrations

	D_a	$D_{\|a\|}$	D_{rms}	$D_{\|a\|,r}$
Calder / LAS	-13.726	17.749	4.651	1.163
Went / LAS	2.764	7.401	3.134	0.975
Don / LAS	-4.649	18.361	6.855	0.577
Aire / LAS	-6.610	18.849	4.278	1.107
Aire / LAS calibrated	-12.801	15.890	3.879	0.937
Calder / Boron	-56.451	56.451	14.589	0.874
Went / Boron	-176.936	176.936	67.369	3.037
Don / Boron	-152.471	152.471	43.770	0.791
Aire / Boron	-46.727	46.884	9.379	0.704

Discussion

The in-stream concentrations do not cover the peak concentration, due to slow dilution, since the model assumes immediate full mixing. Hence, the sites for in-stream monitoring were placed sufficiently remote from points of emission.

The simulation results are within a factor of 2 for most sites and, with the exception of situations of unconsidered background levels, are always within the same order of magnitude.

The relative deviation of simulation results from measured values, as shown in table 6.13 ($D_{|a|,r}$), is quite similar for all scenarios, even for the two different substances (in contrast to final effluents, where boron clearly performs better than LAS, see table 6.12). One exception, however, is the Went / Boron scenario.

For boron, a catchment-dependent background concentration is detected through measurements upstream from any emission point (figures 6.25, 6.27, 6.29 and 6.31 and corresponding precise numbers in tables 6.3: Calder Portsmouth, 6.5: Aire Gargrave and Aire Carleton, 6.7: Rother US Danesmore STW). Concurrent low or almost zero LAS measurements (figures 6.15 and 6.19 and same tables as above) prove that the reason for the presence of boron is not

unconsidered detergent consumption. A geological background level appears to be an explanation.

Outliers At the confluence with the River Aire, the River Spen shows significantly higher concentrations for both substances, LAS and boron, than any other monitored in-stream site in the Calder catchment (figures 6.14 and 6.24). The higher concentration can be explained due to a relatively low dilution in comparison to a served population of almost 80,000.

A similar situation is present for the tributary Little Went, which carries the discharges of Carleton (figures 6.16 and 6.26).

Low diluted effluents are also observed for the upper part of the River Rother, downstream from Danesmoor and downstream from Tupton respectively (figures 6.19 and 6.29). The latter is overestimated by slightly more than a factor of 2. This is not regarded a serious problem since the high effluent volumes are almost not diluted when mixing with low in-stream flow volumes and thus create a higher uncertainty for measurements than for more downstream situations.

For the Aire/LAS scenario, two sites (Apperley and Calverley, figure 6.20) show an overestimation. Both sites are located downstream from Esholt STW, which required a site-specific calibration (figures 6.21 and 6.23). With the calibrated Esholt plant, Apperley and Calverley are no longer outliers (figure 6.22).

In comparison to all other in-stream sites for the Went/Boron scenario, the site upstream from Ackworth shows an extreme underestimation (figure 6.26). This fact is likely to be a consequence of an unconsidered background influence at an upstream location (absolute deviation d_a is almost the same as for Hardwick Beck in table A2.7).

Tendencies Boron is generally underestimated. The median-based assessments of the model's exactness show an almost prevalent underestimation ($|D_a| = D_{|a|}$, table 6.13). Also the direct comparison of mean concentrations shows an underestimation for most sites; only a few are slightly overestimated.

For the Went catchment, the underestimation of boron appears to be caused by an unusually high background concentration of almost 180 $\mu g/L$ (figure 6.27 and table 6.13). A possible additional source of boron comes from flooded coal

mines, which might have increased the natural contribution from the geological background (according to personal communication with local authorities). The high relative discrepancies $D_{|a|,r}$ (table 6.13) are partly due to the Went being a small catchment highly loaded with additional constant input.

Site calibration The site calibration performed for Esholt STW in the Aire catchment improved the simulation for the site itself and for some distance downstream. The effect of the calibration decreases rapidly for readily degradable substances such as LAS. The site calibration considered the removal efficiency and thus has no effect on boron.

While the profile plot improves dramatically in terms of visual impression (figure 6.21 compared to figure 6.23: simulated concentration profile approaches monitored values), $D_{|a|}$ and D_r improve only slightly (table 6.13). D_a is even worse than before, which is perhaps due to the more emphasized systematic error because of a lower influence by site-specific errors.

6.4.4 Regional PECs

The spatially refined and explicit exposure assessment delivers a high number of site-specific predicted concentrations. In the context of regional comparison and evaluation, an adequate aggregation into single values is considered helpful in many ways. Any approach will stress or neglect certain properties of the scenario under consideration. The actual focus and tolerable disadvantages of a method are ultimately a matter of political weighting. However, such methods must always exclude basic dependencies, such as the scale and resolution of the geographic data.

Two aggregation methods were developed within the GREAT-ER framework (Boeije et al., 2000):

- $PEC_{initial}$: unweighted aggregation of concentrations immediately downstream from wastewater emissions.

  The predicted concentrations at the start of the receiving river segments $C_{sim,start}$ are applied with the assumption of instantaneous full mixing. Hence, they present a kind of a worst-case scenario. It should be kept in mind that even higher concentrations are possible with incomplete mixing. No in-stream removal influences this PEC.

- $PEC_{catchment}$: weighted aggregation of all stretch concentrations. Weighting by flow increment is applied to eliminate the dependency on the river network's geographical detail (the inclusion of a high number of unloaded headwater stretches does not influence the $PEC_{catchment}$) and the dependency of inhomogeneous stretch lengths (weighting compensates the fact that the average C_{sim} decreases with increasing stretch length).

  The only considered unloaded stretches are those directly upstream of discharges. These stretches, by their flow increment (identical to their actual flow), consider all of the unpolluted headwaters for the regional weighting.

Table 6.14 summarizes $PEC_{initial}$ and $PEC_{catchment}$ for all scenarios including the uncalibrated Aire/LAS catchment as computed by the GREAT-ER 1.0.1 software following the equations by (Boeije et al., 2000).

Table 6.14: Regional PECs $[\mu g/L]$

	$PEC_{initial}$	$PEC_{catchment}$
Calder/LAS	76.797	24.312
Went/LAS	111.162	20.796
Don/LAS	97.818	33.279
Aire/LAS	95.089	30.443
Aire/LAS calibrated	91.549	23.545
Calder/Boron	170.501	102.639
Went/Boron	218.680	115.863
Don/Boron	329.257	261.193
Aire/Boron	225.349	126.368

Discussion The results in table 6.14 show the same order of magnitude across all four catchments for each combination of regional PEC type and substance. With the exception of the Don/Boron scenario, all $PEC_{catchment}$ values are within the very small range of 14 $\mu g/L$ for LAS and 24 $\mu g/L$ for boron (Table 6.14). The variations for PEC_{intial} are moderately higher. The reason for this low variance is almost definitely caused by the fact that all four regions are located close to each other, and therefore experience very similar discharges and consumption rates, and are affected by the same climatic conditions.

A higher $PEC_{initial}$ does not automatically lead to a higher $PEC_{catchment}$, as shown by the Went/LAS scenario.

Only the $PEC_{initial}$ of the Don/Boron scenario exceeds the corresponding substance PNEC of 300 $\mu g/L$.

6.4.5 Risk characterization

Risk within the EU Environmental Risk assessment scheme is the ratio between the Predicted Environmental Concentration (PEC) and the Predicted No-Effect Concentration (PNEC). With given PNEC values and the results of GREAT-ER simulations, this ratio can be calculated for all river stretches to distinguish sites potentially at risk from the safe sites. Either a simple true / false pattern or a scheme considering several fractions or factors of the PEC / PNEC ratio can be applied. Figures 6.32 to 6.39 show the regional site-specific risk characterization with three classes, in which the lowest class covers sites with a ratio of less than one tenth. The second class reaches a factor of one and the final class covers any sites which are potentially at risk.

For the PNEC of boron the rather low value of 300 $\mu g/L$ is applied. This value is clearly below the maximum allowed concentration of 985 $\mu g/L$ as used for the award of the Community Eco-label to Laundry Detergents (EU, 1999b). It is also below the current maximum value of 1000 $\mu g/L$ defined for the drinking water directive which is discussed to be lowered significantly to 300 $\mu g/L$ (Metzner et al., 1999).

For LAS a PNEC of 250 $\mu g/L$ is applied. This value is derived in A.I.S.E. / CESIO (1995). A slightly higher value of 300 $\mu g/L$ is used for the award of the Community Eco-label to Laundry Detergents (EU, 1999b).

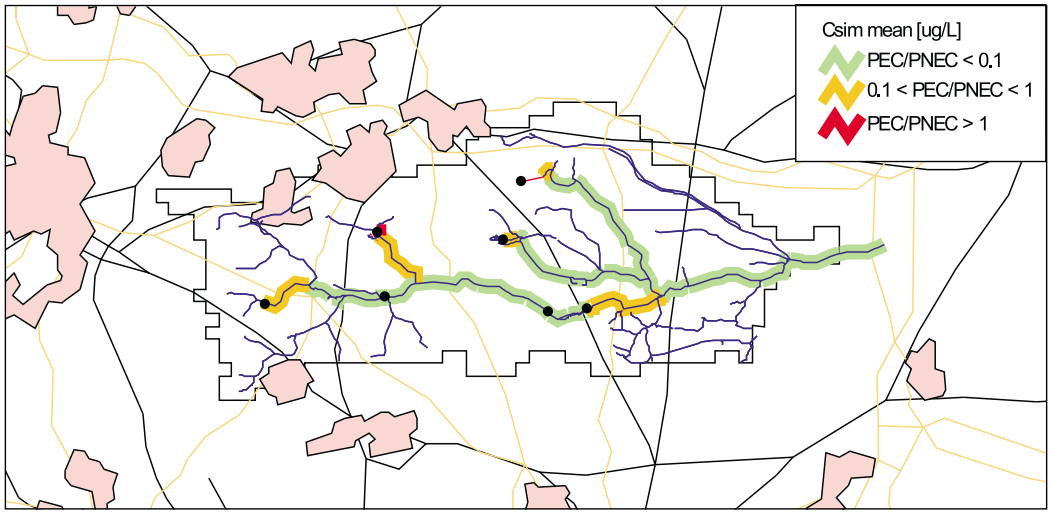


Figure 6.32: *Risk characterization for Went/LAS (PNEC=*$250\mu g/L$*)*

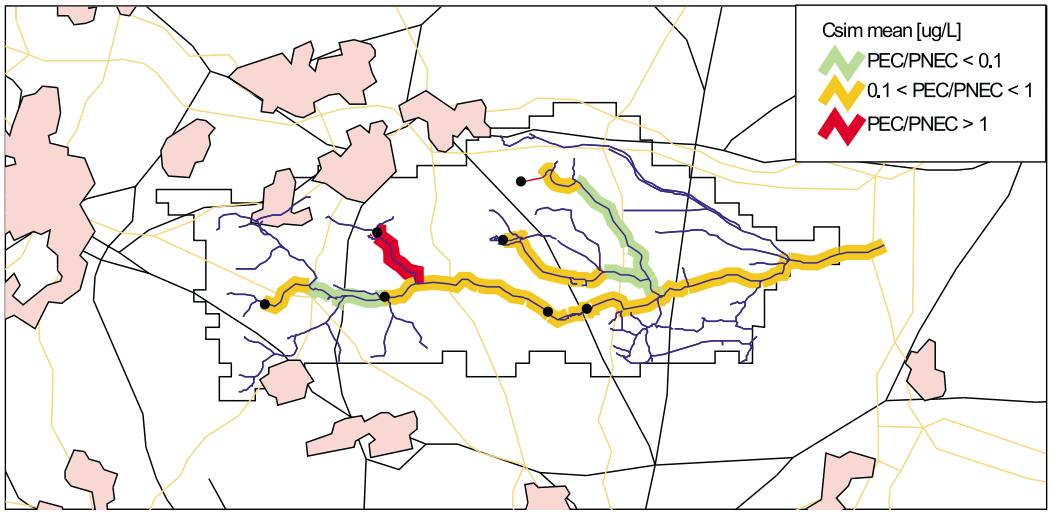


Figure 6.33: *Risk characterization for Went/Boron (PNEC=*$300\mu g/L$*)*

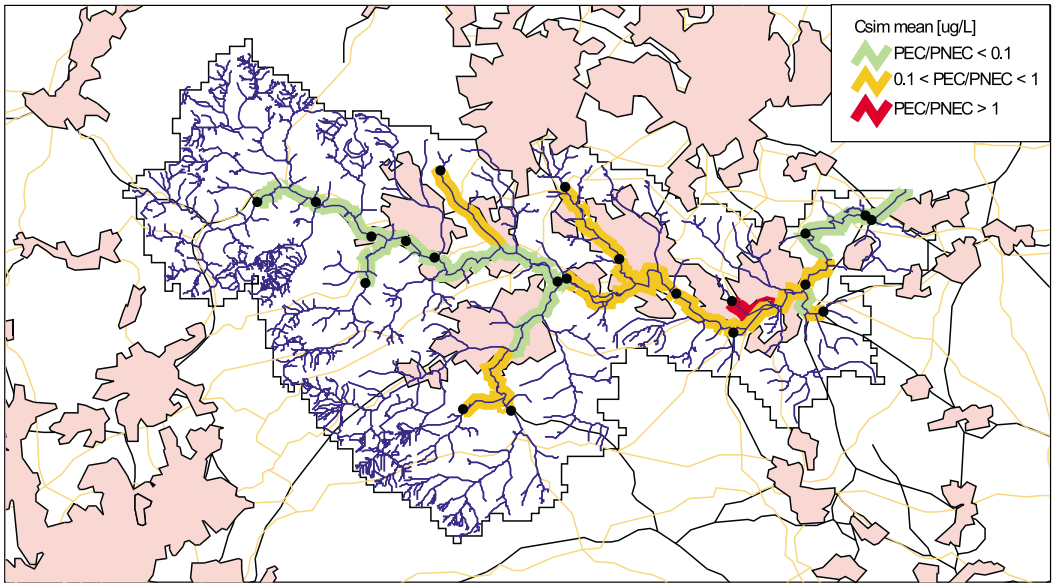


Figure 6.34: *Risk characterization for Calder/LAS (PNEC=$250\mu g/L$)*

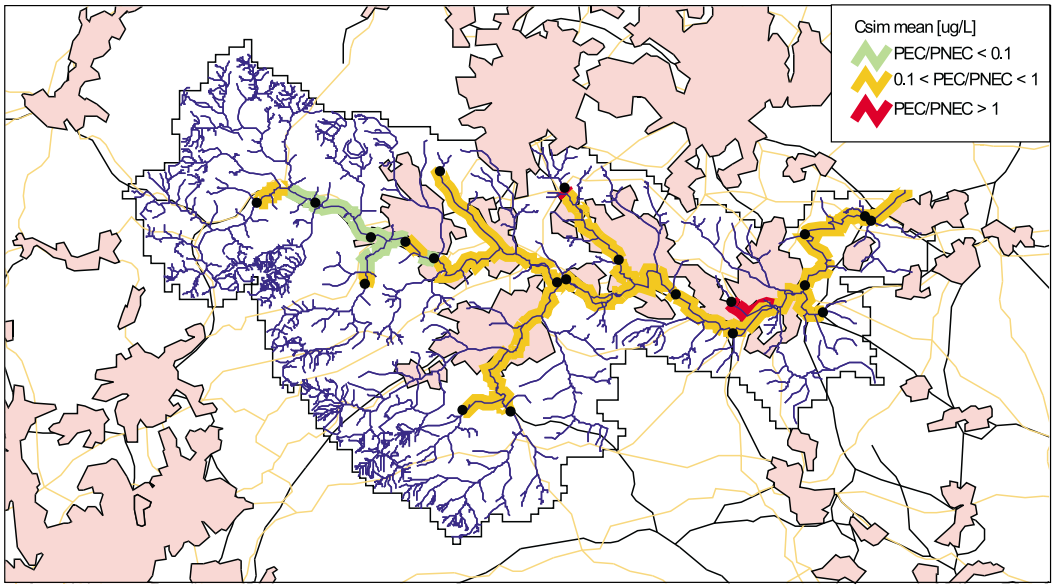


Figure 6.35: *Risk characterization for Calder/Boron (PNEC=$300\mu g/L$)*

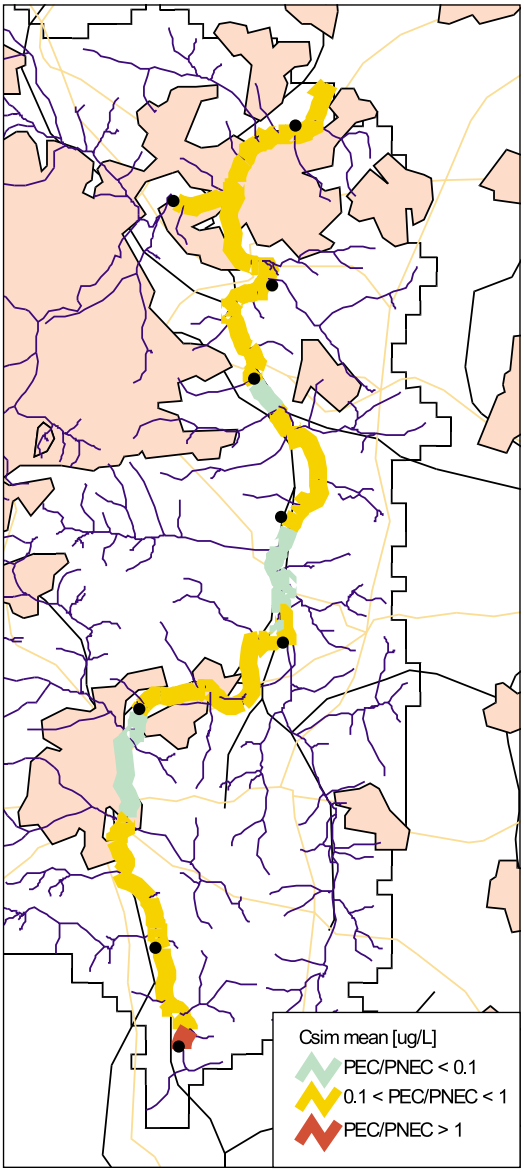


Figure 6.36: *Risk characterization for Rother/LAS (PNEC=250$\mu g/L$)*

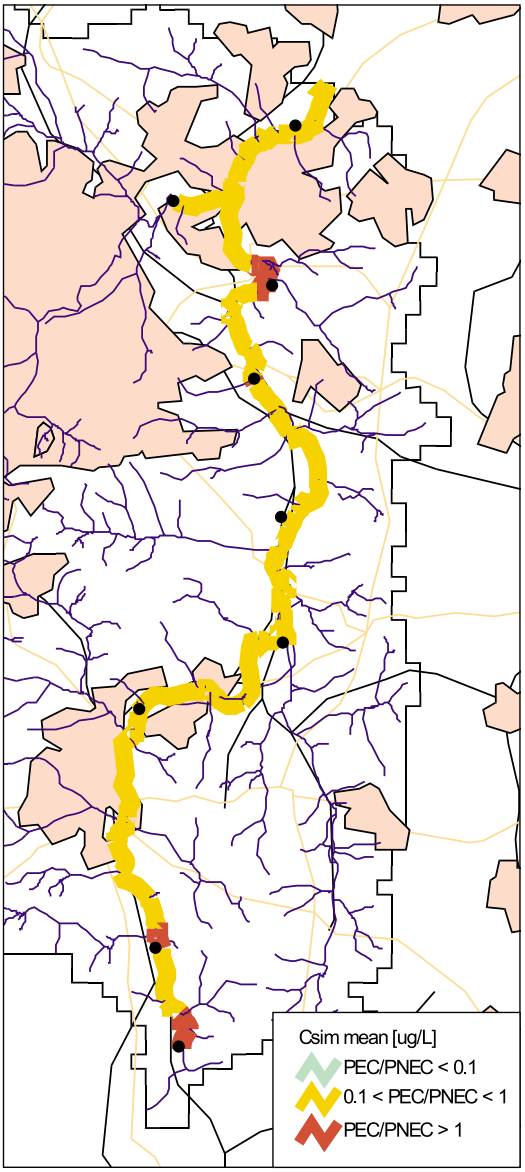


Figure 6.37: *Risk characterization for Rother/Boron (PNEC=300$\mu g/L$)*

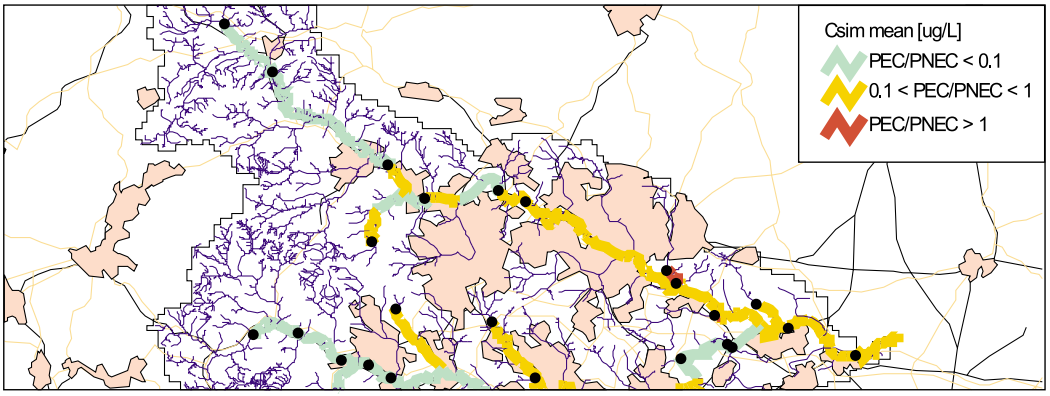


Figure 6.38: *Risk characterization for Aire/LAS (PNEC=*$250\mu g/L$*)*

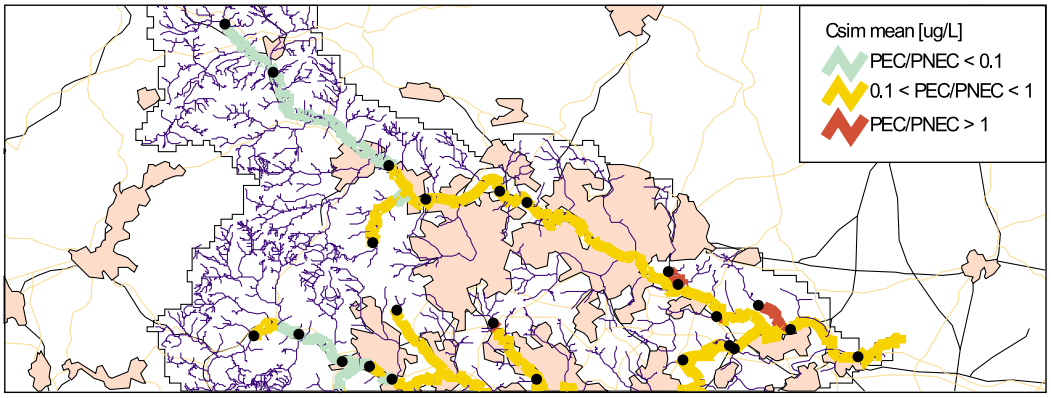


Figure 6.39: *Risk characterization for Aire/Boron (PNEC=*$300\mu g/L$*)*

Discussion The risk characterization with GREAT-ER provides the opportunity to explore site-specific risk estimation within a whole region. Nonetheless, being a spatially refined and spatially explicit method without aggregation of risk characterization it is not directly applicable to an official risk assessment scheme such as for the evaluation of new or existing substances within the EU.

For this purpose, the risk characterization feature needs to be further developed towards a well discussed and agreed methodology to be incorporated into the EU Environmental Risk Assessment scheme.

Chapter 7

Conclusions and outlook

7.1 Conclusions

The development and application of GREAT-ER has proven that geo-referenced exposure assessment is possible in terms of quality and also in terms of practicability.

Such a system enables the review and understanding of spatial aspects in the exposure behaviour of substances. Even some less precise input data will enable one to learn about the spatial aspects. This spatial view is not yet an integral part in discussions between industry and authorities, and is not even widespread in science. Thus a new awareness can be initiated. At a more practical level, decision support for data collection and screening is also provided.

The approach accommodates a higher level of complexity and uncertainty. This requires a careful treatment and understanding of combined uncertainties on the one hand. On the other hand, precise and well documented technologies must be issued to deal with the given complexity and to make the process transparent for anyone who wants to track down a certain behavior (provided he or she has sufficient experience in the corresponding topic). If this transparency is not given, the tool becomes largely useless for any task to which it was assigned.

7.1.1 Adequacy of the developed tool

Construction of a usable tool

GREAT-ER was developed to be a hybrid in terms of its type of usage. An approach was chosen to develop a prototype which can easily evolve into a professional tool, which at the same time is not limited to any scientific in-depth analysis. However, this decision only marginally influenced the usability aspects of the system (e.g. lack of visually perfect user interface).

The principle usability aspects for the task of combining exposure assessment and spatial reference have been addressed. The major issue that usability design has to cope with is the enormous volume of input and output data. This is done on behalf of issuing the capabilities of a GIS realizing a visual interactivity. The user works with dynamic maps retrieving information and partly communicating changes via the map representation (e.g. selection of sites for property editing). This type of interaction does not only provide comfortable handling of the tool, it also enables the user to better understand complex data and their interdependence.

GREAT-ER utilizes adequate technologies and demonstrates how the visual interaction can greatly improve usability. Nevertheless, many further opportunities of the technologies which would additionally enhance the usability of GREAT-ER (e.g. areal selection mode: drawing a box, circle or polygon to select spatial objects) have not been realised.

Another important aspect of usability is the incorporation of expert knowledge. When plugging several expert models together into one framework, there is a need to support users who do not have the same expertise as the system's developers. Good support to prevent wrong application or interpretation must be provided to keep the tool usable. GREAT-ER considered this by implementing parameter error ranges, parameter warning ranges, default values, parameter necessity indicators and parameter comments. These features support performance of an application and provide transparency and comprehensibility of scenarios as well. Further transparency could be gained by making the model itself and the incorporated processes and parameter interconnection visible in terms of dynamic graphical presentation. This has not been realized in GREAT-ER, but is a desirable feature.

Preparing input data

The preparation of purpose-adequate input data is one of the most problematic tasks for geo-referenced exposure assessment. The principle dual-step approach using an intermediate pre-defined state of the required data has provided good support for the data collation and processing. The first step from raw data to the intermediate state still requires manual work and control on the basis of profound topic-related know-how. The automated second step, which prepares ready-to-use data sets, offers anchorage in any plausibility checks. Furthermore, this step is repeatable and thus supports interactive quality improvement. A final quality control has been proven to be unavoidable. The raw data provided for GREAT-ER have been further improved with the introduced methodology: several errors the data owner was not aware of have been detected. However, this methodology still needs further user support in order to achieve a high quality for large numbers of different catchments.

Questions that can be answered with the tool

GREAT-ER addresses the fate of consumer chemicals along the waste water pathway. The following items GREAT-ER is able to deliver refer to this type of application:

- general impression on what is going on
- identification of hot-spots
- quantification and location
- comparison of substances for the same catchment
- comparison of catchments for the same substance
- risk characterization
- usable results even for a small number of point emissions
- selected ”what-if” scenarios

First of all, GREAT-ER provides a good impression of the substances’ fate within the catchment. Thus, understanding of the spatial aspects is supported

by answering also the question of what happens to a substance in a certain catchment.

Next, even with less precise basic data, hot-spots can be identified: the order of magnitude and also a sufficient spatial reference / extent are supplied for potential further investigation, e.g. monitoring.

With adequately refined and complete basic data, the tool offers quantification and location at a level of precision (errors less than a factor of 2 and location within a few hundred meters) that no other simulation method is able to provide.

GREAT-ER allows users to compare chemicals (e.g. substance alternatives for a product) for one catchment, as well as the comparison of a chemical's fate in different catchments. Besides the principle comparison of maximum concentrations, with the $PEC_{initial}$ and $PEC_{catchment}$ support is given to identify and judge differences and similarities.

With a given PNEC, a risk characterization for a substance / catchment combination can be performed. This characterization supports the refinement tier, as proposed by the corresponding EU Technical Guidance Documents.

Simulations with only a very few discharge sites still offer usable results at the same level of quality as for many discharge sites.

Finally, GREAT-ER allows to run a number of "what-if" scenarios: improvement of an STP, increasing / decreasing population, changes in substance consumption.

Questions that can not be answered

Before applying GREAT-ER, one should carefully check whether the aspired answers can be delivered with the tool in its current state. The following can not be answered adequately:

- 2- or 3-dimensional analysis, e.g. transversal in-stream gradient
- analysis close to emission points
- chemical metabolites

- chemical mixtures
- general ”what-if” scenarios
- transferability of results

Many questions will go further - for some GREAT-ER can be adapted, but for others a different tool might be more appropriate.

The underlying model naturally limits the diversity of questions. The model is one-dimensional which means that no analysis distinguishing the left from the right bank can be made. Also, only the water body is considered for the pathway and hence no explicit results for, e.g. fate in sediments, is offered. The assumption of immediate full mixing prohibits a detailed analysis close to emission points. Furthermore, questions regarding mixtures or the fate of metabolites can not be answered with the presently incorporated model.

Besides these model-dependent items, the current data interface is not extended to allow ”what-if” scenarios regarding new plants or different flow scenarios (summer / winter). Each of these ”what-if” scenarios would just require a new execution of the pre-processing. Since GREAT-ER does not incorporate a hydrological flow estimation model, it does not make any sense to allow the insertion of new STPs via the user interface, because flow regimes would be affected and inconsistent situations would be created. No problems arise for explicit flow scenarios.

Above all, it must be kept in mind that the tool can identify situations of high substance concentration, but these are potentially not the only hot-spots in the given catchment (due to incomplete basic data).

Finally, the transferability of the results from one scenario to other areas / catchments is very problematic, because too many catchment-specific parameters are used.

Interpreting simulation results

The quality of the underlying data primarily determines the opportunities and limititations of interpretation. No quantifying methods are available for the actual determination; interpretation needs to be done on the basis of sound scientific argumentation considering all information available. The consultation of background information, such as water quality maps, is quite helpful.

Besides, any interpretation must bear in mind that GREAT-ER computes statistical results, which means that probabilities are given, rather than an explicit prediction for a point in time.

Improving the quality of simulation results

Increasing the number of Monte-Carlo samples is a very limited way to improve the quality of the results. It should only be used as a method to decrease the uncertainty contributed by the stochastic approach below the level of other uncertainties. This also means that there is no need to replace the Monte-Carlo method by an alternative since other factors dominate the uncertainty and the computational effort is negligable anyway.

A dominant role for the quality of simulation results is played by the adequacy of the applied consumption and removal scenario. Compared to this, a further refinement of spatial data has a less efficient cost-use benefit for the overall quality (provided the level of refinement is sufficient for an adequate output precision).

Thus a better review of the basic data in terms of additional plausibility checks and consideration of local experts will contribute to the quality of simulation results, at least by increasing trust in them.

Another option to gain a higher quality is adjustment at certain keypoints (e.g. a discharge's removal efficiency) based on monitoring results.

7.1.2 Further opportunities for the tool

Though it was not explicitly designed for this purpose, GREAT-ER has shown the potential to become a vital support tool for planning monitoring studies. Initial scenario simulations imply where situations need to be verified (or falsified) via measurements. Also broad monitoring campaigns with a limited budget can cover a regional situation on the basis of initial GREAT-ER simulations more efficiently.

7.2 Outlook

GREAT-ER is a fully developed and usable tool for analysing the ”down-the-drain” path of consumer chemicals. It is a prototype for dealing with spatial aspects in exposure modeling in general.

Some aspects need to be refined and some new ideas are generated which need to be evaluated for possible integration.

7.2.1 Internet feasibility

For computer programs and data a central point of maintenance with web access enables users to work with most current software releases and data. On the technical level, two ways to take advantage of this opportunity exist: On the one hand, users can download a computer program for local installation. On the other hand, the Internet allows interaction with a central point (which even includes a remote use of software). The first option is mostly feasible for small- and medium-sized software packages with a large user group. The second option is becoming more important for very large packages and a relatively low number of users, and is also typical for many databases.

Bringing a complex GIS-based application with a voluminous database onto the Web would, though with a limited user interface and user interaction, complement the full GREAT-ER system that requires local installation. The advantages are:

1. The applications are readily available without complicated installation routines. The local system remains unchanged and no software conflicts, disk capacity excess (large databases!), etc. have to be dealt with. Since the local installation of a GIS is not required, no time is wasted, no maintenance skills are required and, of course, no costs for the software package are incurred.

2. Users do not depend on the operating system and / or the hardware platform on which the actual GIS and database are running. It is sufficient to use a browser providing the chosen data access and communication.

3. The centralized architecture means that maintenance of the system is only needed at one point to keep users up-to-date. Besides the reduc-

tion of redundance, this also reduces costs, because the server installation only needs to be developed and maintained for one operating system and hardware platform.

4. Hosting at one place by a widely accepted competent authority will also increase the general acceptance of and trust in the system and provided data.

5. Many data sets are often owned by third parties. Many geographic data, especially in vector format (geographic objects plus attributes), are considerably valuable. Distributing these data via the Web will cause serious licensing and legal problems. On the other hand, the Web also offers a comfortable method to handle data accessability and unauthorized reuse of geographic data: the server installation holds the geographic objects and only sends out images of the requested maps. Data providers will feel more comfortable with this technology and hence can be more easily convinced to provide their data.

7.2.2 LCIA

The application of GREAT-ER 1.0 within a Life Cycle (Impact) Assessment, LC(I)A, can be done in different ways.

The most straight-forward way calculates fate factors for appropriate, i.e. wide-dispersive used ”down-the-drain”, chemicals based on reference emissions that have taken place in reference regions, and which reflect at least in which way wastewater is treated. The mean over all river stretches of normalized concentration increases and can then be used as both substance- and catchment-specific fate factors, i.e. the fate factors can be different for different countries or regions. The key substance parameters are the chemical’s elimination behavior in wastewater treatment and in the surface water, while the main important geographic parameters are the way in which wastewater is treated, as well as the hydrological data, which determines the dilution ratios in the different stretches. The flow velocity is less important.

Alternatively, in cases in which the functional unit can be assigned to a time span, an approach as proposed by (Schulze, 2000) can be followed. There, the functional unit is defined as the use of a product system per time unit. Transforming this on an annual basis allows for the direct use of the emitted mass as an annual load. The mean concentration increase or the percentage of stretches exceeding a defined reference concentration, e.g. a no-effect concentration, are

possible fate or characterization factors, which are then also functions of the emitted mass. Schulze (2000) proposes a critical length as the sum across all river stretches of a catchment and all substances emitted due to the functional unit of mean concentration increases weighted by the length of the stretch and divided by a no-effect concentration. This approach deviates from the framework of equation 1 and is therefore clearly located in a later stage of an LCA: If emissions of ”down-the-drain” chemicals have been recognized as important, such an assessment can be done to gain further information that may be useful to many decisions. The GREAT-ER graphical user interface (GUI) was extended to allow automated simulations of product ingredients based on the above-mentioned procedure.

Appendix A

Simulation results

All tables share the same columns:

- Name: Name of monitored site.
- $\overline{C_{\text{mon}}}$: Mean monitored concentration.
- $\overline{C_{\text{sim}}}$: Mean simulated concentration.
- d_a: = $\overline{C_{\text{sim}}} - \overline{C_{\text{mon}}}$ (absolut deviation).
- d_r: = $|d_a|/min(\overline{C_{\text{mon}}}, \overline{C_{\text{sim}}})$ (relative deviation)

A.1 Final effluents

A.1.1 LAS

Table A1.1: **Calder** - LAS - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Ossett Spa	358	373	15	0.042
Huddersfield	108	219	111	1.035
Brighouse	601	268	-332	1.235
North Bierley	203	285	82	0.409
Caldervale	22	87	65	2.973
Shibden Head	488	356	-131	0.370
Halifax	69	77	8	0.129
Neiley	344	277	-66	0.240
Milner Royd	11	106	95	8.722
High Royd	96	233	137	1.436
Redacre	514	220	-293	1.336
Eastwood	204	125	-78	0.623
Mitchell Laithes	539	280	-258	0.920

See also figure 6.6, page 116.

Table A1.2: **Went** - LAS - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Fitzwilliam	322	305	-16	0.054
Ackworth	6	122	116	19.462
Carleton	413	636	223	0.542
Kirk Smeaton	153	970	817	5.345
Askern Norton	199	382	183	0.923
Cridling Stubbs	197	760	563	2.862
Stapleton Park	196	352	156	0.797

See also figure 6.7, page 116.

Table A1.3: **Don/Rother** - LAS - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Holbrook	218	578	360	1.653
Woodhouse Mill	n.d.	151	-	-
Long Lane	577	562	-14	0.026
Danesmoor	697	535	-161	0.302
Tupton	101	443	342	3.396
Old Whittington	n.d.	117	-	-

See also figure 6.8, page 117.

Table A1.4: **Aire** - LAS - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Gargrave	223	400	177	0.794
Snaygill	218	333	115	0.529
Marley	183	296	113	0.619
Dowley Gap	504	326	-177	0.545
Esholt	14	308	294	21.064
Knostrop	354	329	-24	0.073
Wheldale	173	341	168	0.976
Sutton	361	480	119	0.331
Eastwood	204	125	-78	0.622
Redacre	514	220	-293	1.334
High Royd	96	234	138	1.438
Milner Royd	11	106	95	8.682
Halifax	69	77	8	0.124
Shibden Head	488	356	-131	0.369
Brighouse	601	269	-331	1.233
Neiley	344	277	-66	0.239
Huddersfield	108	219	111	1.037
North Bierley	203	286	83	0.410
Mitchell Laithes	539	280	-258	0.919
Ossett Spa	358	373	15	0.043
Caldervale	22	87	65	2.956

See also figure 6.9, page 117.

Table A1.5: **Aire calibrated** - LAS - Final Effluents

Name	$\overline{C}_{\text{mon}}$	$\overline{C}_{\text{sim}}$	d_a	d_r
Gargrave	223	401	178	0.799
Snaygill	218	334	116	0.533
Marley	183	297	114	0.624
Dowley Gap	504	327	-176	0.541
Esholt	14	15	1	0.101
Knostrop	354	330	-23	0.070
Wheldale	173	342	169	0.981
Sutton	361	481	120	0.334
Eastwood	204	126	-77	0.617
Redacre	514	220	-293	1.328
High Royd	96	234	138	1.444
Milner Royd	11	106	95	8.669
Halifax	69	77	8	0.123
Shibden Head	488	357	-130	0.366
Brighouse	601	269	-331	1.227
Neiley	344	278	-65	0.235
Huddersfield	108	220	112	1.042
North Bierley	203	286	83	0.413
Mitchell Laithes	539	281	-257	0.914
Ossett Spa	358	374	16	0.046
Caldervale	22	86	64	2.951

For Esholt see figure 6.9, page 117.

A.1.2 Boron

Table A1.6: **Calder** - Boron - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Ossett Spa	691	621	-70	0.113
Huddersfield	393	365	-27	0.074
Brighouse	416	447	30	0.074
North Bierley	484	476	-7	0.017
Caldervale	682	465	-216	0.465
Shibden Head	458	592	133	0.292
Halifax	598	415	-183	0.442
Neiley	413	461	48	0.117
Milner Royd	262	570	307	1.174
High Royd	239	389	150	0.628
Redacre	276	366	89	0.324
Eastwood	218	209	-9	0.044
Mitchell Laithes	1116	467	-649	1.389

See also figure 6.10, page 118.

Table A1.7: **Went** - Boron - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Fitzwilliam	891	508	-382	0.752
Ackworth	785	654	-130	0.200
Carleton	747	1060	312	0.419
Kirk Smeaton	1092	1617	524	0.480
Askern Norton	720	637	-82	0.130
Cridling Stubbs	1156	1267	110	0.096
Stapleton Park	795	586	-208	0.355

See also figure 6.11, page 118.

Table A1.8: **Don/Rother** - Boron - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Holbrook	831	963	131	0.158
Woodhouse Mill	987	807	-179	0.222
Long Lane	987	937	-50	0.054
Danesmoor	986	891	-94	0.106
Tupton	1057	739	-317	0.430
Old Whittington	704	623	-80	0.130

See also figure 6.12, page 119.

Table A1.9: **Aire** - Boron - Final Effluents

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Gargrave	455	665	210	0.462
Snaygill	450	554	104	0.231
Marley	504	492	-11	0.023
Dowley Gap	427	542	115	0.271
Esholt	459	513	54	0.118
Knostrop	495	548	53	0.108
Wheldale	592	568	-23	0.042
Sutton	837	799	-37	0.047
Eastwood	218	209	-9	0.044
Redacre	276	366	89	0.324
High Royd	239	389	150	0.628
Milner Royd	262	570	307	1.174
Halifax	598	415	-183	0.442
Shibden Head	458	592	133	0.292
Brighouse	416	447	30	0.074
Neiley	413	461	48	0.117
Huddersfield	393	365	-27	0.074
North Bierley	484	476	-7	0.017
Mitchell Laithes	1116	467	-649	1.389
Ossett Spa	691	621	-70	0.113
Caldervale	682	465	-216	0.465

See also figure 6.13, page 119.

A.2 In-stream

A.2.1 LAS

Table A2.1: **Calder** - LAS - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Calder Methley Br.	39	20	-18	0.904
Calder Portsmouth	3	0	-3	-
Calder Heptonstall	35	15	-19	1.246
Calder Hebden Br.	21	6	-14	2.034
Calder Brearley Weir	28	9	-18	2.097
Calder Sowerby Br.	20	11	-8	0.779
Calder Copley Br.	27	7	-19	2.441
Calder North Dean	35	15	-19	1.222
Red Beck at Brookfoot	17	26	9	0.571
Calder Rastrick Br.	23	10	-12	1.121
Calder Cooper Br.	26	8	-17	2.003
Colne at Colne Br.	21	6	-14	2.130
Calder Battyeford	64	36	-27	0.747
Spen A644	80	103	23	0.296
Calder Dewsbury	61	31	-29	0.962
Calder Horbury Br.	78	34	-43	1.241
Calder Kirkgate	68	29	-38	1.337
Calder Stanley Ferry	59	23	-35	1.503
Calder Methley Br.	39	20	-18	0.904

See also figures 6.14 and 6.15, page 123.

Table A2.2: **Went** - LAS - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Went US Hardwick Beck	31	0	-31	-
Went at Ackworth	6	6	0	0.011
Went at Standing Flats Br.	10	13	3	0.397
Little Went at Hardwick Road	137	137	0	0.001
Went at Went Br.	12	20	8	0.716
Went at Walden Stubbs	8	13	5	0.626
Went at Sykehouse	10	15	5	0.503

See also figures 6.16 and 6.17, page 124.

Table A2.3: **Don/Rother** - LAS - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Rother US Tupton	32	44	12	0.383
Rother DS Tupton	46	97	51	1.124
Drone DS Dronfield Bypass	86	0	-86	-
Rother at Newbridge Lane	25	5	-19	3.355
Rother at Cow Lane	14	0	-14	-
Doe Lea at Renishaw	21	37	16	0.796
Rother at Renishaw	25	22	-2	0.106
Rother at Holbrook	30	38	8	0.274
Rother at Woodhouse Mill	26	47	21	0.814
Rother DS Woodhouse Mill	59	31	-27	0.869
Rother at Canklow	50	38	-11	0.298
Don at BSC Rotherham Gate 14	15	0	-15	-
Don at Rawmarsh Road - Rotherham	21	29	8	0.395
Rother US Danesmoor	3	0	-3	-
Rother DS Danesmoor	239	195	-43	0.225

See also figures 6.18 and 6.19, page 125.

Table A2.4: **Aire** - LAS - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Aire Gargrave	-1	0	1	-
Aire Carleton	-1	0	1	-
Aire Crossflatts	22	27	5	0.236
Aire Cottingley Br.	17	23	6	0.374
Aire Salts Weir	25	25	0	0.025
Aire Buck Br.	15	17	2	0.183
Aire above Esholt	13	15	2	0.229
Aire Apperley Br.	15	89	74	4.948
Aire Calverley Br.	15	69	54	3.645
Aire Kirkstall Br.	21	51	30	1.442
Aire Leeds Br.	22	36	14	0.672
Aire US Thwaite Mill	30	30	0	0.023
Aire Fleet Weir	124	71	-52	0.738
Aire Allerton Bywater	96	59	-36	0.600
Aire Beal Br.	46	25	-20	0.780
Calder Portsmouth	3	0	-3	-
Calder Heptonstall	35	15	-19	1.256
Calder Hebden Br.	21	6	-14	2.046
Calder Brearley Weir	28	9	-18	2.108
Calder Sowerby Br.	20	11	-8	0.785
Calder Copley Br.	27	7	-19	2.453
Calder North Dean	35	15	-19	1.237
Red Beck at Brookfoot	17	26	9	0.568
Calder Rastrick Br.	23	10	-12	1.133
Calder Cooper Br.	26	8	-17	2.020
Colne at Colne Br.	21	6	-14	2.139
Calder Battyeford	64	36	-27	0.752
Spen A644	80	103	23	0.294
Calder Dewsbury	61	31	-29	0.967
Calder Horbury Br.	78	34	-43	1.246
Calder Kirkgate	68	29	-38	1.342
Calder Stanley Ferry	59	23	-35	1.509
Calder Methley Br.	39	20	-18	0.908

See also figures 6.20 and 6.21, page 126.

Table A2.5: **Aire (calibrated)** - LAS - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Aire Gargrave	-1	0	1	-
Aire Carleton	-1	0	1	-
Aire Crossflatts	22	27	5	0.253
Aire Cottingley Br.	17	23	6	0.392
Aire Salts Weir	25	25	0	0.036
Aire Buck Br.	15	17	2	0.192
Aire above Esholt	13	16	3	0.238
Aire Apperley Br.	15	18	3	0.200
Aire Calverley Br.	15	15	0	0.036
Aire Kirkstall Br.	21	11	-9	0.803
Aire Leeds Br.	22	8	-13	1.589
Aire US Thwaite Mill	30	7	-22	3.201
Aire Fleet Weir	124	60	-63	1.057
Aire Allerton Bywater	96	50	-45	0.887
Aire Beal Br.	46	23	-22	0.979
Calder Portsmouth	3	0	-3	-
Calder Heptonstall	35	15	-19	1.226
Calder Hebden Br.	21	6	-14	2.018
Calder Brearley Weir	28	9	-18	2.078
Calder Sowerby Br.	20	11	-8	0.768
Calder Copley Br.	27	7	-19	2.431
Calder North Dean	35	15	-19	1.221
Red Beck at Brookfoot	17	26	9	0.576
Calder Rastrick Br.	23	10	-12	1.123
Calder Cooper Br.	26	8	-17	2.008
Colne at Colne Br.	21	6	-14	2.122
Calder Battyeford	64	36	-27	0.738
Spen A644	80	104	24	0.301
Calder Dewsbury	61	31	-29	0.954
Calder Horbury Br.	78	34	-43	1.233
Calder Kirkgate	68	29	-38	1.330
Calder Stanley Ferry	59	23	-35	1.499
Calder Methley Br.	39	20	-18	0.901

See also figures 6.22 and 6.23, page 127.

A.2.2 Boron

Table A2.6: **Calder** - Boron - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Calder Methley Br.	244	145	-98	0.679
Calder Portsmouth	26	0	-26	-
Calder Heptonstall	56	33	-22	0.675
Calder Hebden Br.	33	16	-17	1.088
Calder Brearley Weir	40	22	-18	0.840
Calder Sowerby Br.	48	28	-20	0.740
Calder Copley Br.	54	27	-27	1.012
Calder North Dean	144	74	-69	0.922
Red Beck at Brookfoot	154	107	-47	0.439
Calder Rastrick Br.	150	68	-82	1.197
Calder Cooper Br.	128	65	-62	0.951
Colne at Colne Br.	97	34	-63	1.850
Calder Battyeford	151	98	-53	0.545
Spen A644	416	267	-149	0.557
Calder Dewsbury	168	108	-59	0.552
Calder Horbury Br.	206	123	-83	0.679
Calder Kirkgate	213	125	-88	0.706
Calder Stanley Ferry	221	138	-83	0.602
Calder Methley Br.	244	145	-98	0.679

See also figures 6.24 and 6.25, page 128.

Table A2.7: **Went** - Boron - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Went US Hardwick Beck	179	0	-179	-
Went at Ackworth	202	20	-182	8.903
Went at Standing Flats Br.	315	87	-228	2.615
Little Went at Hardwick Road	530	431	-98	0.229
Went at Went Br.	311	123	-188	1.533
Went at Walden Stubbs	269	109	-159	1.449
Went at Sykehouse	244	118	-125	1.058

See also figures 6.26 and 6.27, page 129.

Table A2.8: **Don/Rother** - Boron - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Rother US Tupton	308	172	-135	0.788
Rother DS Tupton	311	286	-25	0.088
Drone DS Dronfield Bypass	266	0	-266	-
Rother at Newbridge Lane	180	42	-138	3.264
Rother at Cow Lane	317	0	-317	-
Doe Lea at Renishaw	408	218	-190	0.872
Rother at Renishaw	295	252	-42	0.166
Rother at Holbrook	328	261	-67	0.257
Rother at Woodhouse Mill	299	367	68	0.229
Rother DS Woodhouse Mill	334	286	-47	0.167
Rother at Canklow	353	300	-52	0.176
Don at BSC Rotherham Gate 14	287	0	-287	-
Don at Rawmarsh Road - Rotherham	441	216	-225	1.044
Rother US Danesmoor	106	0	-106	-
Rother DS Danesmoor	512	474	-37	0.079

See also figures 6.28 and 6.29, page 130.

Table A2.9: **Aire** - Boron - In-stream

Name	$\overline{C_{\text{mon}}}$	$\overline{C_{\text{sim}}}$	d_a	d_r
Aire Gargrave	20	0	-20	-
Aire Carleton	30	0	-30	-
Aire Crossflatts	99	68	-30	0.446
Aire Cottingley Br.	109	64	-44	0.695
Aire Salts Weir	107	77	-30	0.397
Aire Buck Br.	112	61	-51	0.840
Aire above Esholt	113	59	-53	0.893
Aire Apperley Br.	177	195	17	0.098
Aire Calverley Br.	157	194	37	0.240
Aire Kirkstall Br.	164	188	24	0.151
Aire Leeds Br.	163	177	13	0.082
Aire US Thwaite Mill	167	182	15	0.091
Aire Fleet Weir	246	217	-29	0.133
Aire Allerton Bywater	279	220	-59	0.269
Aire Beal Br.	259	189	-69	0.365
Calder Portsmouth	26	0	-26	-
Calder Heptonstall	56	33	-22	0.667
Calder Hebden Br.	33	16	-17	1.078
Calder Brearley Weir	40	22	-18	0.831
Calder Sowerby Br.	48	28	-20	0.732
Calder Copley Br.	54	27	-27	1.003
Calder North Dean	144	75	-68	0.914
Red Beck at Brookfoot	154	107	-46	0.433
Calder Rastrick Br.	150	68	-81	1.187
Calder Cooper Br.	128	66	-62	0.942
Colne at Colne Br.	97	34	-63	1.836
Calder Battyeford	151	98	-53	0.539
Spen A644	416	268	-148	0.552
Calder Dewsbury	168	108	-59	0.546
Calder Horbury Br.	206	123	-83	0.672
Calder Kirkgate	213	125	-88	0.699
Calder Stanley Ferry	221	138	-82	0.596
Calder Methley Br.	244	146	-98	0.672

See also figures 6.30 and 6.31, page **??**.

Appendix B

Errors detected in GREAT-ER 1.0.1

During the detailed application and analysis of GREAT-ER 1.0.1, some errors in the data were identified and corrected for the present analysis:

- Went / FE-Monitoring: Carelton wrongly connected to StretchID 8882 - should be 25775.
- Went / FE-Monitoring: Criddling Stubbs wrongly connected to StretchID 25804 - should be 7107.
- Went / FE-Monitoring: Kirk Smeaton wrongly connected to StretchID 26709 - should be 26773.
- Don / FE-Monitoring: Long Lane wrongly connected to StretchID 26573 - should be 8278.
- Don / FE-Monitoring: Holbrook wrongly connected to StretchID 26782 - should be 126800.

References

A.I.S.E. / CESIO (1995). Environmental Risk Assessment of Detergent Chemicals. Proceedings of the A.I.S.E. / CESIO Limelette III Workshop on 28-29 November 1995.

Baumgarten, G., Reiter, B., Scheil, S., Schwartz, S., Wagner, J.O. (1998). CemoS User's Manual. In: (Trapp & Matthies, 1998).

Berding V., Schwartz, S., Matthies, M. (2000). Scenario analysis of a Level III multimedia model using generic and regional data. Environmental Science and Pollution Research 7: 147-158.

Boeije, G., Schowanek, D. & Vanrolleghem, P. (1998). Adaption of the Simple-Treat chemical fate model to single-sludge biological nutrient removal waste water treatment plants. Water Science and Technology 38(1), 211-218.

Boeije, G. (1999). Chemical Fate Prediction for Use in Geo-Referenced Environmental Exposure Assessment. PhD thesis, University of Ghent, Belgium.

Boeije, G., Wagner, J.O., Koormann, F., Vanrolleghem, P.A., Schowanek, D.R., Feijtel, T.C.J. (2000). New PEC definitions for river basins applicable to GIS-based environmental exposure assessment. Chemosphere 40(2000), 255-265.

Chapra, Steven C. (1997). Surface Water-Quality Modeling. McGraw-Hill, New York.

de Oude, N.-T. (1992). Detergents. The Handbook of Environmental Chemistry, Volume 3, Part F. Editor: O. Hutzinger, Springer Verlag, Berlin.

Deutscher Kanu-Verband (1985). Deutsches Flußwanderbuch, 22. Auflage, DKV Wirtschafts- und Verlags GmbH, Duisburg.

Dyck, S. & Peschke, G. (1989). Grundlagen der Hydrologie. Verlag für Bauwesen, Berlin.

ECETOC (1994). Technical Report No. 61 - Environmental Exposure Assessment.

ECETOC (1999). Special Report No. 16 - GREAT-ER User Manual.

EU (1980). Council Directive on the Quality of Water for Human Consumption 80/778/EEC. Official Journal of the European Communities No. L229/1, 30.8.1980.

EU (1995). Recommendations for a Council Directive on the Quality of Water for Human Consumption. Official Journal of the European Communities No. C131, 4.1.1995.

EU (1996). Technical Guidance Documents in Support of The Commission Directive 93/67/EEC on Risk Assessment For New Notified Substances and the Commission Regulation (EC) 1488/94 on Risk Assessment For Existing Substances. ECB, Ispra, Italy.

EU (1999). Common Position (EC) No 41/1999, Official Journal of the European Communities 1999/C 353/01.

EU (1999b). Commission decision of June 1999: Establishing the Ecological Criteria for the award of the Community Eco-label to Laundry Detergents, Official Journal of the European Communities 1999/476/EC.

Fedra, K. (1996). Distributed Models and Embedded GIS: Strategies and Case Studies of Integration. In: Goodchild, M.F., Steyart, L.Y., Parks, B.O., Johnston, C., Maidment, D., Crane, M. and Glendinning, S. [Publ.], GIS and Environmental Modeling: Progress and Research Issues. S. 413-417. GIS World Books, Fort Collins, CO, USA.

Feijtel, T.C.J., Boeije, G., Matthies, M., Young, A., Morris, G., Gandolfi, C., Hansen, B., Fox, K., Holt, M., Koch, V., Schroeder, R., Cassani, G., Schowanek, D., Rosenblom, J. & Niessen, H. (1997). Development of a Geography-referenced Regional Exposure Assessment Tool for European Rivers - GREAT-ER. Chemosphere, 34(11), 2351-2374.

Fox, K. (1998). Personal communication with Dr. Kay Fox, Unilever in October 1998.

Gayler, S. (1998). Bewertung computergestützter Umwelt-Simulationsmodelle. PhD thesis, University of Osnabrück, Germany.

Gustard, A., Bullock, A. & Dixon, J.M. (1992). Low flow estimation in the United Kingdom. Institute of Hydrology report 108, Wallingford.

Haberer, K. (1996). Bor und die Trinkwasserversorgung in Deutschland, gwf - Wasser / Abwasser Vol.137, No. 7, pp 364-370.

Han, S., Evans, R.G., Hodges, T. & Rawlins, S.L. (1995). Linking a GIS with a Potato Simulation Model for Site-specific Crop Management. J. Environ. Qual. 24, pp 772-777.

Holt, M.S., Fox, K.K., Burford, M., Daniel, M. & Buckland, H. (1998). UK Monitoring Study on the Removal of Linear Alkylbenzene Sulphonate in Trickling Filter Type Sewage Treatment Plants. Contribution to GREAT-ER Project # 2. Sci. Tot. Environ. 210 / 211, 255-269.

Jager, D.T., Visser, C.J.M. & Van de Meent, D. (1994). Uniform system for the evaluation of substances IV. Distribution and intake. Chemosphere 29 (2), 353-369.

Klepper, O., den Hollander, H.A. (1999) A comparison of spatially explicit and box models for the fate of chemicals in water, air and soil in Europe. Ecological Modelling, 116 (1999) 183-202.

Koorman, F. (1997). Organische Schadstoffe im Fließgewässernetz Rhein Modellierung, Simulation, Visualisierung von chemischen Zwischenprodukten. Diplomarbeit, University of Osnabrück.

Koormann, F., Matthies, M., Trapp, S. (1998). Simulation and Visualisation of Spatial Exposure Patterns: Intermediates in the Rhine River System, In: Regulation for Chemical Safety in Europe: Analysis, Comment and Criticism (Eds. D.M. Pugh, J.V. Tarazona), Kluwer Academic Publishers, Dordrecht, pp. 65 - 96.

Leopold, L.B. & Maddock, T. (1953). The Hydraulic Geometry Channels and Some Physiographic Implications. Geographical Survey Professional Paper 252, Washington DC, USA.

Loague, K. & Green, R.E. (1991). Statistical and graphical methods for evaluating solute transport models: Overview ans application. J. Contam. Hydrol, 7 pp. 51-73.

Louvar, J.F. & Louvar, B.D. (1998). Health and environmental risk assessment: fundamentals with applications. Prentice Hall, Upper Saddle River, NJ 07458.

Luckyanov, N.K. (1993). On adequacy of ecological models. Ecological Modelling, 69, pp. 51-65.

Mackay, D. (1979). Finding fugacity feasible. Environ. Sci. Technol. 13, 1218-1223.

Mackay, D. (1991). Multimedia Environmental Models: The Fugacity Approach. Lewis, Chelsea, MI, USA.

Mayer, D.G. & Butler, D.G. (1993). Statistical validation. Ecological Modelling, 68 pp. 21-32.

Metzner, G., Lind, G. & Nitschke, L. (1999). Survey of boron levels in aquatic environments in Germany. Tenside Surf. Det. 36 (1999) 6 pp. 364-378.

Miller, S.N., D.P. Guertin and D.C. Goodrich (1996). Linking GIS and Geomorphologic Field Research at Walnut Gulch Experimental Watershed. Proceedings of the AWRA's 32nd Annual Conference and Symposium: "GIS and Water Resources", Sept. 22-26, 1996, Fort Lauderdale, FL. pp. 327-335.

Miller, S.N., M. Hernandez and L.J. Lane (1997). GIS applications in the spatial extrapolation of hydrologic data from experimental watersheds. Proceedings of the Conference on Management of Landscapes Disturbed by Channel Incision, Oxford, MS, USA, May 20-22, 1997.

Morgan, M.G. and Henrion, M. (1990). Uncertainty. Cambridge University Press, Cambridge, UK.

Römpp Lexikon Umwelt (1997). Version 1.0, Georg Thieme Verlag, Stuttgart / New York.

Round, C.E., Young, A.R. & Fox, K. (1998). A Regionally Applicable Model for Estimating Flow Velocity at Ungauged River Sites in the UK - Contribution to GREAT-ER #4. Water and Environmental Management 12, pp. 402-405.

Schulze, C. (1998). Modellierung und Visualisierung der Konzentration von Waschmittelinhaltsstoffen in der Itter. Diplomarbeit, University of Osnabrück.

Schulze, C., Matthies, M., Trapp, S., Schröder, F.R. (1999). Georeferenced Fate Modelling of LAS in the Itter Stream. Chemosphere, 39(11), 1833-1852.

Schulze, C. (2000). Modelling and evaluating the aquatic fate of detergents. PhD thesis, University of Osnabrück, Germany.

Simons, D. B. & Albertson, M. L. (1960). Uniform water conveyance channels in alluvial material. Proc. Am. Soc. Civ. Engg., J. Hydraulics Div., 86, pp.33ff.

Struijs, J. (1996). Simple Treat 3.0: a model to predict the distribution and elimination of chemicals by sewage treatment plants. Report 719101025. National Institute of Public Health and Environmental Protection (RIVM), Bilthoven, The Netherlands.

Tim, U.S., Jolly R. (1994). Evaluating Agricultural Non-Point-Source Pollution Using Integrated Geographic Information Systems and Hydrologic / Water Quality Models. J. Environ. Qual. 23:25-35.

Trapp, S, Matthies, M. (1998). Chemodynamics and Environmental Modeling. Springer Berlin.

Van de Meent, D. (1993). SIMPLEBOX: A generic multimedia fate evaluation model. Report 672720001. National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

Van de Plassche, E.J., De Bruijn, J.H.M., Stephenson, R.R., Marshall, S.J., Feijtel, T.C.J. & Belanger, S.E. (1999). Predicted no-effect concentrations and risk characterization of four surfactants: linea alkyl benzene sulfonate, alcohol ethoxylates, alcohol ethoxylated sulfates and soap. Environmental Toxicology Chemistry, Vol. 18 No. 11, pp. 2653-2663.

Wagner, J.-O. (1996). Modelldaten-Austausch-Server für die Interaktion eines Geographischen Informationssystems mit Umweltsimulationsmodellen. Diplomarbeit, University of Osnabrück.

Assertion

Hiermit erkläre ich, diese Doktorarbeit selbständig verfasst und keine als die angegebenene Quellen und Hilfsmittel verwendet zu haben.

(I hereby declare that I have written this Doctoral thesis by myself, and that Ihave only used the given sources and aids.)

Osnabrück, June 17th, 2001,

(Jan-Oliver Wagner)