Exploratory Data Analysis in Water Quality Monitoring Systems

MSc Thesis

Ruurd Maasdam

Dissertation submitted in part fulfilment of requirements for the Degree of Master of Science in Geographical Information Systems

University of Salford

May 2000

Abstract

This thesis reviews some of the conceptual and technological issues associated with changes in water management and water monitoring. In the past many monitoring programs have been characterised as "data-rich but information poor", because the existing data were not used effective and efficient. This thesis treats this shortcoming of the past as the challenge for the future, and additionally strengthen the spatial content of the data.

This thesis advocates the concept of explorative analysis, with a special emphasis in those developments that make it possible to explore the data in a visual, spatial, dynamic and interactive way. These technologies allow the user to gain insight and stimulate the generation of ideas ("private visual thinking"). In this new concept maps are an tool of analysis and not just the end-product (e.g. presentation) of the analytical process.

A case study with Regional Water System Report illustrates that this new concept and tools are valuable in current monitoring programs. In this case ArcView and Cartographic Data Visualiser (CDV) are used as tools to allow the user to explore data and results in an interactive and dynamic way.

Table of contents

List of tables

List of figures

Disclaimer

The results presented in this thesis are based on my own research in the Department of Geography at the University of Salford, Great Britain. All assistance received from other individuals and organisations has been acknowledged and full reference is made to all published and unpublished sources used.

This thesis has not been submitted previously for a degree at any Institution.

Signed: Leeuwarden, May 2000

Ruurd Maasdam

Motto

Explorers are we intrepid and bold,

Out in the world amongst wonders untold,

Equipped with a wit, a map and a snack

We're searching for fun and were on the right track

The Indispensable Calvin and Hobbes

A Calvin and Hobbes Treasury

by Bill Watterson

1. Introduction

1.1. Prologue

In ancient times the monitor was a man at the front of a (war)ship, sitting in the crow's nest at or near the top of the central mast, or on the bowsprit in front of the ship, watching and warning for land, reefs floating mines, whales, friends and enemies. The monitor informed the captain and the crew. The system to be managed comprised the sea, the weather, the boat, the crew, the cargo and the captain sailing with a specific mission. This metaphor can help to see the use of monitoring and the need for information to get a goal. In fact, the monitor searches and warns for differences that matter to the captain, in this case, the differences that matter for the water manager. Monitoring is essentially transmitting 'differences' of qualities of a water system in time and space, which are relevant for developing or maintaining an ideal or preferable situation.

The ultimate goal of monitoring is to provide information, not data. In the past, many monitoring programmes have been characterised by the "data rich, information poor syndrome" (DRIP-syndrome; Ward *et al.*, 1986) . There should be more attention on the analysis and further use of collected data so that the end product of monitoring is information. Another notion that became obvious over the last decade is that the connection between the *data collected* by monitoring and its use within management and policy is an important element in the success of any (water quality) monitoring system design. Water quality monitoring (information) systems should be a balanced combination of data collection and information generation. This is illustrated in the monitoring cycle (Figure 1, Timmerman and Hendriksma, 1997), which illustrates that monitoring is a sequence of related activities that begins with the definition of information needs and ends (and starts again) with the use of information products. Too often water monitoring has been viewed as only the first three steps listed above. In other words, once data are stored in a computer, the monitoring task is completed. Data are hereby viewed as the final “product”. The basic premise of this thesis is that at that point of obtaining water (monitoring) data, one is only half way towards the goal of having information about watersystems.

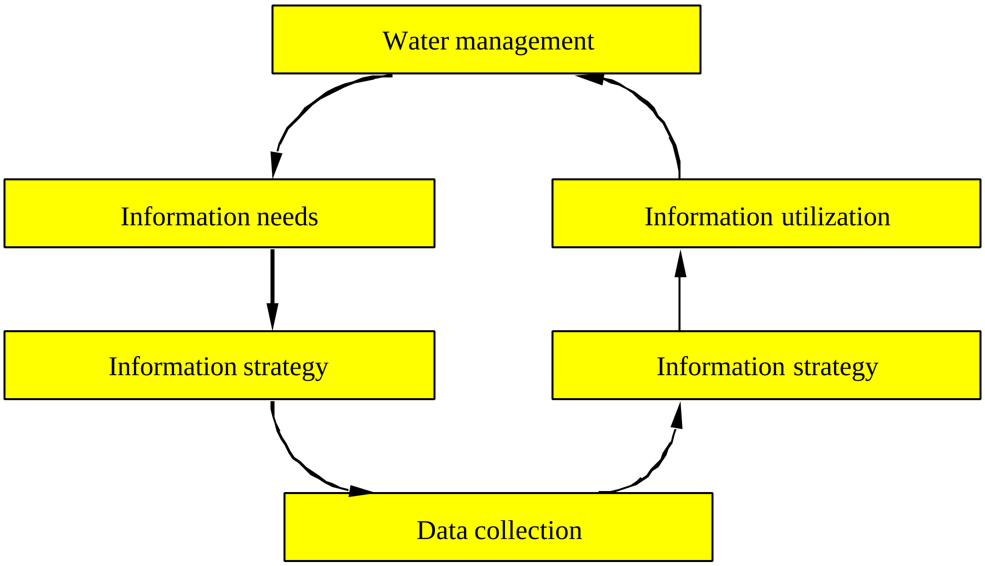


Figure 1 The information cycle (after Timmerman and Hendriksma, 1997)

In order to make monitoring systems more effective and efficient there is a large concern making monitoring tailor-made (e.g. Adriaanse *et al.*, 1994; Ottens *et al.*, 1997) or optimising monitoring (e.g. Blind and Van der Wiele, 1998; de Vree and Blind, 1998). The main focus of this discussion is to design a monitoring program that is fitting the information needs and consequently a best-fit of monitoring strategy and (statistically) network design (Ward *et al.*, 1990). In general the focus on these approaches is on the data-collection side of monitoring systems. An other approach to overcome the DRIPS-syndrome is by starting to use data more efficiently (Openshaw, 1991).

Tailor-made implies 'providing the right information to make management decisions' (Adriaanse, 1997). The focus in this thesis will be on the information generating side of monitoring systems, especially on data analysis. This element of the monitoring process is regarded as a somewhat separate world of expertise. Recent developments in computing hardware and software made it possible for a broader public to use data more effectively and obtain almost instantaneously results of simple data analysis. These technological and scientific improvements in recent years have not been institutionalised in many monitoring programs.

In many monitoring agencies databases (and information systems) are not efficient and not effectively used for information processing, analysis and visualisation of data and decision-support functions.

Water quality data, today, are in many ways just lying about waiting to be combined, analysed and interpreted in more meaningful and relevant ways for the public and managers.

This has two types of implications;

- data are not easily accessible for management purposes;
- water quality programs remain largely invisible because of the lack of highly visible data products.

As a result such programs often fail to win managerial and political support (Ongley, 1997). Monitoring programs could directly (initially) benefit if existing data are analysed, interpreted and reported to the public. Water quality data however, due to the many constituents and resulting large volumes of numbers, can be overwhelming to analyse, interpret and understand. New information technology (e.g. data-mining capabilities, Geographical Information systems (GIS), analytical tools (statistics, graphics etc.), decision-support capabilities and visualisation capabilities) could be valuable attributes to analyse existing water quality data. These technologies should be operational for a monitoring audience (with a minimum of training). On the other hand monitoring professionals should adopt new data analysis, interpretation and reporting procedures to their existing monitoring systems, where the historical focus has been simply collecting data.

1.2. Monitoring Systems are Geographical Information Systems

Most published definitions of "geographic information systems" refer to both data and operations, as in "a system for input, storage, manipulation, analysis and output of geographical referenced information". In turn, geographical referenced information can defined fairly robustly as information linked to specific locations on the Earth's surface. In almost all types of water monitoring information samples or measurements are located in a hydrological cycle. Generally these efforts attempt to describe water conditions on a specific (measuring) point or over larger areas (e.g. watersystems, catchment areas). This difference of location on which the data/information is being collected is sometimes referred as the micro- and macrolocations of monitoring systems (Ward *et al.*, 1990, Sanders *et al.*, 1987). Water monitoring systems are geographical information systems. Information systems in the sense that they are a collection of *procedures*, *activities*, *people* and *technology* set up to the collection of relevant data, its storage until is required, its processing to help provide answers to a specific set of questions, and *communication* of the resulting information to the *people* who need to *act upon it* (Silk, 1991; p. 84). Geographical in the sense should provide information on the earth's phenoma. "GIS" at its broadest has become a term to refer to any and all computer-based activities that focus on geographic information; "GIS data" is often used as shorthand for digital geographic information; and the redundant "GIS system" is becoming the preferred term for the software itself (Goodchild, 1999). The "GIS system" requirements of most (current) monitoringsystems are, in fact quite small (usually limited to handling georeferenced site information, spatial mapping, and limited map overlaying).

The general believe in water management is that GIS systems are only for specialised activities because the learning requirements are substantial, the hardware and software cost are high and only specialists can efficiently use such systems (Ongley, 1997). Geography and maps are not considered as a natural (somewhat special) extension of information systems for monitoring.

The science that serves as a basis for monitoring water systems is evolving rapidly. This science is necessarily broad and complex. It has to cover the nature of environmental decision making, aquatic ecology, the statistics of analysing data, the chemistry of water, the toxicity of chemicals to biological organisms, hydrology, data management hardware and software and many other areas of sciences. Invariably the people operating monitoring systems are not experts in all areas necessary for the successful design, implementation and operation of a total water (management) information system.

A combination of water and electricity usually leads to short circuiting. Something similar seem to occur when water managers (of monitoring systems) are confronted with innovations in the field of information technology, even if these innovations are not the most up-to-date.

This thesis will provide a brief discussion on developments in "GIS" beneficial to monitoring systems. Emphasis will be placed on developments in "visualisation" and "exploratory analysis" because these developments have the biggest parallel with the developments in monitoring discussed before. The basic premises is that there should be a balance between a technology-driven path and the demand driven path (Figure 2). This thesis will discuss that the demands of monitoring (and water management) are changing, this has implications on how technology could be used to support these demands.

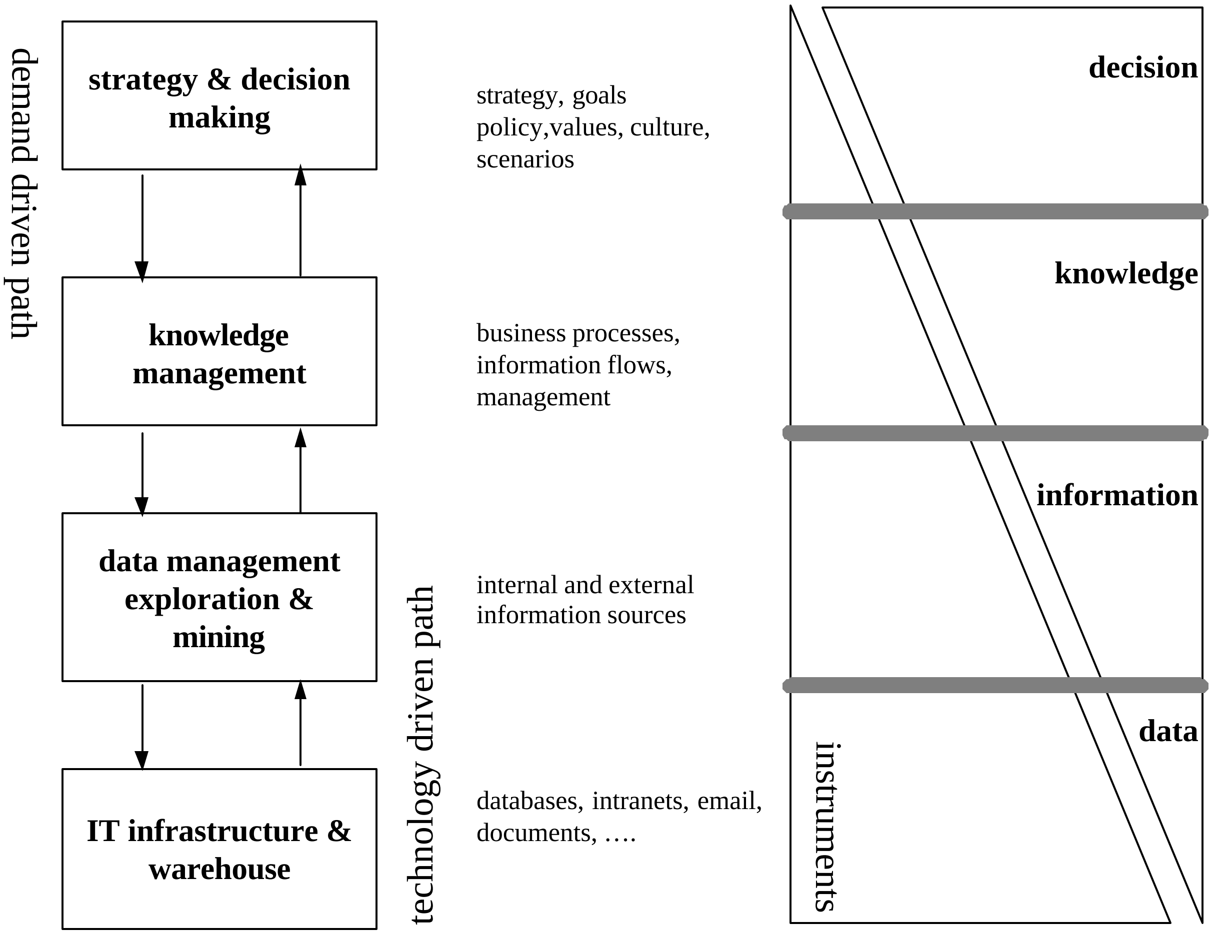


Figure 2 Technology driven versus demand driven changes in information systems.

This discussion gives some perspectives offered by GIS for monitoring systems and could give insight into the benefits of this technology. Emerging technologies in themselves are no solution when human and organisational factors make the difference (Hainje, 1999), but the situation at this moment is that many of these developments are simply not known (or considered).

1.3. Objectives and outline of the study

The overall objective of this thesis is to explore the possibilities which developments of “visual spatial data analysis” offer for water monitoring systems. Visual (spatial) data exploration is defined as a class of techniques for exploring data and information graphically, as means of gaining understanding and insight into the data and information.

Visual spatial data exploration is that field of (geo-)research where map-based scientific visualisation (e.g. MacEachren & Kraak, 1997, Kraak, 1999) and exploratory spatial data analysis (e.g. Anselin, 1999, Fisher *et al.* , 1996) come together. A new branch of science where the map, as such, is a tool to aid the user's (visual) thinking process (DiBiase, 1990), combined with the principles behind interactive and dynamic graphical data analysis (Cleveland and McGill, 1988). This new combination of scientific visualisation (McCormick *et al.* 1987) and exploratory data analysis (Tukey, 1977) is an area of very active research, where interaction and dynamics with maps as an integral part of the process of data analysis are key-concepts. Recent developments in computing hardware and (GIS) software have made it possible to use this visually and dynamic/interactive approaches to data analysis.

The overall objective is reached by taking a number of steps. Each step has it own objective (reflected in the structure of the thesis). These steps and their objective are:

- The description of developments in water management and water monitoring systems. The objective of this step is to understand the user-demand oriented developments (see Figure 2) and to provide an integrated view on water management and monitoring (chapter 2).
- The description of shortcomings and new possibilities for monitoring offered by new technologies. The objective of this step is to understand technology driven developments (see Figure 2) and to provide a new paradigm for monitoring. The notion of this new paradigm is that it is "data driven", "technique-driven" and related with "learning from data" (chapter 3).
- The description of the role maps are getting in spatial data handling. The objective is to give an overview how maps are becoming tools for exploration that facilitates thinking, problem solving and decision making. This includes an overview of software development for visual exploration (chapter 4).
- The assessment of visualisation strategies in a case of monitoring for water management in the Netherlands. This case is based on the new systematic and quantitative assessment methodology for water systems (Regional Water System Report) and illustrates the possible use of visual exploration within this methodology. The objective of this step is to show the importance and perspective of the new concepts and technologies (chapter 5).

Figure 3 shows the structure of the thesis. Besides the introduction (chapter 1) and the discussion (chapter 6) this thesis is divided into two parts.

Part 1 is theoretical (conceptual) and deals with the concepts of water management, monitoring and technology to support this. At the conceptual level several specific concept that link visual spatial data exploration and water monitoring systems will be discussed. This level reviews the changing ideas behind monitoring systems and visual spatial data analysis and their relation. Many of these ideas and concept have been developed recently and are still evolving. Therefore no attempt is made to get a comprehensive overview.

Part 2 describes a practical case involving (spatial) visual exploration. In this case study the Regional Water System Report for the waterboard of Marne-Middelsee (province of Friesland, The Netherlands) is used. The Regional Water System Report is a Dutch initiative to provide aggregated monitoring data to the policy and decision making process in water management. The case evaluates visualisation strategies used in the current practise and demonstrates the new explorative strategy of visualisation. In the case study ArcView and Cartographic Data Visualiser (CDV) are used as software developments to illustrate the changing role of maps in this systematic and quantitative methodology to explore water systems.

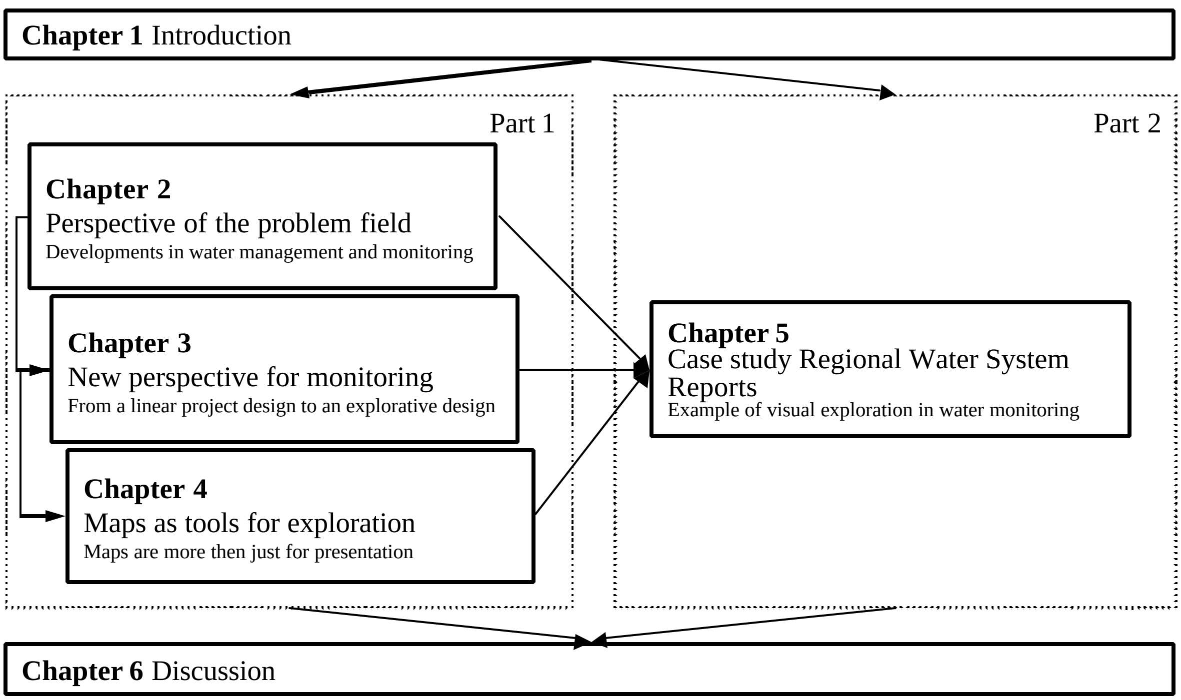


Figure 3 Diagram showing the structure of the thesis.

1.4. Scope and limitations

This thesis will **not** deal with

✗ The "third dimension", space as a three-dimensional phenomenon

- ✗ The “fourth dimension”, time in monitoring programs and visualisation
- ✗ Modelling, remote sensing and bioassay as techniques for data-collection and information generation
- ✗ Truth and proof (“in God we trust, the rest have to bring data”).
- ✗ Perception and visualisation as mental cognitive processes.
- ✗ The (essential) differences between data, information, knowledge and decisions.

"The tools we have most commonly used to interact with data, such as the "desktop metaphor" employed by the Macintosh and Windows operating systems, are not really suited to this new challenge. I believe we need a "Digital Earth". A multi-resolution, three-dimensional representation of the planet, into which we embed vast quantities of geo-referenced data" (Al Gore, 1998)

This thesis deals with the spatial dimensions of watermonitoring consisting of two geographical axes. Watersystems (and waterprocesses) in the real world are, however, typically three dimensional in physical space. In this thesis the third dimensions is not included, but visualisation with three dimensions is a rapid growing interest. The study of environmental phenomena in the third-dimension is a rapid growing research field.

“The passage of time is important only because changes are possible with time” (Shoham and Goyal, 1988, p420.)

Changes en trends and therefore time is an important aspect in water management. To monitor changes and trends is one of the basic objectives of most monitoring programs (Niederländer *et al.*, 1996). The evolution of processes and phenomena through time has always been recognised as a fundamental fact underlying information systems. For lots of applications using a GIS, there is a need to represent evolutionary processes. The requirement to incorporate the temporal dimension in GIS is undisputed and well-documented (Langran, 1992) Time has traditionally been thought of a “fourth dimension”, after three spatial dimensions, and this tends to suggest that time and space should be modelled using the same techniques (Langran and Chrisman, 1988). It is certainly the case that, just as an object occupies a certain range of space, time is only valid over a certain range of time. Despite this apparent similarity between temporal and spatial data, there are also some important differences:

- Time is often "open-ended" so that data are valid from a given time up to the present (or into the indefinite future). The concept of the present is a curious one, since it naturally keeps changing. Spatial data are almost always closed.
- In a similar way, some time is considered a continuum, so that the start of one period implicitly defines the end of the previous one.
- Time has a useful natural ordering.

In this thesis "the fourth dimension" is not further included. An overview of methods to visualise spatio-temporal information on maps is given by Vasiliev (1997). An example of the use of spatial-temporal visualisation in water management is included in Hogeweg (2000).

"Information can be obtained from in-situ monitoring, surveys (relatively short-time inventories) or from rules of thumb, expert judgement, from models, decision support systems or from other system knowledge, such as historical databases. Of course, also combinations of these are possible." (Claessen, 1997)

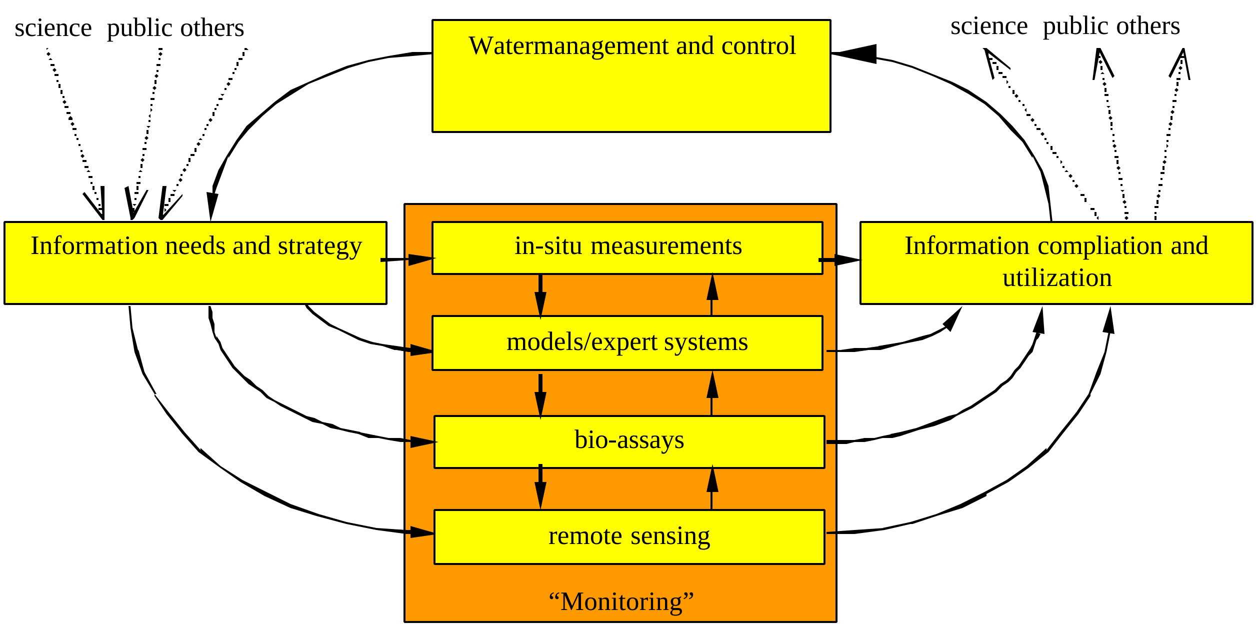


Figure 4 Water monitoring information systems

Figure 4 shows the general information system for water management. Information can be produced in different ways. Information for water management can be obtained from primary sources as monitoring programmes, computations and predictions with models and knowledge-based systems (expert systems), and other sources e.g. surveys (relatively short-time field inventories) or from the rules of thumb, (historical) databases containing statistical or administrative information. Of course, also combinations of these are possible and can offer optimal conditions for cost-effectiveness. This thesis will focus on examples of information provided by monitoring networks (e.g. in-situ measurements). In the broader sense all elements mentioned before can be part of monitoring.

"In God we trust all others must bring data" (Bailey and Gatrell,1995)

"We must appreciate that what we perceive is a function of what we see and what we expect to see. Seeing itself is a complex process. Indeed much of the seeing process is counter-intuitive and appreciating what we see even of simple objects requires an understanding of the eye-brain linkages." (Petch, 1996)

"Effective data visualisation is an outcome of the proper combination of graphical technique, research goal and data type" (Yu and Behrens, 1995)

The promise of visualisation is based on the assumption that human vision and cognition has powerful information synthesis and pattern seeking capabilities that can effectively complement the raw information processing power of digital computers (MacEachren, 1998). While there is a solid base of knowledge about perception and cognition as it relates to static maps, there is less known about the cognitive and perceptual issues associated with dynamic displays. Psychological studies concerning advanced dynamic graphical representation have been unable to support these assumption (e.g. Marchak & Whitney, 1990; Marchak & Marchak, 1991; Wickens, Merwin & Lin, 1994). Many scientific researchers have the (implicit) believe that dynamic visualisation is a promising data analysis tool, that will facilitate science and will play a role in "better" decision-making. While anecdotal evidence may seem to support these assumptions, anecdotal evidence by its nature is generally positive. As long as there is little systemically collected empirical evidence to either support or refute the claim the central notion in this thesis is that proper visualisation is effective. An overview of cognitive aspects of visualisation is given by Petre *et al.* (1998).

Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth. Truth is not beauty. Beauty is not love. Love is not music. Music is the best. Wisdom is the domain of the wish, which is extinct. (Frank Zappa in Joe's garage).

Information, knowledge and decisions are abstract constructions , which have important differences. This thesis will not deal with these differences. It is important to mention that more data will not (automatically) lead to more information, more information will not (necessary) lead to new knowledge and more knowledge not lead to better decisions.

Decisions are often made on the basis of different information sources. In water management there are also enough examples were decisions are based on “accidents”, although the “knowledge” and “information” was present long before:

- Strict emission regulation from drins, cadmium, mercury, PCB’s after cases of lethal poisoning (Japan and The Netherlands)
- Strict shipping regulations (MARPOL) after disasters with the Torrey Canyon and the Amoco Cadiz.
- A 50% load reduction to the Rhine of several contaminants as a result of the Sandoz affair (the Rhine Action Program)
- A 50% load reduction to the North Sea of several contaminants and the acceptance of the “precautionary principle” at the Second North Sea Ministerial Conference in response to phenomena as toxic algae, anaerobic conditions in the German Bight in 1981-193, fish diseases and large scale fish mortality (the North Sea Action Program).

“Management by accident” instead of “management by knowledge and information” (Luiten, 1995). This thesis will treat data, information, knowledge and decisions as a rational, quaternary unity.

1.5. Position statement

In scientific writing there is a general agreement that writing should be in the third person. It is not done to use opinions as facts and in the following chapters personal opinions are perfectly hidden by quoting or citing other researchers. This chapter will somewhat break with this tradition and contains an position statement (much like those which on the internet site of the Varenius-project (see for instance www.ncgia.ucsb.edu/conf/sa_workshop/papers.html). This position might, in the tradition of explorative analysis, change when it is researched further and hypotheses (presupposition or proposition) are rejected.

This thesis has it is roots in my personal interest in exploratory data analysis and GIS. My initial interest in GIS came after an research on the first 2 year's physico-chemical data from the New Zealand's National River Water Quality Network (Maasdam and Smith, 1994; Smith and Maasdam, 1994). Where after the publication someone asked me why we hadn't used GIS to relate the environmental variables (such as catchment erosion, soil, rock type and land use) to the chemical data.

The analytical functions of current off-the-shelf GIS software are however insufficient to analysis the wealth of multivariate data offered by many monitoring programs. Hopeful was the lecture of Henk Scholten (UNIGIS workshop spatial analysis) that exploratory data analysis was the most likely canditate to be added as spatial analysis in the GIS world (see Fisher, Scholten and Unwin, 1996 or

Scholten and LoCascio, 1997). This led to the investigation of the exploratory spatial data analysis (ESDA).

I am working as a research co-ordinator for Wetterskip Fryslân. This is a regional waterboard in the Northern part of the Netherlands. By the virtue of my profession I am responsible for the water quality monitoring activities of this waterboard. In this capacity I have always stressed that the data produced by these monitoring programmes should be validated, archived and made accessible. In many waterboards (e.g. Waterboard Friesland) the weakest link in monitoring programmes still is the proper storage of data. If data are not accessible and complete with respect to the conditions and qualifiers pertaining their collection and analysis or properly validated, the data will never be able to satisfy any information needs. Many waterboards are working hard to improve this data management. See for instance Hainje (1999) on the implementation of a GIS-system (e.g. INTWIS) in the Friesian Water Boards.

In the next phase the data have be to converted into information, this involves data analysis and interpretation. Routine data analysis should be directed towards obtaining information on common conditions, trends or changing conditions and testing for compliance with a standard. Each one of the information goals is based upon a statistical techniques. However, there are several data characteristics, data limitations or statistical characteristics which complicate the proper statistical analysis of water quality data (e.g. missing data values, changing sampling frequencies, multiple observations, censored data, small sample size, data outliers, non-normality, seasonality or serial correlation). Because of this limitations most of the data analysis is carried out using multivariate (e.g. Maasdam and Smith, 1993) or descriptive (e.g. Maasdam and Claassen, 1993; 1994) explorative methods. My experience is that given the above limitations many commercial available software packages (such as SPSS, Systat) are only in a limited situations suited for analysis of water quality data. In the U.S.A. (for instance WQSTAT of Colorado State University) or U.K. (at WRC) there are few systems developed specifically for water quality applications, but these applications are not used within the Dutch waterboards. Given the availability of hardware and software (and the ever-growing amount of data) my personal focus remains on the explorative data analysis.

Interpreting and assessing water quality is only one step in converting data information and presenting it to an audience for whom the information is meant. Monitoring programs produce large amounts and often very diverse type of data. It is useful to aggregate the data to a higher and more understandable level. Data aggregation is already inherent to most water quality assessment methods (e.g. ecological indices). The concept of indices remains in my opinion the most promising way for conveying the results of monitoring to a non-technical users and to the public. Regional Water System Reports is the most recent attempt to develop an assessment method for water systems. Currently I am a member of an working-group attempting to implement this methods in the Friesian water management organisations (waterboards and province). In 1997 this working-group was the first to experiment with the new methodology (Witteveen en Bos, 1998), resulting in the proposal of an new presentation method (Van der Straten *et al*., 1998).

The most prominent advantage of the RWSR is that it converts the results of several variables to one quality evaluation which creates a rapid manner in which to express the complex concept of "water systems" as a single evaluation figure. This advantage is also a big draw-back, because such a high level of aggregation is not obtained without loss of information (of lower levels of aggregation) and discussion what data should be included into the few indices. Within the national working-group of Regional Water System Report there is a tendency to develop a manual and a procedural framework which is rigid and standardised. In my view this is not practical or desirable since it will not lead to a representative situation. Regional Water System Reports should be defined by each monitoring organisation (within a general framework) and should be flexible to meet the goals of regional water management.

The information system to support Regional Water System Reports should support an explorative way of working. A better insight is obtained if, instead of predefined calculations, "indices" can be examined (e.g. leaving out locations, observations or variables). Aggregations by means of indexing is practical but proper graphical presentation is ever so important. Providing a better view on the results and allowing any trends, outliers, oddness to be recognised at a glance.

This thesis is a strong plead to use maps and graphics in an interactive and dynamic way to explore data collected by water monitoring systems. This focus to explorative data analysis is regarded as the first (and essential) step to overcome the shortcoming of the data-rich-but-information-poor environment of many monitoring programs. This proposition is a however a presupposition and is not the conclusion of comparative research. This thesis should on the first place be regarded as a plea, an *advocacy piece*.

PART 1

2. Perspective of the problem field

2.1. Introduction

A good way to address the problems related to the inadequacy of monitoring systems to profite information for water management is putting the problem field in perspective.

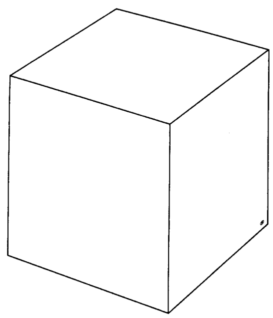
This first chapter attempts to summarise the relevant developments in water management and monitoring systems in the context of time. Major developments are occurring in the sphere of water management, watermonitoring (and information technology) in the last ten years. The crucial point of this perspective is that developments in water management, watermonitoring and information technology is that there might be shift in paradigm; a tendency towards the exploration paradigm .

2.2. Developments in water management

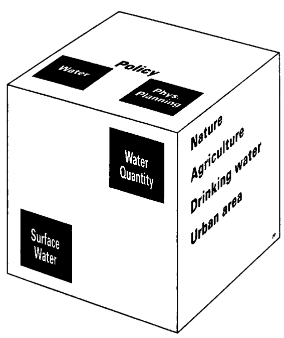
"Water management can be defined as: research, policy and operational activities, pursued to maintain, to control, to manage and to develop functions and (potential) uses of the Watersystem involved" (Claessen, 1997).

Major developments have occurred within the sphere of water management for thousand of years. A similar situation exists in The Netherlands today and it speed at which developments have followed on each other has made the last decades so special. Although the emphasis has been put on the Netherlands, the scope of the development is more or less also of global value.

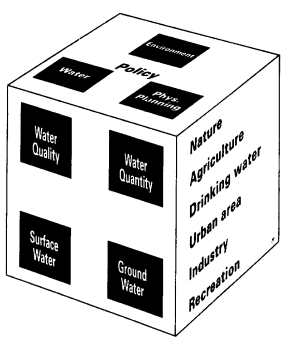
The development of water management until now can be defined into five stages (Van Rooy and de Jong, 1995; van Rooy, 1997). (van Rooy and de Jong, 1995). Every stage can been visualised using a typical model of the watercube (van Rooy, 1993, van Rooy *et al.*, 1993).



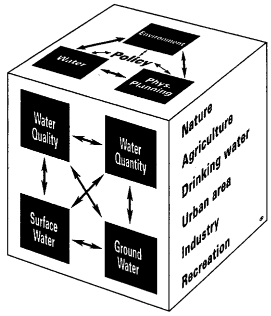
1. *Water without management*; The first deals with a time when neither water management nor the necessity for it existed: a period when man lived in a certain harmony with his environment.



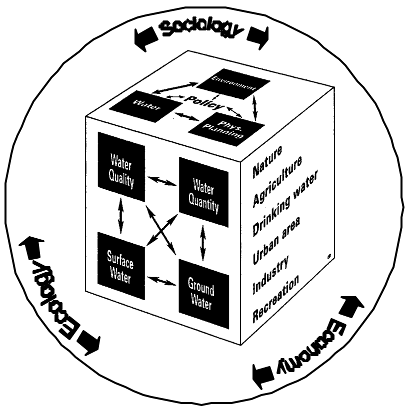
2. *Basic water management*; The second stage concerns man's first conscious and large-scale use of water, called basic water management. Although the number of uses was limited, by various technical means greater quantities of surface water came under man's influence.



3. *Sectoral water management*; As the number of water uses increased, more specific technology had to be developed and implemented to meet the different demands. This third stage, that of sectoral water management, is marked by a growing realisation that these increasing demands were leading to a shortage of good quality water as well as causing serious damage to the environment.



4. *Integrated water management*; The response of these fundamental problems has been the concept of integrated water management. This fourth stage involves adjusting water management, water uses and related policy areas to each other. The key words in this stage are "integrated", "sustainable" and a "systems approach" .



5. *Comprehensive water management;* Although integrated water management has yet to be put into practise on a broad scale the following stage has already begun to manifest itself conceptually. Stage five comprehensive water management, set integrated water management in an interactive framework with ecology, sociology and economy.

The five different stages are distinguished by basic, sectoral, integrated and comprehensive water management. No fixed time limits are defined between these different stages. They are phased transitions in time that are closely related to the space in which they occur (see Figure 5)

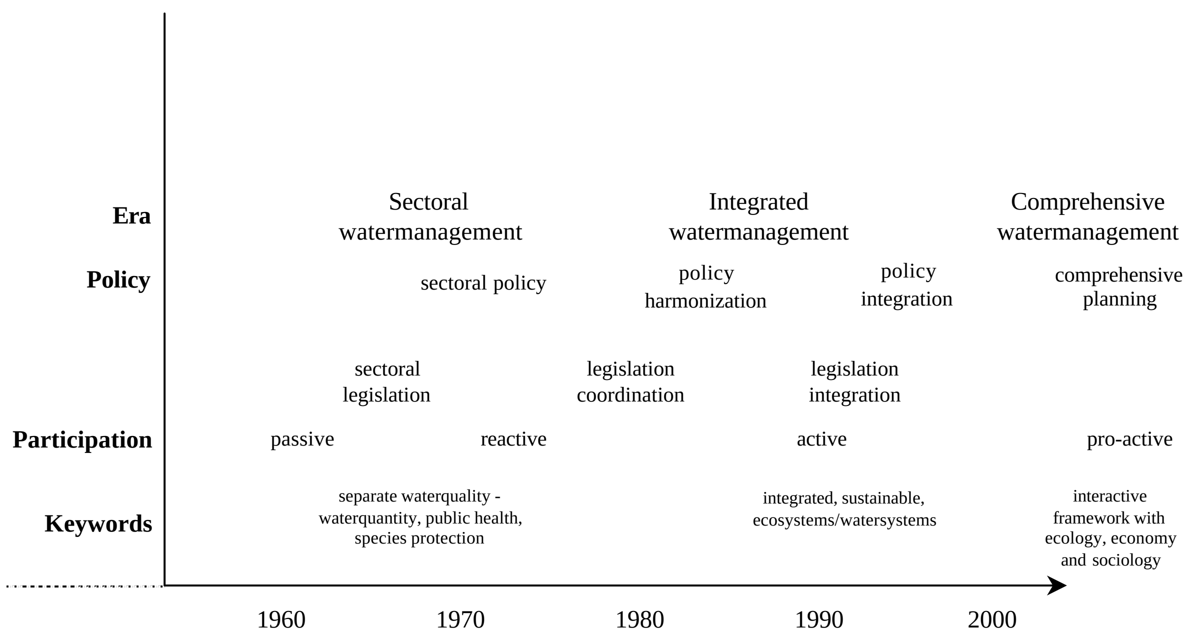


Figure 5 Overview of development in water management in the Netherlands (after Van Rooy, 1997; Misseyer, 1999)

2.3. Transition of integrated to comprehensive water management

In contrast with the sectoral management approach, the elements of water (water quantity, water quality, surface water and groundwater) are no longer regarded as separate entities but as being inextricably linked together as water-systems. A water system being defined as *the water, banks and* *underwaterbeds, including relevant ground water systems, also including the existing ecosystems, and the technical infrastructure*.

Water systems do not always have strict and permanent boundaries. They are limited in space and time by morphological, ecological and functional characteristics. Within one system there are specific goals and/or problems related to specific characteristics. The goals and problems should determine the schematisation of a water system and its relevant environment.

Another consequence of regarding water as a system for a given area, the relationship between element and uses, and between the uses themselves, is taken into account. Furthermore, water is no longer seen as a separate component of the environment; an interaction between relevant policy areas comes into play – namely water, environment and physical planning. Integrated water management thus also means managing the water systems by gearing the different policy and management levels to one another as good as possible (de Jong *et al*., 1995).

The transition towards integrated (and comprehensive) water management will not evolve naturally (Van Rooy,1997; De Jong *et al*., 1995). There are many bottlenecks that hamper the implementation of integrated water management; the way in which water management is organised (institutional bottlenecks), the way in which the parties concerned communicate (communication bottlenecks) and the sharp contrast between our current use of (water) resources and that which is actually sustainable (socio-political bottlenecks) (see Table 1 for an overview).

Categories	Bottlenecks
Institutional	*Fragmentation:* large number of responsible organisations *Setting of tasks*: overlapping and claims on exclusivity *Secondary goals*: protectionism instead of participation *New goals*: lack of support *Set of rules and regulations*: procedures regarding, water, space and environment not tuned. *Mismanagement*: “management by accident” instead of “management by knowledge”
Communication	*Organisation culture*: failure to acknowledge differences *Fields of knowledge*: mutual incomprehension due to differences in culture *Science*: centralising the observer *Advisors*: insufficient transfer of knowledge and know-how *Civilians*: lack of integrated and adequate information
Socio-political	*Use of resources*: exhaustion and consumption *Rationalism*: alienation from reality

Categories	Bottlenecks
	Static approach: denial of the dynamics of water systems *Self overestimation*: denial of limitations

Table 1 A schematic view of the bottlenecks, divided into the three different categories with the key points summarised. (after van Rooy, 1997 and De Jong et al., 1995)

2.4. Definition of monitoring

"Monitoring is any or various devices for checking or regulating the performance of a machine, aircraft, guided missile, etc. " (Webster Dictionary, 1990)

Monitoring is a sort of information system in which during a certain time on a systematic way data are being collected, handled, managed, analysed and presented. The goal of monitoring is to provide adequate information for the recognition, formulation, implementation and/or control of policies and management. Monitoring in the broader sense can be placed in the information-cycle (Figure 1). The guiding principle is that monitoring (and assessment) should be seen as a sequence of related activities that starts with the definition of information needs and ends with the use of the information products. This cycle of activities is called the "monitoring cycle" (Figure 6; Task Force on Monitoring and Assessment, 1996).

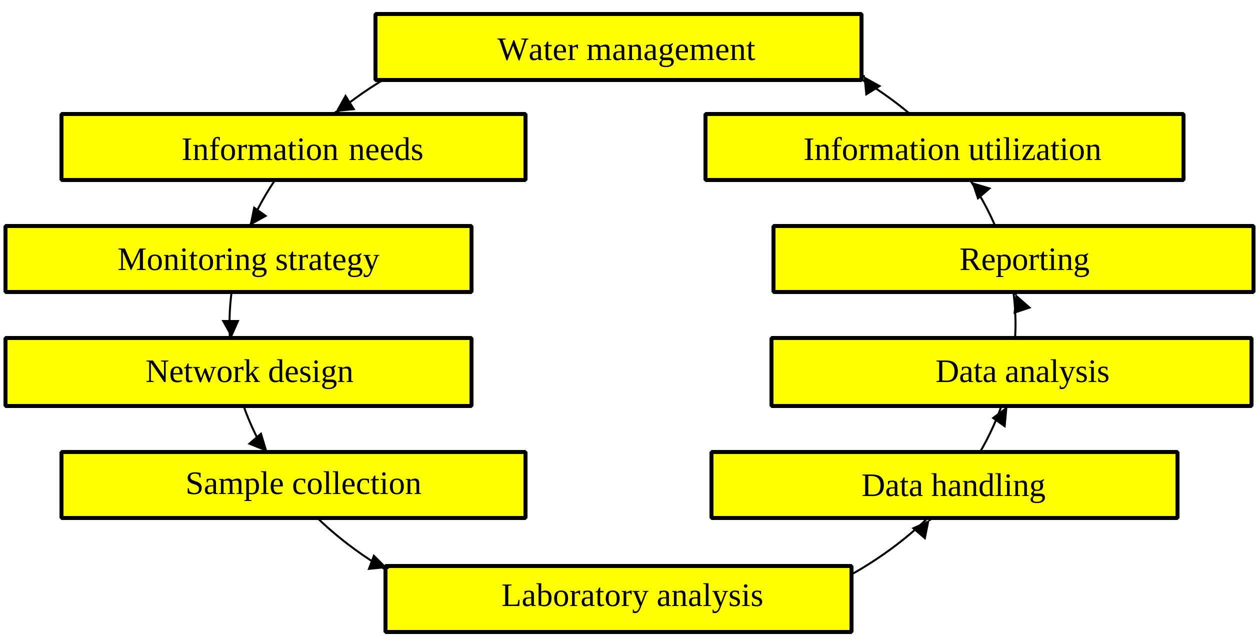


Figure 6 The monitoring cycle

Schemes from the past (and also recent one's) show a more top-down sequence of a restricted number of activities, starting from a rather arbitrarily chosen network and having an open end concerning the production of data.

Dynamic informationneeds require a regular rethinking of information system. It is essential to add, to cancel, to revise and update the concept. Feedback is accomplished if the results of the first cycle of activities are used as a starting point for the next cycle (Cofino, 1994). The ideas behind the monitoring cycle are discussed in two international symposia "Monitoring Tailormade" (Adriaanse *et al.*, 1995; Ottens *et al.*, 1997).

This cycle emphasis that the ultimate goal of monitoring is to provide information, not only data. As in the past many monitoring programmes have been characterised by the "data rich, but information poor" syndrome (Ward *et al.*, 1986), attention should be directed towards the end-product of monitoring, i.e. information.

2.5. Developments in water monitoring

Watermonitoring has also evolved rapidly during the last decades. Misseyer (1999) studying environmental monitoring (in the Netherlands) distinguishes in chronological order registering- , forecasting- and monitoring-related activities respectively. The dominating themes in these activities are "what-was", "what-if" and "what-is". Hotto *et al.* (1997) and Ward, Loftis and McBride (1990) give a discussion on the evolution of water (quality) monitoring systems in the United States. Ongley (1995, 1997, 1998) gives an impression on the evolution in developing countries. The evolution goes more or less along the same line although the timeframe is much smaller.

The development of water monitoring can be defined into five stages (after Hotto *et al.*, 1997; Misseyer, 1997). These four last different stages are distinguished monitoring networks, monitoring systems, information systems as no by basic, sectoral, integrated and comprehensive water management. No fixed time limits can be defined between these different stages, there is a phased transitions in time (see Figure 7).

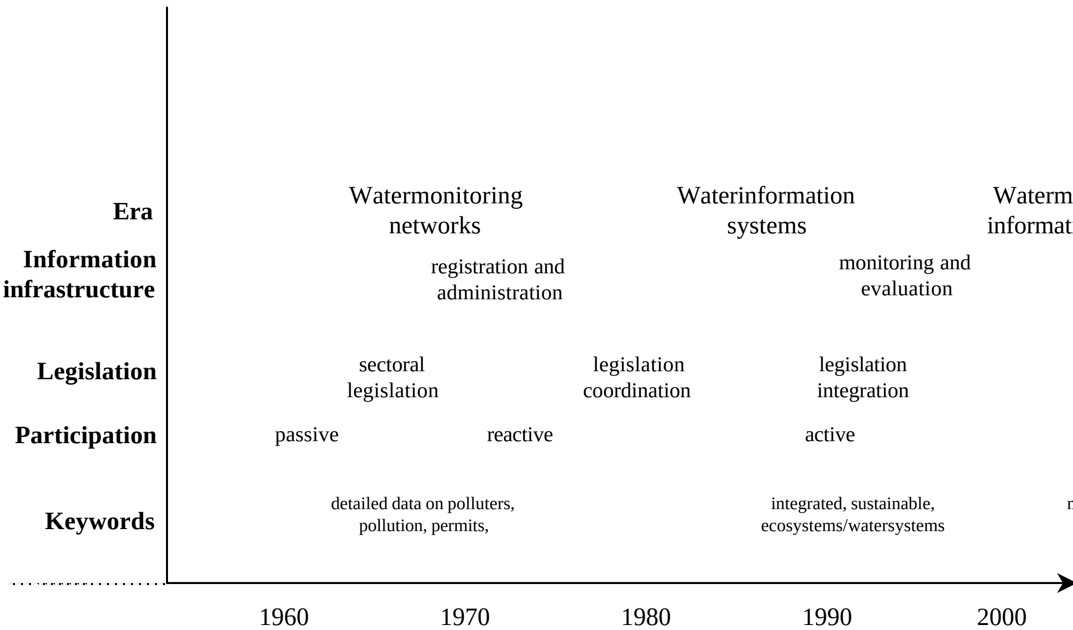


Figure 7 Overview of developments in monitoring systems

Water monitoring systems have evolved through several phases:

1. *No formal water quality data collection or information generation*

2. *Monitoring Networks*: Early twentieth-century water quality data collection networks (U.S.A) were designed to collect data to answer narrow technical questions with the respect to limited areas and limited time periods. Many early efforts focus on the prevention of nuisances. Over time, monitoring network design embodied improved data collection and storage techniques, and an increased appreciation of information needs and the stochastic nature of water quality. The focus began to shift to pollution control, especially when the relationship between water quality and public health began to influence legislation. Monitoring networks really taking a flight when laws on the water pollution are placed in action (USA 1940-1950, The Netherlands 1970). Much of the monitoring efforts is to determine whether water quality is sufficient for a set of designated uses. Samples were collected, checked against criteria for an established use, and determined to be acceptable or non-acceptable. Schemes show a more top-down sequence of a restricted number of activities, starting from a rather arbitrarily chosen network and having an open end concerning the production of data.

3. *Monitoring Systems*: The 1980s witnessed continuing development of data collection and storage methods, refined statistical characterisation of water quality variable behaviour (especially

directed toward trend-detection) and a merging appreciation of the need to design water quality networks according to systems principles. The distinction between data and information was emphasised. Systematic design procedures developed included a 12-Step Process (Sanders *et al.*, 1979), a 5 Step-Process (Sanders *et al.*, 1987), a formal (4-task) Systematic Design Procedure (Mar *et al.*, 1986), the New-Zealand National Network design (Smith *et al.*, 1989), a 5 step (optimisation) process (Schilperoort and Groot, 1983) and “wheel and axle” framework (Payne and Ford, 1988).

4. *Information Systems*: The late 1980’s and early 1990s witnessed continued development of data collection and statistical analyses, and of frameworks to guide the design of more comprehensive water information systems. The guiding principle in this is that monitoring (and assessment) should be seen as a sequence of related activities that starts with the definition of information needs and ends with the use of the information products. This cycle of activities is called the “monitoring cycle” (Figure 6). Increasingly, their potential contribution to broad water management systems and ecological management systems has been recognised. To begin to address the problem, the concept of a “water quality information system” was proposed by Ward (1979) and Ward (1986). Examples of these efforts include a “5 step framework for designing water quality information systems (Ward *et al.*, 1990) and the development of “Data Analysis Protocols” (Atkins, 1993).

5. *Management Information Systems*: Recent water management trends (see evolution of water management) have set the standard for a new era. Water monitoring and information systems must become Water Management Information Systems. As management information systems, these programs must:

- derive from goals and associated water management objectives
- satisfy expanded water management objectives and information needs.
- link directly to specific water management decision needs and decision processes (“monitoring for action”)
- be dynamic, i.e. designed to be continuously reviewed and improved (Hotto *et al.*, 1997; Misseyer, 1999).

2.6. Transition of information systems to management information systems

Too often in the past monitoring was characterised by the “Western” data paradigm (Ongley 1995;1998; Adriaanse *et al.*, 1995), at the moment a new paradigm seems to emerge (see table 2 for an overview). Although not fully specified, this new approach has new characteristics. Data that do not lead to management action, or whose use cannot be stated explicitly, are increasingly being labelled as "not needed".

PAST	FUTURE
Being data rich, but information poor	Value and use of information justifies monitoring
Monitoring occurring without information objectives	Monitoring must be accountable to an information purpose
More is better; money in- data out	Monitoring must be cost effective
Using high-tech measurements must be an improvement	New data-collection techniques supported by information need
"Know how" is better than "know why"	Information produced must support action

Table 2 Trends in monitoring systems

In the past information for water management was inadequate. Monitoring programmes provided policy makers lots of data, pure relevant information was rather scarce (Ward *et al.*, 1986; Laane and ten Brink, 1990). Now that the problems are growing bigger, knowledge is increasing rapidly and water issues stand on the political agenda, the time has come for a more coherent approach. A continuous flow of adequate information must be created. The aquatic outlook (Luiten, 1995) and Regional Watersystem Reports (Witmer, 1995;1997) (see also chapter 5) are two Dutch initiatives to develop a "management information systems".

2.7. An integrated view on the future of water management and monitoring

In the previous sections water management and monitoring have been addressed separately. This chapter gives an integrated view on the future of water management and monitoring.

Predicting future developments is always somewhat tricky. The chapter on the evolution of monitoring and water management contains already visions on what the future will bring; water quality management information systems (chapter 2.5) and comprehensive water management (chapter 2.2). Some of the general "trends" increasing scale of problem, delayed and irreversible environmental responses, growing role of non-point source emissions, unforeseen interactions among different pollutants and media, "surprises", are observed today, and they are likely to be continued. (Somlyódy, 1995).

PAST	FUTURE
Control Type and Treatment	
"End of Pipe"	Source control, closing material cycles, land use management, concern on large projects
"Traditional technology"	Reuse and recycling
	Spatial treatment methods (biological-chemical treatment, upgrading, natural treatment, small-scale treatment) and emerging new traditions
Planning and project evaluation	
Poor/narrow definition of objectives	More "sensible" definition
Short-term view	Long-term view
Cost evaluation	EIA, risk and multiobjective evaluation, social and political impacts
Little concern on failures and adjustment needs	The future is never certain: reliability, resiliency, robustness and vulnerability
Science and Engineering	
Science does not drive actions	"Science for Action"
Problem isolation and engineering solutions	Improved planning
Interdisciplinary gaps and barriers	Integration of quantity, quality, hydrology, and management
Legislation, Decision making	
General rules and rigidity	Specific rules and flexibility
Fast implementation (a misbelieve)	Process view
Little enforcement	Improved enforcement
Command and control	Polluter (and user) pays
Confusing institutional settings	Clearer structures and less barriers
Decisions by politicians and administration	Public awareness and participation, NGOs and enhanced communication
National policies	International policies
Monitoring	
Monitoring does not drive actions	Monitoring for action
Individual issues (processes, control, operations, planning etc.)	Integration (model library, DSS, GIS, expert systems etc.)
Use by experts	Use in administration, meetings, by public and organisations
"Western data-paradigm", data rich but information poor	"new monitoring paradigm", value and use of information justifies monitoring.
Local measurement	Networks, remote sensing, continuous measurements
Conventional (chemical) variables	Special variables (eco-toxicology, biomonitoring etc.)
Monitoring of water	Integration of effluent and ambient monitoring and environmental monitoring
Poor data availability –poor analysis	Improved availability (databases, digital maps, telecommunication) and data analysis

Table 3 Trends in water management and monitoring (after Somlyódy, 1995; Van Rooy, 1995; Villars, 1995).

Overlooking the trends in water management and monitoring there are number of concerns:

- Water managers will face fewer decisions based only upon technical criteria and more and more decisions will also involve non-technical, qualitative considerations (e.g. social, economic. legal and political factors)
- Water managers will be held more accountable (by the public, watersystem stakeholders and public organisations) for the operation of their management systems and the contribution of those systems to the actual quality of water.
- Recourses available for water management purposes will become increasingly constrained and uncertain as regulatory philosophy is reviewed and as state agencies, waterboards, municipalities and private industry all try to operate under reduced budgets.
- Large system perspectives (e.g. watersheds, river basins, water systems, ecosystems) and complex problems will call for integration and co-ordination of systems, more information sharing and more participative, open and diverse decision making processes.

To meet public water (quality) goals and corresponding management objectives, water managers (in the future) will require effective and efficient monitoring information systems more then ever.

3. A new perspective for monitoring

3.1. Introduction

In the ideal world a monitoring systems, like a scientific research, begins with clearly stated objectives and problem. Some decision must be made, some question of scientific, social or environmental concern must be resolved. For generation of "academic" researchers, the formulation of research hypotheses has been followed by choice of a data collection method, (a designing a survey or sampling schedule as appropriate), identifying a sample design, piloting, field collection of data (with verification and resampling), collation of results, analysis and report-writing. Goodchild and Longley (1999) speak of the sequential 'linear project design' which is a model for contemporary research in natural and social science.

The monitoring cycle (see chapter 2.4) is only a slight variant on the 'linear project design', like most robust designs it also includes feedbacks and checks to ensure that the principles of good design are not overly compromised in practical implementation and to check whether the information needs have not changed to much in time.

Goodchild and Longley (1999) suggest that there are circumstances and imperatives that will lead to the emergence of a new kind of spatial analysis in the coming years. Many of there observations are valid for water management and monitoring. This chapter will discuss some of the new developments that could lead to a new paradigm. This new paradigm is here loosely called the 'exploration paradigm' as it has much associations with 'explorative data analysis' (Tukey, 1977) and 'greater statistics' (Chambers, 1993). The notion in this new paradigm is that it is 'data driven', 'technique-driven' and related with 'learning from data', from the planning, collection of data to the analysis, presentation and reporting.

In contrast to the coherent research design, the terms 'data driven' and 'technique-driven' are highly infouvarible (pejorative) in research generally, as are such phrases as 'a technique in search of a problem' –in this ideal world, the statement of the problem strictly proceeds with the collection of data and the performance of analysis (Goodchild and Longley, 1999). But probably the best way to overcome the data-rich but information syndrome is to start using data (Openshaw, 1991).

3.2. The linear project design

For long period a essentially linear project structure with recursive feedback's underlies student dissertations, governmental reports and research activities. The sequential events in this design together constitute a holistic research project and the feedback's are all internal to the research design. This classic research design is illustrated in Figure 8.

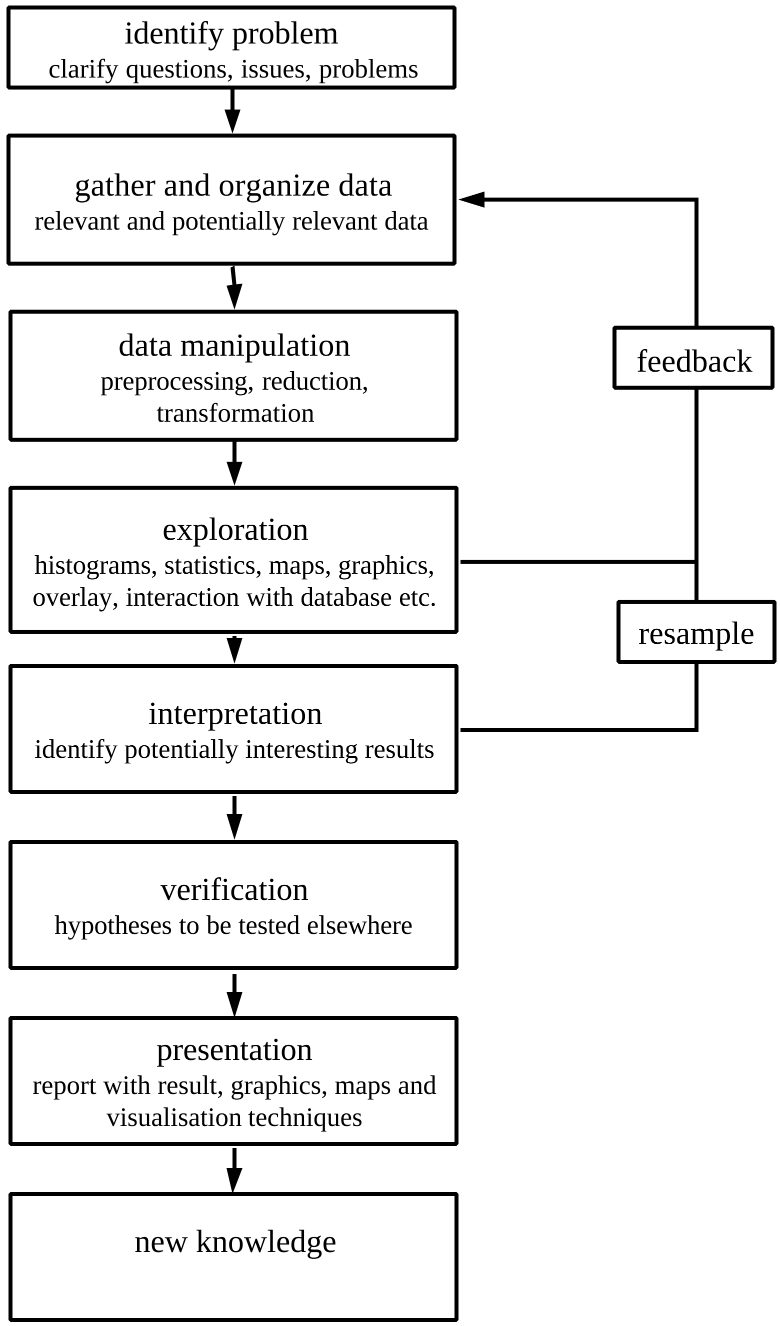


Figure 8 Sequential stages of a typical project design (after Goodchild and Longley, 1999 and Getis, 1999)

Once the project has been initiated, the availability of existing data has no further influence upon problem definition; methods of analysis that are consistent with the type, quality and amount of data to be collected are identified at the design; the sample design is not guided by considerations and priorities outside the remit of the research (Goodchild and Longley, 1999). Getis (1993, 1999) noted that in a GIS-environments there are two important adjustment to the sequential linear project design.

In the traditional approach the hypothesis-testing was straightforward with little opportunity to recast hypotheses. Geographical technology has added the opportunity for guided hypothesis inquiry, and most steps have expanded to include more opportunities to assess the data from different (vantage) points. The data manipulation step was added, because the geographical technology present the analyst with opportunities to use larger samples, view data over a series of map scales, and generally to extent the collection of data to relevant and potential relevant data.

In the linear project design a study must meet a number of strict requirements. It should:

- have a clearly stated aim
- have well-defined populations, observation process, hypotheses to be tested and parameters to be estimated
- use random sampling/ random assignment of treatment
- have previously specified probabilities of making false decisions
- use a study design based on the requirements above
- have conclusions that concern only the questions stated in the study.

Studies not meeting these requirements (and most research does not meet this requirement) should be labelled 'exploratory', since they do not produce conclusive, statistically sound statements.

In the linear project design the emphasis is on testing some hypothesis. Though often forgotten, the proper way to conduct research is to chose the way of analysis before the data are gathered so their collection can be appropriately designed. To a large extent the choice of analysis is determined by the objectives of the project and the nature of the data to be analysed (and to a lesser extent by what is available on a statistical package or GIS). Opposed to exploratory analysis, it is unacceptable to do a number of different types of test for the same null hypothesis on the same data set, and then pick the one that gives a "significant" result. Hypothesis test have to be chosen with an eye to the underlying model, that fits the situation (type of data, type of distribution etc.) and then accept the outcome. Fishing for significant results is not permitted.

3.3. Shortcoming of the linear project design

3.3.1. Finding the problem

John Tukey (1980) has stressed, “Science .. does not begin with a tidy question”. It is a truism that “finding the question is often mere important than finding the answer”, but research generally emphasises the latter, perhaps because it is easier to understand. Tukey observes that questions are generated mainly “by quasi-theoretical insights and exploration of the past data”. His 1977 text

Exploratory data analysis (EDA) introduced a number of robust techniques for making sense of complex data sets. More than a collection of techniques, EDA is in essence an attitude “a willingness to look for what can be seen, whether or not anticipated.” (see also chapter 4.3)

Many monitoring programmes keep persisting in measuring variables which are “easy to determine-why not” (Literathy, 1997) and “we are good in this” (Ongley, 1998). Traditionally the aims are mainly defined as a collection of scientifically measurable variables. These variables have in particular ascribed a formal status (e.g. standard or norm). In many countries there is a legal remedy to create list of chemicals for monitoring, even when it is neither efficient nor effective (Ongley, 1998). As Lijcklema (1997) points out, “there is a tendency among managers to focus their objectives on attaining the norm. The standards become the highest aim and blur to focus the underlying real objective. (..) A certain myopia is not foreign to a number of employees of the water authorities; their creativity is not challenged and maybe it is more convenient to shelter behind rules than to struggle with the complex matter where so many interests and view have to be reconciled.” An approach that is principally based upon reasoning, rules and regulations does not act as an ideal breeding-ground for greater creativity when dealing with water systems (van Rooy, 1995).

3.3.2. Data is everywhere

Today's research environment is characterised by datasets which are collected by many different means and which pass through many hands (see also chapter 3.4.1). More and more data are second hand and more data are collected using unscientific research designs. Often principally not collected for "research" at all (Longley, 1999).

Many water quality monitoring programs are still dominated by a chemical monitoring. At the same time new developments in the field of toxicity parameters, biomonitoring, remote sensing, continuous measurement may offer an alternative for the chemical—specific approach (Villars, 1995; Zwart, 1995; Tonkes and van de Guchte, 1995; van Loon and Hermans, 1995) or may more adequately fulfil the future information needs.

3.3.3. Spatial is special

Anselin (1990) explains in considerable detail why its necessairy to treat spatial data differently than other types of data. In essence, the point is that spatial effects complicate any staightforward understanding of spatial data. "Spatial effects" have two interrelated meanings (Anselin and Getis, 1992). The first is that embodied in Tobler's First Law of Geography (Tobler, 1979) where "everything is related to everything else, but near things are more related than distant things". The more torublesome second meaning is that relationship within or among variables can be determined by the size and configuration of spatial units as by the nature of the variables.

3.3.4. Spatial analysis virtually absent

Spatial analysis might be defined as a set of methods useful when data are spatial (Goodchild and Longley, 1999), in other words when data are referenced to a 2- (or 3-) dimensional frame. To distinguish analytic methods from more general (mundane) operations, they might be defined as methods for processing data with the objective of solving some scientific or decision-making. (Getis 1999). One might assume that Geographical Information Systems (GIS), specifically designed for the storage, manipulation and display of geographical data, would provide an useful environment for spatial analysis. But unfortunately in current GIS the focus tends to be on displays, organisation and simple manipulation of information in spatial data bases. The lack of analytical capacities of a GIS is by now a familiar complaint in research literature (e.g. Goodchild, 1987; Burrough, 1990)

3.3.5. The Well-informed analyst is absent

In the traditional linear project design there was the well-informed analyst who knew what methods of analysis should be performed to test the predesigned hypothesis. Within water monitoring this is a panacea, there is hardly any personal with a proper statistical education.

However monitoring professionals will be asked to create more publicly relevant information from existing water quality data. This will require monitoring professionals assuming a leading role in adding data analysis, interpretation and reporting, for public consumption, to their existing monitoring system operations. The resources (computer system and data analysis software) and skills (statistical, public relations and GIS) required to do this will have to be increased in many monitoring system operations where the historical focus has been simply collecting data (Ward, 1997). The number of fields of knowledge necessary for successful design, implementation and operation of a watersystem information system are broad and complex. It has to cover the nature of environmental decision making, aquatic ecology, the chemistry of water, the toxicity of chemicals and organisms, hydrology, datamanagement hardware software, the statistics of analysing data and many other areas of science. Invariably the people to implement monitoring systems are not expert in all areas necessary for the successful design, implementation and operation of a total monitoring system.

The advent of the digital computer has changed the world fundamentally because it became possible to perform a method of analysis that monitoring professionals do not know everything about. Methods emerged, beginning in the 1970's and particularly in the area of multivariate analysis that are impossibly to perform by hand. Fundamentally there was a shift towards the use of analytical techniques by studying the nature of its response, rather than studying the procedure which generated the response (Goodchild and Longley, 1999). Especially the innovations of computer graphics and windows, icons, mice and pointers (WIMP's) have created a more intuitive environment for exploratory data analysis in a spatial context (Anselin, 1999 ; Kraak, 1999).

3.4. Elements of a new perspective

This chapter explores factors affecting the uprise of a new perspective for monitoring (and environmental science).

3.4.1. Data sharing

In the past (nearly) every organisation dealing with one or more aspects of water management, built its own monitoring network. Moreover for operating water management systems and for devices, such as manipulating sluice gated, inlet works for water supply, reservoir management, water level control different monitoring networks were constructed. Initially these were often single-purpose operating systems. As a result separate monitoring networks were developed at different geographical scales, for different scientific disciplines, for different sectors of water users and data where not shared (Claessen, 1997).

The huge number of monitoring programs is not by definition problematic. The notion of integrated water management has increased the need for data sharing and data exchange as part of interaction between organisations. The need for good co-operation and an open attitude towards data and information has led to a declaration of co-operation signed by all the major Dutch water organisations (convent informatievoorziening voor de sector water http://www.waterland.net/index2.html).

3.4.2. New computer architectures

Mainstream Information Technology is being driven by user requirements and users look for enterprise wide access to information supporting them in their daily tasks (Roodzand, 1999a). It is important to provide this interoperability and scalability in the IT infrastructure. Traditionally (Geo)graphically oriented systems (CAD, GIS, AM/FM) formed a separate environment in the organisations (different databases, different user-groups). The inaccessibility and interoperability were the driving forces behind the Open GIS Consortium (http:/www.opengis.org)

Modern information technology makes it possible to merge the traditionally highly technical specialised, separated technology via the enterprise database with the other business applications (Roodzand, 1999b). The main characteristics of this integrated spatial technology are:

- Central databases for the whole organisation. Spatial extensions to industry standard databases allow the storage of spatial-data alongside the alpha-numeric data (e.g. Oracle Spatial Cartridge or Sybase's SQS)
- Better integration provided by open systems (e.g. "open" products like Geomedia)
- Predictable, user-friendly and easy-to-learn applications as a result of Windows-dominance and the growing popularity of the internet (see for developments like CommonGIS)

Extending "administrative" applications with the "picture" of spatial information will especially benefits the user-viewer group with better quality, more comprehensible and easier accessible information ,allowing them to make decision quicker and probably better (Roodzand, 1999a).

3.4.3. New techniques for analysis

The combination of vast new sources of data and high-speed computation have led to new methods of analysis. These include OLAP (On-line analytical processing, MQE (Managed Query Environments), Data mining (e.g. neural nets, fuzzy sets , decision-trees, knowledge discovery in Databases (KDD)) (Fayyed *et al*., 1996) (see table 4 for a short overview).

Technology tool	Question	Sample output	User
GIS, SQL, Querying, interrogation and reporting, ESDA	"What is ..?" "What happened ..?" "Where is..?"	Monthly reports Asset histories Maps	I'm interested in history. I'm interested in visualisation.
On-line analytical processing (OLAP)	"What happened and why did it occur?"	Comparison of different monthly reports	I'm interested in proving what I expect to be true.
Data-mining	"What is interesting, what might happen?" "Tell me about relations and clusters in my data"	Advanced predictive modeling, to find out about next month Pattern discovery Classification Business rules	I'm interested in discovering new and interesting patterns and predicting new results from my data, to help me make better decisions.

Table 4 The differences between GIS, OLAP and data-mining (after Taylor and Hänni, 1998)

Figure 9 summarises the processes identified in data mining and knowledge discovery (Fayyed *et* al., 1996). Andrienko and Andrienko (1999) concluded that the actual data-mining step takes no more than 20% of the total workload in the "data-mining" process. This relative small step is best supported methodologically and by software (John, 1997), while the rest is considered more then an art as a science (Kodratoff, 1997).

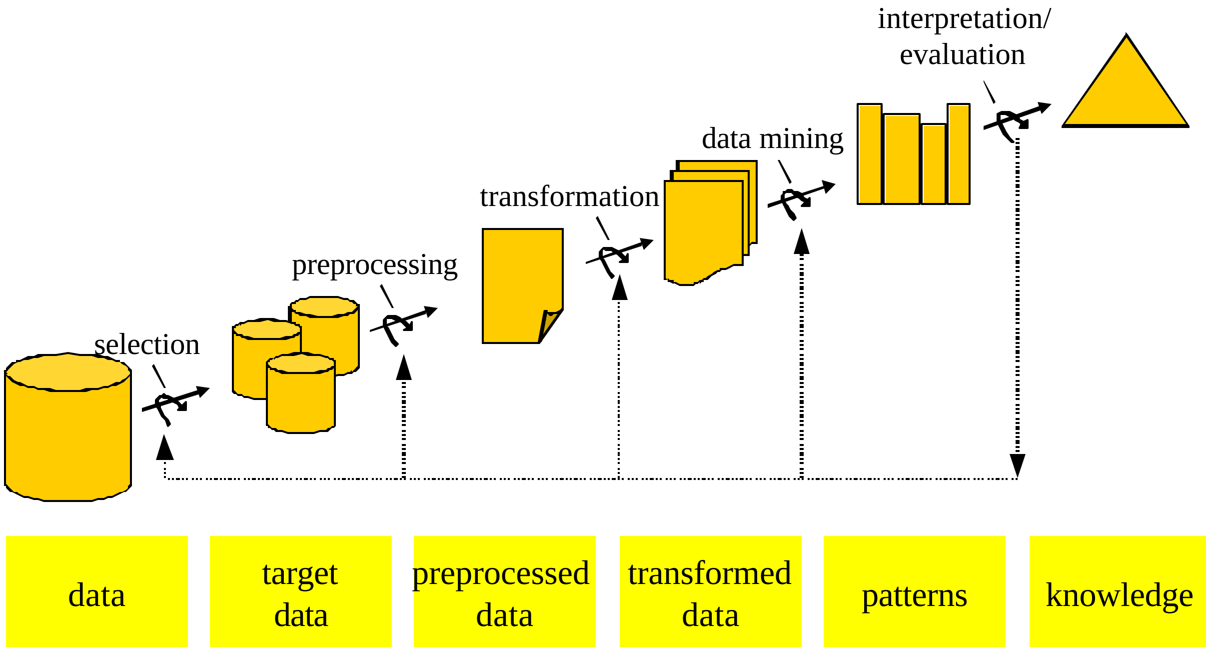


Figure 9 The data-mining process (after Fayyad et al. 1996)

These techniques emerged outside the GIS or spatial analysis but can well be extended to the 'map' and spatial data (e.g. Openshaw and Alvanides, 1999; Fisher, 1999; Andrienko and Andrienko, 1999). The NCGIA Project Varenius (Miller and Han, 1999) has recently brought together experts to develop and apply new techniques for exploring large and geographic datasets. The Varenius workshop on Geographic knowledge discovery (GKD) or the SPIN-project (http://www.ccg.leeds.ac.uk/spin/spin.html) are indications that more enhanced data analysis implementation will appear in the GIS-community in the near future.

These new information technologies are slowly integrated into monitoring systems, where in the past the emphasis was on datacollection techniques(Ward, 1997).

3.4.4. Extending the functions of analytic software

The computer is no longer part of the research environment, we are rapidly approaching a world in which the computer is the research environment. Most (spatial) statistical techniques, such as test for spatial autocorrelation and regression models are primarily static in nature, allowing only limited interaction between the data, the models and the analyst. In contrast dynamic or interactive approaches to data analysis stress the user interaction with the data in a graphical environment (Buja *et al*., 1996; Cleveland, 1993; Cleveland and McGill, 1988)

The current commercial GIS-user interfaces is characterized by the use of windows, icons, menus and pointing devices (WIMP-style interfaces). Their prevailing paradigm of interaction is that of querying, browsing a geographical database and presenting the results (after analysis) in maps and tables. The current development of multimedia technologies (with video, CD-ROM's and WWW), virtual reality systems will give uprise to new interaction and collaboration styles (e.g. interviewing, updating, experiencing) (Egenhofer and Kuhn, 1999).

3.4.5. Non-threatening statistics

Quotes about statistics like “Statistics are like bikinis; what they conceal is more important than what they reveal.” or “There are three kind of lies: lies; damned lies and statistics” are familiar phenomena. There seems to be a persistent thought that statistics are used to deceive and many people harbor a vague distrust of statistics as commonly used. One of the reasons is that statistics requires to consider things from a probabilistic perspective, using rather technical concepts as “confidence", “reliability’, “significance”. This is in contrast to the way people of cast problems; logical, concrete, often dichotomous (right or wrong, large or small, black or white, this or that). On the other hand there is a strong believe that numbers are, or at least should be, unquestionably correct. This awe is strengthened by the way people are exposed to math problems in school; there is a clearly defined method for finding the answer, and that answer is the only acceptable one. It comes then as a shock to many naive observers that statistics not always provide reliable (in the nontechnical sense) indicators of reality after all. And the logic goes, if statistics aren’t “right”, they must be “wrong”.

Exploratory data analysis (particularly the use of descriptive statistics and statistical graphics) are in many ways a substitute to perform “non-threatening statistics”. Though statistical, this approach is most of the time hypothesis free, making use of common sense, yet systematic, and “low tech”. Graphical, exploratory methods are felt to be more intuitive for non-specialist to use than methods of numerical spatial statistics. (Wise *et al.*, 1998) (see also chapter 4.4 for an overview of software).

3.4.6. Public participation

The new goals of integrated water management places much more emphasis on the need for water quality information to inform, and in the process, educate and motivate the public. More specifically, such information needs to describe, for example the impact of citizens individual (and collective) actions on water quality, as well as the overall status and trends in water quality in the community. In addition this information needs to be delivered in a way that is readily comprehensible to citizens (van Rooy, 1995)(see also chapter 2.7). This will require that water quality monitoring professionals view their purpose and tasks differently and develop a client-oriented approach. Many monitoring professionals, like scientists, have a way of writing that tends to hold them back from conclusive statements about the trends and status of the environment (Ward, 1997).

Public awareness about water issues is also reflected in the rapidly developing volunteer monitoring groups. Good examples of these initiatives can be found in the United Kingdom and Australia (Streamwatch)(http://www.streamwatch.org.au/) and E.P.A's volunteer monitoring (http://www.epa.gov/owow/monitoring/vol.html). Volunteers monitor all types of water bodies and collect physical, chemical, biological and habitat data. Because volunteer monitoring organizations can be strong partners in monitoring in the U.S.A. volunteer monitoring is integrated into existing and planned monitoring programs. (ITMF, 1995)

3.5. New paradigm

As a reaction to the shortcoming of the coherent linear design and the uprise of new elements there is a shift to a data exploration paradigm (illustrated in Figure 10). The terms 'data-driven' and 'technique-driven' are highly disreputable in science, but in society this paradigm is gradually becoming the new practice. With the vast amounts of data now available in digital form, this is a shift from the deductive approaches (generally using small amounts of data), towards a inductive approach where models are not so much hypothesized as observed form the data itself (Gahegan, 1999). From a pessimist, skeptics standpoint, this paradigm is a procedure, a technique or data in search of a problem, the return of the naive empiricism. For the optimist this paradigm shed light upon a wider range of social and environmental research problems (Goodchild and Longley, 1999).

Exploratory research implies the generation of questions followed by the inspection, creation, and interpretation of data looking for clues that could give possible answers and ideas. Once a question based on a vague idea or insight has been transformed into an explicit question, a scientist turns from a exploratory to a confirmatory mode of analysis. Confirmatory techniques, which allow researchers to confirm or reject hypotheses. The explorative paradigm is thus not a denial of the importance of confirmatory analysis, but a valuable extension. One of the great intellectual products of our century is that science has produced this ability to reject hypotheses about a population on the basis of a sample with a estimate of the probability of an erroneous choice (Tukey, 1997).

Exploration analysis is about generating many questions, finding initial answers and ideas, but also about discarding most (and maybe all) of them. Questions are designed to encourage thinking and learning, they do this by posing a problem which requires an answer (Nyerges and Golledge, 1997). Answering involve creatively integrating, rearranging, or manipulation bits of data (or information). In order to answer questions, exploratory analysis requires individual skills of observing, defining, classifying, analyzing, inferring, reasoning, integrating and associating phenomena (Slater, 1982). Since this skills can not be demonstrated on paper, in this part of the thesis only questions are posed.

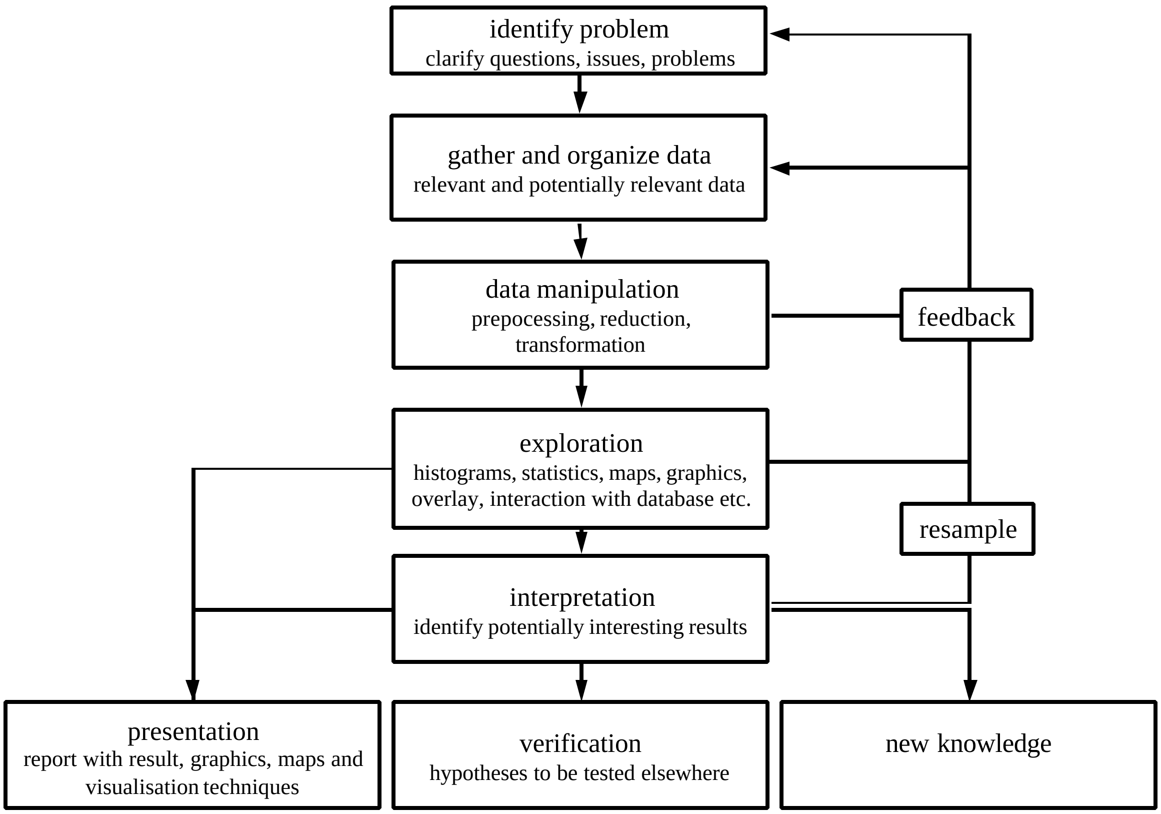


Figure 10 The exploratory (spatial) analysis paradigm (after Goodchild and Longley, 1999 and Openshaw and Clarke, 1996).

There is an underlying (usually not explicit) thought on positioning the explorative paradigm. This underlying model is illustrated in figure 11.

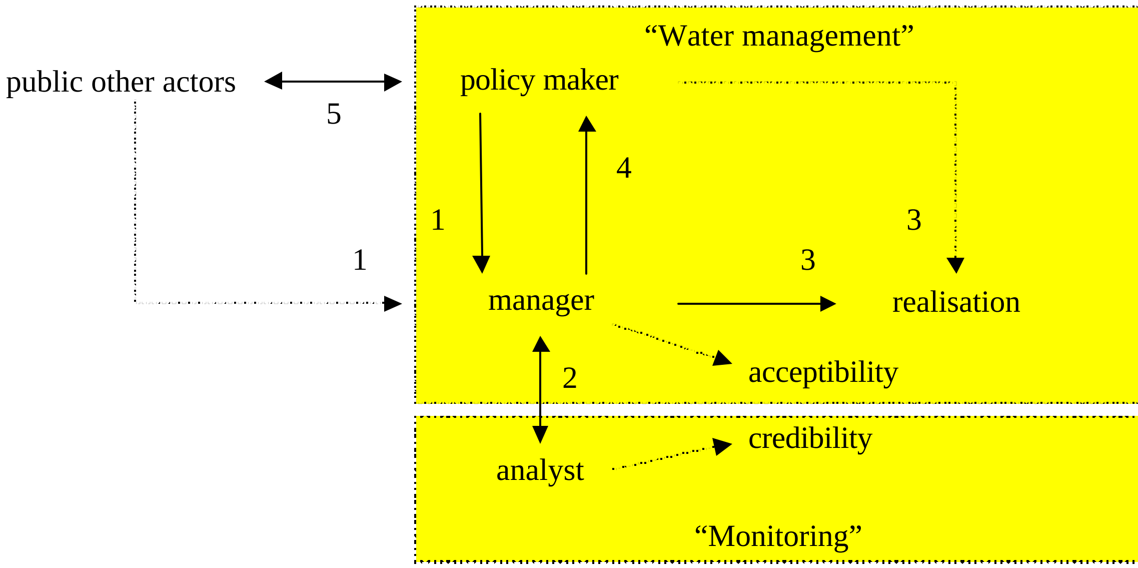


Figure 11 Positioning of credibility and acceptability and the actors in the explorative paradigm. The (operational) manager derive objectives form 'the field' or objectives may be imposed by policy makers (1). Next managers initiate an monitoring program (2). Results are put into effect (3) or passed on to policy makers (4), who may negotiate with the public (5) before implementation of the outcome.

Water management (like many other environmental) issues are complex problems and studying such issues is also dealing with uncertainties. A crucial point in dealing with this uncertainty is the distinction between 'credibility' and 'acceptability' (van der Molen, 1999). Credibility is the more technical appropriateness of the outcome (of a model ; van der Molen, 1999) and acceptability is the perception of the manager of its practical value. Credibility should be addressed by the analyst (the user), while acceptability is the responsibility of the audience (the manager). The distinction between acceptability and credibility plays an individual level ("a willingness to look for what can be seen", Tukey, 1977), a group level (van der Molen, 1989) and a organizational level (van Rooy, 1997) and beyond. Although the situation is a rather simplified model it represent common practice.

3.5.1. Explore the problem

What are central questions and issues? Is there ambiguity and vagueness of statement? Can the question, problems issues be restated in clear, precise, unambiguous terms? Are there elements of a question, problem, issue in need for further clarification? What are the areas in need of investigation? Can the type of questions, investigations be distinguished (e.g. direct/hypothetical, investigative/predictive, observational/experimental, exploratory/confirmatory)? What is the appropriate procedure to investigate the problem, question, issue? Is it likely there will be an answer? Is it worth to be investigated?

The first phase of every study is the most critical part of any study: a mistake at this stage can ruin the entire project. If a mistake is made during analysis, the data can be reanalyzed; if the whole project is not well-considered then the best analysis becomes problematic and at worst a "mission impossible". There is no single recipe for a well designed project, but a well designed project saves time, effort and money at every stage. The data collection is efficient, the analysis simple and the results emerge as an organic outcome of the study, self organizing, directly interpretable and simple to present.

If the study is exploratory then the emphasis is on isolating interesting patterns or trends. Since statistical tests can not be used to impress the prospective audience it is essential to discuss the objectives. Do the objectives match their requirements? Will they be interested, satisfied, convinced by the (explorative) results? Will the results be acceptable in the perception of the user(s)? (see also the distinction between "credibility" and "acceptability" in chapter 3.5).

There are innumerable relevant questions that should be asked at the first stage, but at the end it boils down to the final questions: Is it going to work? Is it going to be worth proceeding it? There is little point in continuing with a project that is unlikely to produce anything useful.

3.5.2. Gather and organize the data

What kind of data are required?
What are possible sources of data? Who has data?
What kind of data gathering techniques (e.g. sampling, surveys, remote sensing, inventories and so on) are useful?
What are appropriate data, should it be validated or "raw" data?
What is the agreement on coding on the data (and the meta-information)?
What is the content and the format of data exchange?
What is the (a priori) defined format of the data for analysis?
How should the data be stored (for later analysis)?

Once the problem and the objective of the project have been stated it is possible to develop a schema that summarizes accurately the procedures to collect data. In the linear project design the development of this schema is often called the 'experiment design' or the 'study design' and is conventially stated in terminology of statistics, but it basically involves choosing sample characters (parameters or variables to be measured), sampling units (study type and sampling strategy).

When the data are gathered and organized it is tempting to dive straight into the analysis. No matter how close the deadline, time spent checking and preparing the data can save lots of time (and embarrassment) later. The last step in gathering and organizing data (or first step of analysis) is checking the data visually. No matter how much attention is paid data sets of any size contain some errors in it. Generally the most disastrous errors are those that generate outliers, a slipped decimal point for example. These can often be detected by preparing the univariate displays (e.g. histograms, stem-and-leaf or boxplot) combined with maps (e.g. chloropleth or nominal point map). Not discovering any mistake is something to be worried about. Any data set of any size has some errors. Typos, observations missed out, observations repeated all happens somewhere. Tools for visual exploration (see chapter 4.5) are helpful in this stage to check the data visually. Simple univariate methods, histograms, frequency plots can often be useful to detect data that stand out. The chief problem is that most of the time the data are multivariate and spatial and it is virtually impossible to visualise the data cloud properly. At this stage it is worth to plot each pair of variables in turn and link this with a map.

3.5.3. Process the data (manipulation, exploration, interpretation)

What form or structure do the data have?
What is relevant (and unrelevant) data?
What basic information (spread, central tendency) can be retrieved from a single data source?
What less obvious information (trends, outliers) can be retrieved from a single data source?
What complex information can be retrieved from a single data source?
Which relationships (causal, chronological, geographical, concurrent etc.) can be detected?
What facts, opinions or speculation can be made can be made by comparing different sources?
What is the line of argument (especially where there is an unfamiliar and unconventional point of view)?
What unwarranted conclusions (e.g. assertions, inferences) can be detected from the data?
What warranted conclusions (e.g. predictions, trends, consequences) can be made from the data?
What hypotheses can be formulated to account for effects observed from the data?

It is important to display the data not only for the detection of errors and outliers but also to give insight in the data. Questions to be asked are: Is there any evidence for structure in the data, do any patterns emerge when you compare means between variables, or are there spatial clusters when comparing groups of observations? If the values of each variables are mapped do any patterns emerge? It may be that the patterns or results are so obvious that no sophisticated analysis is necessary. Subtle, truly multivariate spatial patterns will not emerge, but the grosser patterns and relationships could be apparent. Though before making conclusions always have second thoughts; what is obvious to a true believer may not be obvious to a skeptical audience. Make sure the results will be "credible".

In an exploratory study there is no objection using a number of similar techniques; if the same pattern emerges in all of them then there is greater confidence in the robustness of the results. Combining techniques, maps and graphics may improve the display of the results. By looking at the data from a number of viewpoints it is less likely to miss interesting patterns - though using techniques blindly it is also likely that some observed patterns are artifacts. Tools for visual exploration (see chapter 4.5 and 5) are useful in this stage to seek for patterns and trends by viewing the data in several different ways.

3.5.4. Reach and apply generalizations (presentation, verification, new knowledge)

What are common elements in the data?
What relationships in the data lead to generalizations?
What kind of generalization can be presented?
What conclusions can be made in relation to the questions/issues/problems?
What is the best way to present these conclusions?

In the explorative paradigm the main goal is to search for trends in the data, to detect patterns in the arrangements of the data values, to identify unusual data values. Many of the best results will not require statistical testing, the results are convincing ("credible and acceptable") without the crutches of test statistics and significance levels. In an other case the interesting patterns or trend lead hypothesis formulation from the data (see for an analogue with data-mining and knowledge discovery figure 9). These hypothesis can be validated with other data.

A second stage of investigations can follow up the explorative stage where the main aim is to estimate the size of the patterns or trends under investigation, not just to show that it is non-zero. Tools for visual exploration are in this stage used to communicate patterns or trends and have same function as traditional forms of graphics and maps (see table 7).

3.5.5. Re-evaluate (resample, feedback)

What factors could improve the process?
What limitations and qualifications can be recognized in the data?
Are there mistakes, faults, errors, missed observations in the data? What is the reliability of the data?
Where are limitations, deficiencies and gaps in the data and the used techniques?
What is inappropriate use of the data?
What factors may affect the accuracy?
What factors may introduce inadequency?
What could be artifacts of the techniques used?
Is there inconsistency in the data, arguments, conclusions?
How can be dealt with these factors?

Since statistically sound results are no recipe for "acceptability" analyst are not longer searching for the "golden truth". They must however produce a credible results by carefully pursuing the analysis. Uncertainties and errors pop up in all stages of the research. In the GIS-society there is an important focus on data quality (e.g. accuracy, precision, scale, bias, completeness, temporal consistency, logical consistency, semantic consistency, repeatability) (Trodd, 1996) or 'product quality' (Couclelis, 1991). Dealing with quality (and thus uncertainties) is a laborious, complex job, there is no uniform procedure and it is not possible to be complete and free of value-judgement when water management issues are concerned. Every (GIS-) action, from conceptualization of the data model to processing of the data through to output, has the potential to generate errors and compound existing ones (Trodd, 1996). In the explorative paradigm there must be "natural" tendency to be cautious and an sense when garbage is produced.

The explorative paradigm implies performing operations on data knowing that they are (in some cases) inappropriate or misleading and ending up with results that are partial, out of date, inappropriate, inadequate, or just plain wrong. The credibility of the results necessitates accounting for uncertainties and errors. The proper way to do this is to re-evaluate the results, making uncertainties explicit (van der Molen, 1999). An additional approach is to reconsider each aspect in the analysis, and to look for alternative procedures and feedback. This makes it possible to allow a fair consideration when acceptability comes into sight.

4. Maps as tools of exploration

4.1. Introduction

Looking at data to provide information is an old subject. Hieroglyphics from prehistoric times show that data analysis was important to early man. Abstract display of information (graphs, plots, maps) are a more recent innovation at around 1750-1800 (Tufte, 1985). Both Tukey (1977), writing about exploratory data analysis, and Tufte (1985) writing about visual presentation of information, stress the importance of showing the data. Visualisation is critical to data analysis (Cleveland, 1993). Visualisation in general scientific inquiry can take many forms, such as graphs, charts, tables, figures and images. In monitoring programs there is an emphasis to present data in tables, graphs and charts (Niederländer *et al.*, 1996, Ward *et al.*, 1990). Data interpretation is in many cases limited to performing descriptive statistics, calculation of trends and loads and testing for compliance with standards (Niederländer *et al.*, 1996; Ward *et al.*, 1990). In most monitoring programs the spatial content of the data is ignored and visualising data on a map somewhat neglected (Demayo and Whitlow, 1993).

However science and society are undergoing a profound change. Data exploration is getting a key issue of scientific endeavor (see also chapter 3). The most natural form to visualise spatial data is the map, which, in addition to showing the quantity and/or type of data, also shows the spatial distributions. Maps are getting tools of exploration (Kraak, 1998;1999; Andrienko and Andrienko, 1999).

This change is stimulated by a combination of technological developments and user-oriented developments (see chapter 3). First, technological developments in field such as databases, computer graphics, multimedia, internet, virtual reality and GIS, have boosted interest in graphics and stimulated (spatial) data presentation. From this perspective it appears that there are almost no barriers left. Second, user-oriented developments, often as an explicit reaction to the technological developments have stimulated scientific visualisation (McCormick *et al.*, 1987) and exploratory data analysis (Anselin, 1999; Kraak, 1999).

Result of these changes is that users have the ability to explore, manipulate, analyse and visualise data from different kind of sources, such as statistical databases and spatial databases. Users are trying to find answers to (spatial) questions and maps are important tools in this process. They are used to visualise spatial data, to reveal and understand spatial distribution and relations. Maps are now no longer the final products they used to be. Maps are becoming an integral part of spatial data handling (Kraak, 1998). This chapter will discuss how maps are becoming tools for exploration (that facilitate thinking, problem solving and decision making). The key concepts ("(scientific) visualisation", "geographic visualisation", "exploratory data analysis") will be explained in this chapter. This chapter ends with a list of software developments on the area of "visual spatial data exploration".

4.2. Scientific Visualisation

Viewing the data helps to gain understanding of the data and is an important step towards interpretation and understanding. Visualisation is, like a telescope or microscope, a tool to power visual thinking (DiBiase, 1990) or visual reasoning (Clarke, 1999).

While most scientists would agree that "visualisation" has become a scientific practice, they often disagree on the definition (Petch, 1996). Terms as visualisation (with an "s"), visualization (with an "z"), ViSC (visualisation in scientific computing) visual analysis, visual representation and visual display are often confused and confusing. Visualisation seems to have at least three meanings (or aspects): the production of (carto)graphical representations of data, the use of (advanced) display technology that acts as an interface between data and user, and the production of cognitive (mental) representations (Petch, 1996).

In the most general sense all these terms have the meaning of "to make visible". McCormick *et al.* (1987) defined visualisation as "a method of computing .. a tool both for interpreting image data fed into the computer, and for generating images from complex multi-dimensional data sets .., " the goal of which is " ..to leverage existing scientific methods by providing new insight through visual methods". Scientific visualisation as the "use of the human visual processing system assisted by computer graphics, as means for the direct analysis and interpretation of information" (Friedhoff and Benzon, 1989). Some scientist counter that visualisation is not only a method of computing. In contrast to visualisation as the creation of concrete displays (whether via computer or otherwise), the term can also refer to making visible in the sense of mental images (MacEachren, 1995, p356). Visvalingham (1994) distinguishes visualisation (with an "s") as the mental process of prompting visual images from visualization (with a "z") as the sophisticated computing technology to create visual displays that facilitate thinking and problem solving. In this thesis the treatment of visualisation is catholic, it is seen as an unity of issues. However the cognitive aspects of visualisation will not further get attention in this thesis (see chapter 1.4).

It is important to note that the aim of (scientific) visualisation is not to analyse the data per se, but rather to present the data to the user in a way that promotes the discovery of inherent structure and relationships (Gahegan, 1998) or to promote visual emphasis (“pop-out effects") (Csinger, 1992).

4.2.1. Geographic Visualisation

Over the past decades considerable attention was devoted to understanding how presentation maps work (what Freitag (1993) terms the "*communication function*" of maps). Relatively little is known about how maps that facilitate thinking, problem solving, and decision making work (uses that, according to Freitag's (1993) typology, are representative of the "*cognitive function*" and the "*decision support function*" for maps), nor what the implications of "working" are in these contexts (implications associated with Freitag's (1993) "*social function*"). It is to the cognitive and decision-support functions that much of the new geo-information technologies (e.g. GIS) are directed, particularly those maps with dynamic and interactive components. It is also in these functions that scientific visualisation and cartography share the greatest overlap - an area that has been labeled "Geographic Visualisation" (see MacEachren and Monmonier, 1992; MacEachren, 1994; MacEachren and Kraak, 1997).

Following publication of the McCormick (1987) report on visualisation in scientific computing, several cartographers took up the challenge of implementing this new (or renewed) reliance on visual representation in science. DiBiase (1990) borrowed from the exploratory data analysis (see chapter 4.3) literature of statistics to propose a graphic model of stages in map-based scientific visualisation. The model presented visualisation as a four-stage process consisting of two “private visual thinking” stages (exploration and confirmation) and two “public visual communication” stages (synthesis and presentation) (Figure 12). From the investigator's point of view, graphic change from reasoning tools to communication tools as an investigation expands from a private to a public endeavor. The downward slope of the curve suggests that visual thinking involves higher-order cognitive tasks (for the investigator) than visual communication does.

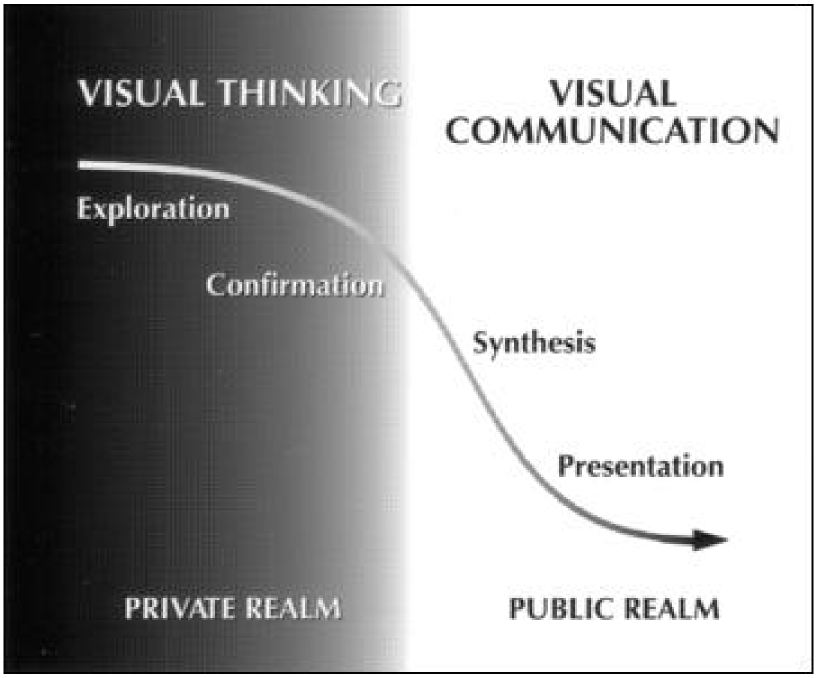


Figure 12 The functions of graphics in the research sequence. (taken from www.geovista.psu.edu/publications/dibiase90/swoopy.html

This model focused on the need to direct attention to the role of maps at the early (private) stages of scientific research where maps and map-based tools are used to facilitate data sifting and exploration of extremely large data sets (table 5 and table 6).

Private visual thinking	Public visual communication
refers to the situation where graphics/maps are used to prompt insight, look for patterns in data, highlight anomalies; any situation where you have more data than ideas or insight. Private, informal, focus on many different views of same data are important elements	refers to the situation of using graphics/maps to make a point, to communicate a message and to let people know something via a graphic form; any situation where you know what you want to express and want or need to use graphics to express it. Public, formal, focus on optimal view of data are important elements

Table 5 Overview about "private visual thinking" versus "public visual communication".

1. Exploratory Uses of Graphics:	private, informal, hypothesis formation, questions raised; focus on devising graphic methods which assist in exploratory graphic use; highly interactive; what if scenarios
2. Confirmatory Uses of Graphics:	graphics used to confirm or counter patterns, anomalies, ideas, etc. found when exploring data or ideas
3. Synthesis Uses of Graphics:	graphics used to pull together data and ideas that are in essence "confirmed" or that you are sure about but need to express in some meaningful manner

4. Presentation Uses of Graphics:	once you have synthesized your data and ideas you need to make many basic design and layout decisions which will help communicate what you want to express to others.

Table 6 Overview of strategies of visualisation by DiBiase (1990). Confirmation and synthesis are sometimes summarized as analysis (Kraak and Ormeling, 1996).

Considering these four different fields of visualisation it can be noticed that tools for synthesis and presentation are the most highly developed in most GIS-systems (Robinson *et al.*,1995) and tools for exploration are poorly available (Goodchild, Haining and Wise, 1992; Anselin and Getis, 1992; Anselin, 1999) .

Building from the perspective on georeferend/cartographic visualisation outlined above, a conception of visualisation emphasizing use of visual displays was developed (MacEachren, 1994). This continually evolving map use-based approach to visualisation has been a driving force behind creation of the ICA Commission on Visualisation (see also http://www.geog.psu.edu/ica/ICAvis.html) and the initial research activities initiated by that commission.

The approach treats map use as a "space" (referred to by MacEachren (1994) as [CARTOGRAPHY][3]; a reference to the three axes) along which map use was characterized (Figure 13).

In this map use cube private versus public map use is just one of the axes. Along the other axes the revelation of the unknown versus presentation of the known, respectively high versus low interaction are plotted.

In this space, visualisation is considered to be the complement of communication. All map use involves both visualisation (defined loosely as the prompting of visual thinking) and communication (defined loosely as the transfer of information), but map use can vary considerably in which is emphasized. The axes of the use space are delineated as private versus public, high interaction versus low interaction, presenting knowns (i.e., simple information retrieval) versus revealing unknowns. Past communication-oriented cartographic research has emphasized the use of static maps designed for public consumption with the emphasis on extraction of specific pieces of information (e.g., research on communication effectiveness of textbook or topographic maps). The sequence of goals that can be facilitated by visualisation methods emphasis on distinguishes among use goals that may different visualisation strategies.

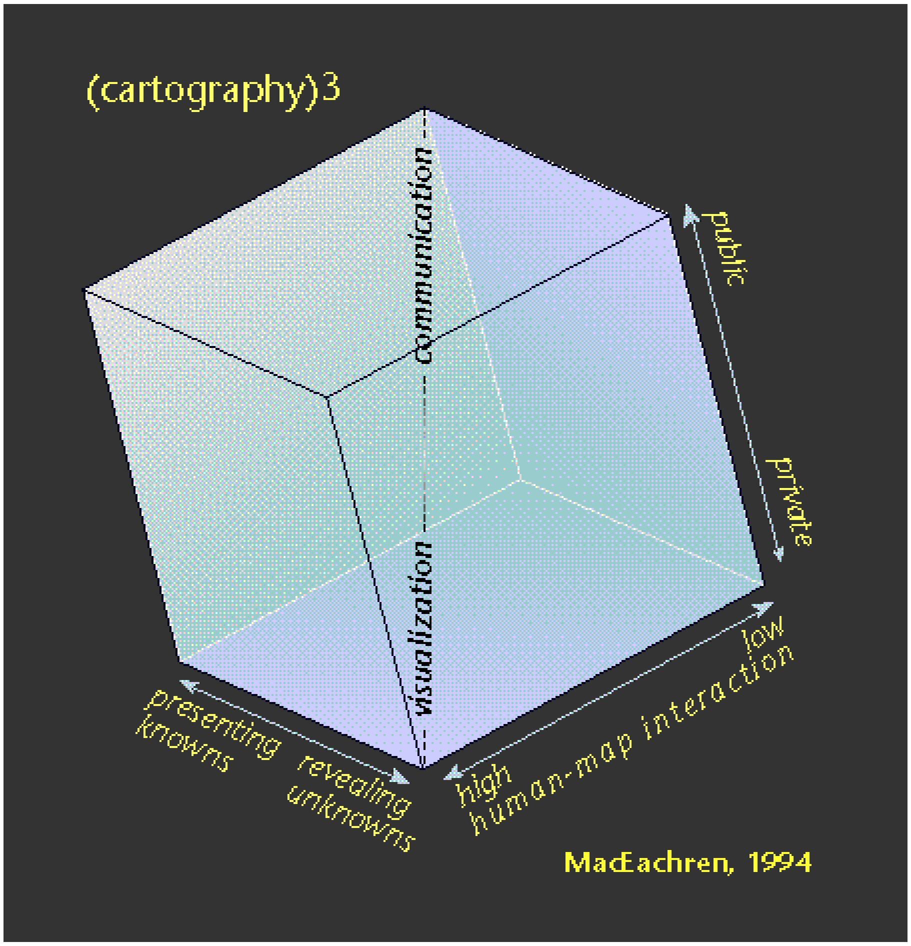


Figure 13 [Cartography]3 - a representation of the "space" of map use and the relative emphasis on visualisation and communication at various locations within this space. This presentation deals, not with kinds of maps, but with the kinds of map use. (http://gis.psu.edu/MacEachren/MacEachrenHTML/MacEachrenVIS.html

Work on visual data exploration tools for spatial data has taken place largely within the framework of statistical and cartographic software, independently of Scientific Visualization community and software such as EVS (Environmental Visualization System; http://www.evs.com) or AVS (Advanced Visual System; http//www.avs.com). This may be in part due to the way which visualisation of the type of spatial data has grown out of initial concern with the lack of spatial data analysis facilities in GIS systems. Whatever the reason it would be interesting to see whether VISC products could in fact be used for the visualisation of spatial data (Wise *et al.*, 1998). Some initial research in this area has taken place by Gahegan (1999), Rhyne (1996) and Hogeweg (2000).

4.3. Exploratory data analysis

The visualisation of (statistical) data has long been existing in one form or another in science and technology. Since the pioneering work of Tukey (1977), exploratory data analysis (EDA to its advocates) has gained considerable influence as a paradigm in applied statistics and it now forms the basis for many of the visualisation and graphical features of modern statistical software (Good, 1983; Cleveland, 1993).

EDA is more of a philosophy (or an art) than an analysis, ”exploratory data analysis is an attitude, a flexibility, and a reliance on display, NOT a bundle of techniques” (Tukey, 1980, p23). The approach is based on a “recognition that the picture-examining eye is the best finder we have of the wholly unanticipated. Clearly by relying on such a (over)sensitive pattern recognition machine, EDA is aimed at the generation of hypotheses, not their test, to identification of the unexpected, not the confirmation of the already known” (Tukey, 1980). To a large extent this chapter is concerned with methods that make spatial data available to that “picture-examining eye” ; techniques that allow the data to be examined, patterns or relationships detected and ideas generated. The techniques discussed are sophisticated and sometimes difficult to comprehend. The spirit of EDA, on the other hand, is simplicity. Simple methods, such as scatterplots, histograms, boxplots and the like, can often be useful in gaining insight into the data (e.g. Maasdam and Smith, 1993). The focus in this thesis will be on these simple methods. Such simple graphical methods are certainly more useful than tables of (summary) data, and are a lot easier to read. The books by Tufte (1985; 1990) and Cleveland and McGill (1988) give an useful review of the principle ideas and these simple (and some not so simple) data presentation methods.

As opposed to traditional hypothesis testing designed to verify a priori hypotheses about relations between variables, exploratory data analysis (EDA) is used to identify systematic relations between variables when there are no (or not complete) a priori expectations as to the nature of those relations (Tukey, 1977). In a typical exploratory data analysis process, many variables are taken into account and compared, using a variety of techniques in the search for systematic patterns.

4.3.1. Interactive graphical data analysis

A still more recent innovation is the concept of interactive displays of information, displays which users can change with a minimal effort and small latency. These interactive graphical tools have only recently been realizable, as they require the use of a computer. Now with the advent of powerful desktop machines capable of rendering tens of thousands of symbols in a fraction of a second, these techniques can be used by a wide range of people for a variety of tasks. Interactive Graphical Methods (a.k.a. Direct Manipulation Graphics or Dynamic Graphics) as the class of techniques for exploring data that allow the user to manipulate a graphical representation of the data. Becker *et al.* (1987) defines the process thus: "the data analyst takes action through manual manipulation of an input device and something happens on the screen. These computing capabilities provide a new medium for the invention of graphical methods for data analysis."

Various interactive manipulation techniques for the visual data exploration were proposed in the area of statistical graphics (Cleveland, 1993). The most widely known idea of visual linking of several graphical displays by means of "brushing". This is when selection of an object in a graph (or map) automatically highlights the corresponding element in the other graphics. Depending on the view in which one selects the object, there is geographical brushing (clicking in the map), attribute brushing (clicking in the diagram or table), and temporal brushing (clicking on the time line) (Kraak, 1999). Most of today's spatial data viewing software has elaborated on these principle (see also chapter 4.4 for an overview).

4.3.2. The spatial element in Exploratory Data Analysis

Exploratory Data Analysis has become increasingly popular as a methodology to generate insight into patterns and associations in data (especially large data sets), without strong prior assumptions and taking into account the potentially misleading influence generated by "extreme" or "atypical" observations. However, none of the traditional tools of EDA are especially geared to dealing with spatial data, in the sense that the effects of location, spatial dependence and spatial heterogeneity are ignored. Anselin (1994) distinguishes several uses of EDA to spatial data. The first is non-specialised EDA to spatial data (see also Anselin and Getis, 1992). This is were standard EDA (such as scatterplots or boxplots) are applied to spatial data (e.g. exporting information form a GIS to a statistical package). Spatialised EDA (Anselin, 1994;1999) is one step further in the sense that location is combined with a graphic description of the data in the form of a bar chart, pie chart, or various icons.

In contrast, Exploratory *Spatial* Data Analysis (ESDA), focuses explicitly on these spatial effects and consists of techniques to describe spatial distributions, to identify atypical locations (spatial outliers), to discover patterns of spatial association (spatial clustering) and to suggest different spatial regimes or other forms of spatial instability (spatial non-stationarity) (Anselin, 1994). Central to ESDA is the concept of spatial autocorrelation, that is, the phenomenon where locational similarity (observations in spatial proximity) is matched by value similarity (correlation). Spatial autocorrelation has been conceptualised from two main perspectives, one prevalent in physical science (the so-called geostatistical perspective, by means of the variogram), the other in the social sciences (the so-called lattice perspective, by means of spatial weight matrix). Table 7 gives an overview of ESDA techniques, see Cressie (1993), Anselin (1999) or Getis (1999) for review of these techniques. Recently there is also a focus on detecting local rather then global patterns of association (e.g. "local indicators of spatial analysis" or LISA) (Anselin, 1999). These techniques focus on local patterns of spatial association, indicating local non-stationarity and discovering islands of spatial heterogeneity (Anselin, 1995; Cressie, 1993).

In this thesis there is limited on the combination of EDA and spatial data, without the explicit exploratory spatial data analysis, so ignoring the typically complicating effects of spatial autocorrelation.

GIS has not been developed with ESDA in mind but rather in terms of data management (e.g. AM/FM applications), cartographic modelling (e.g. whole map operations such as sieve analysis to locate areas) and some selected forms of spatial analysis (e.g. network analysis). Nonetheless GIS does contain some E(S)DA-type facilities (Godchild, Hailing and Wise, 1992; Anselin, 1999). Exploratory (spatial) data analysis is one of the areas where statistical spatial data analysis techniques can strengthen current GIS practice (Getis, 1999; Anselin 1999). The clear need for a quantitative exploratory style of analysis which can complement the map-orientated nature of GIS, is especially due to the data rich and theory poor GIS environment.

	Geostatistical perspective	Lattice perspective
Visualising spatial distribution	• Spatial cumulative distribution function	• Box map • Regional histogram • Spatial exploratory analysis of variance
Visualising spatial association	• Spatially lagged scatterplot • Variogram cloud plot • Variogram box plot	• Spatial lag charts • Moran scatterplot and map
Local spatial association	• Outliers in variogram box plot • Outliers in variogram cloud plot	• LISA maps • Outliers in Moran scatterplot

	Geostatistical perspective	Lattice perspective
Multivariate spatial association	• Multivariate variogram cloud plot	• Multivariate Moran scatterplot

Table 7 Overview of ESDA techniques

4.4. Software for visual spatial exploration

The development of visual spatial exploration applications has its roots in the dissatisfaction expressed by a number of authors about the poor (spatial analytical) tools available in GIS (Goodchild ,1987; Burrough, 1990; Goodchild, Haining and Wise, 1992). One might assume that GIS, specifically designed for the storage, manipulation and display of geographic data, would provide a useful environment within to explore the data visually. But unfortunately in current GIS, the number of visual exploration options under the direct control of the user is small. It can be argued that most GIS still maintain the characteristics of traditional maps and graphics: less interactive, less dynamic, less explorative (see table 8) as one might want. Today a fully (interactive) visual spatial exploration functionality is not yet part of any commercial GIS.

Traditional forms of graphics and maps	New forms of graphics and maps
Static and permanent	Dynamic and temporary
Aimed for presentation	Aimed for exploration or understanding
Are selective	Try to use as many data as possible
Demonstrate the known	Detect the unknown
Used many times	Used once (or twice)
Restricted dimensions (x,y,z,time)	Multidimensional
One view of the data	Multiple view of the data
Intended for many viewers	Used by one person

Table 8 Comparison of traditional and new forms of graphics and maps. Today's GIS maintain most of the traditional forms of graphics and maps.

A number of researcher began to explore the possibilities of building functionality focuses on the potential of exploratory methods to spatial data. Here the aim is to use a range of largely graphical or map-based methods in order to explore data sets and suggest hypotheses, but without the need of model building or formal testing. Table 9 contains a list of the software developments. Some of this software have been developed from the ideas of E(S)DA and have a more statistical approach, other software developments have been developed from the ideas of visualisation of spatial data and implemented more interactive graphical methods. Table 9 summaries the main characteristics of this new approach to graphics and maps.

Package	Reference	Description
Spacestat	Anselin, L and S. Bao (1997)	SpaceStat is principally designed for spatial modelling, but includes a visualization capability through an ArcView extension (and to IDRISI). The tool of Spacestat emphasis on the spatial component of ESDA in the sense that spatial dependence of the data is taken into account explicitly.
Infomap	Bailey T.C. and Gatrell A.C. (1995)	INFOMAP. Pc-based mapping and Spatial Analysis package distributed on disk with textbook on spatial analysis.
	Brunsdon C. and M.E. Charlton (1996)	Prototype spatial analysis system written using public domain software system Xlisp-Stat. http://www.ncl.ac.uk/~ngeog/GWR/
Xgobi	Cook, D. J.J Majure, J. Symanzik and N. Cressie (1996) Symanzik at al. (1994). see for a overview of the Xgobi environment, Buja *et al.* (1996).	Link between ArcView GIS and XGobi graphics system. Has ability to handle both vector and raster data in graphical windows linked to statistical graphs. This approach focuses on the use of XGobi for exploratory data analysis , such as visualization and brushing of scatterplots and cumulative distribution functions. This effort is an extension of the one-directional close coupling of XGobi and ARC/INFO. Here information from Arc/INFO GIS is efficiently passed to the Xgobi software for exploring multivariate data by means of dynamic graphics, brushing, linking and the grand tour.
CDV	Dykes, J. (1996), Dykes, J. (1997a), Dykes, J. (1997b)	CDV (Cartographic Data Visualiser) is principally an map visualisation toolkit with modest statistically capability. The bulk of CDV's facilities consists of graphical tool for viewing data, including a range of mapping options (chloropleth maps, point symbol maps and cartogram) and some additional statistical graphs (dotplot, histogram and scatterplot). Most of the facilities are aimed at the exploration of non-spatial elements of data. Exploration of spatial patterns is done by the linking of the graphs with the map, rather then by specialist spatial graphs or tools.
SAGE	Haining R.P., Wise S.M. and Ma J. (1996)	SAGE adds ESDA capability to the ARC/INFO GIS, also includes spatial modelling capability so that ESDA can be applied to raw data and model residuals generated with the package.
SPIDER REGARD	Haslett J., Wills G. and Unwin A.R. (1990)	One of the key pieces of software that implement dynamic graphics for exploring spatial data is contained in the Regard (formerly Spider), and was a forerunner of MANET. Runs on a Macintosh platform and demonstrated how highly visual exploratory methods could produce useful insights into spatial data.
	MacDougall E.B. (1992)	Polygon explorer. Mac-based software for map display and analysis.
MANET	Unwin A., Hawkins G. Hofman H. and	MANET (acronym of Missings Are Now Equally Treated) build on the earlier, innovative packages of SPIDER and REGARD. MANET is

Package	Reference	Description
	Siegl B. (1996)	highly visual and allows for the visualisation of the distribution and associations between the data for any subset of locations selected on a map display. The central objective is to provide the user with a suite of tools for exploring multivariate spatial data and dealing with missing values. While highly dynamic in its statistical graphics, the SPIDER-REGARD-MANET approach is still somewhat limited in the terms of the spatial aspects of the data. They are based on a fixed map and do not take advantage of the "GIS-functionality" such as specialized data models to facilitate spatial queries and overlays.
	Nagel M. (1996)	System built using data analysis language ISP. Includes a range of map types (chloropleth, isoline) and spatial analysis techniques.
	Batty M. and Xie Y, (1994)	Extension of ARC/INFO with non-spatial EDA tools, such as scatterplots.
SPLUS, S+SpatialStats	Mathsoft (1996a)	A commercial implementation that exists to date of on integrated data analysis and GIS environment, the S+Gislink between the S-plus statistical software and ARC/INFO or ArcView. The linkage allows users to call S-Plus statistical functions within ARC/INFO or ArcView. The S-Plus –ArcView is limited to traditional non-spatial EDA, S+SpatialStats also contains some spatial exploration possibilities.
(IRIS)-Descartes, DialoGIS	Andrienko G.L. and N.V. Andrienko (1998)	Descartes (earlier called IRIS) is a software system designed to automated presentation of data on maps and offers interactive facilities to dynamically manipulate the maps and supplementary graphic displays linked with them. Descartes runs as a Java applet on standard WWW browsers. The DialoGIS extension for ArcView is the commercial version of Descartes.

Table 9 Software developments for visual spatial exploration.

Apart from the self contained (integrated) approach in SPIDER-REGARD-MANET and CDV, most implementations are extensions of existing GIS-systems by means of macro-language scripts. On the other hand several ideas from the methodology of dynamic and interactive statistical graphs are reflected in the design of current GIS and mapping software. For example the ArcView GIS (ESRI, 1996) is organised around several linked 'views' of the data (a map, a table and several types of charts) (see figure 18) This allows users a limited degree of dynamic interaction in the sense that a selection made in any of the views is reflected in all other views. In the case-study of this thesis (see part 2, chapter 5) ArcView and CDV are used to demonstrate the potentials of such dynamic, interactive and explorative approach.

An other promising research direction to develop tools for interactive spatial data analysis is the use of the Internet to facilitate interactive mapping and visual data exploration. The Web Cartography Forum (Blok and Köbben, 1998) give an overview of the activities going on in this area. A promising development is the CommonGIS project (Perdigão, 1998). The key-thought of this is "GIS for everyone, from everywhere", by concurrently applying and integrating several technologies - WWW, mobile-code, GIS and expert systems. For comprehensive data analysis, problem solving and decision making the user needs complementary views on the same data, the opportunity to compare several maps and to access source data through maps, and rich facilities to manipulate data and representations. CommonGIS aims to provide this requirements (Perdigão, 1998; Andrienko and Andrienko, 1999). Current off-the-shelf GIS systems still require the special cartographical knowledge for their design and interpretation (Andrienko and Andrienko, 1997). Current GIS systems ask users to provide commands to get a good visualisation. This is not appropriate for the common user ("viewers") who have no GIS knowledge. The use of knowledge based support of visualisation may be a future, especially appropriate for common users. The users may be provided with a simple icon so they can focus on interacting and exploring the data. Andrienko and Andrienko (1999) have provided with their Descartes evidence that the use of knowledge based technique is efficient in visualisation. Descartes produces high-quality thematic maps for statistical data selected by the user. The system applies cartographic grammar to present statistical data in thematic maps.

PART 2

5. Regional Water System Report in Friesland: a case of the explorative paradigm

5.1. Introduction

This chapter gives a case-study of Regional Water System Report in the waterboard of Marne-Middelsee (province of Friesland, The Netherlands), with a special focus on salination. This case-study is an example application of the explorative paradigm (see Figure 15). In this case-study ArcView and Cartographic Data Visualiser (CDV) are used to illustrate dynamic visualisation within explorative analysis.

5.2. Context of the case-study

Human activities exert a strong pressure on freshwater resources (Saeijs and van Berkel, 1995). This has led to increased demands on policy and decision makers at (inter)national, regional and local levels to develop well-fundamental strategies and solutions. In water management and policy there is an increasing need for assessment methodologies for diagnosis and prognosis purposes in which an integrated watersystem approach is considered (Witmer, 1995; van Rooy, 1997). Such assessment methodologies should aggregate operational (monitoring) data to a comprehensive, strategic, preferably simple quantitative form, to support the policy and decision making process. The "Dutch Aquatic Outlook" (WaterSysteemVerkenning; WSV in Dutch) (Luiten,1995) and "Regional WaterSystem Report" (Witmer, 1995; Witmer, 1997; IPO, 1996) are two methods to provide such a systematic and quantitative exploration of the Dutch water systems.

In this case the Regional WaterSystem Report which is developed by provinces (and waterboards) is illustrated, for the watersystems in waterboard Marne-Middelsee (province of Friesland, the Netherlands) (see Figure 18). Large areas of the watersystems, which are situated behind the dykes of Waddensea and Lake IJsselmeer are effected by seepage of saline water. This salt water effect the conditions for agriculture, on the other hand this salt is the basis for brackish nature reserves. In this case the issue of salination is used to illustrate RWSR (see Figure 20).

Chapter 5.3 gives a description of the methodology. It gives both an overview of the methodology of Regional WaterSystem Report as an overview of the tools in this case-study. The rest of the chapter will follow the structure of the explorative paradigm (chapter 3.5).

5.3. Methodology

5.3.1. Description of Regional Water System Report Methodology

Regional WaterSystem Report is a method developed for the monitoring and evaluation of the water management on a regional scale. It concerns all aspects if water management, surface water and groundwater, quantity and quality of water, morphology and maintenance of watercourses and biological assessment (IPO, 1998) .

In essence the methodology follows the following steps (see Figure 14):

1. The boundary of watersystems are defined according to the national guidelines (CIW, 1998);
2. For each watersystem is being determined what formal functions (e.g. swimming, agricultural use, drinking water etc.) there are in national, provincial and waterboard policy plans. For each of the functions the objectives are being determined;
3. The objectives are being translated into measurable indicators with matching standards;
4. In each watersystems measurements/inventories are carried out to investigate the present state;
5. The measuring data are transformed into testdata which can be compared with the standards;
6. Testdata and standards are tested using yardsticks to get indicatorvalues ;
7. Indicatorvalues are clustered, aggregated, integrated to get an overall score according to compartment, function and watersystem;
8. The overall score are presented in tables, diagrams (see for instance Figure 16 and Figure 17) and on maps (see Figure 18);
9. Results are reported (and evaluated) in a watersystem report.

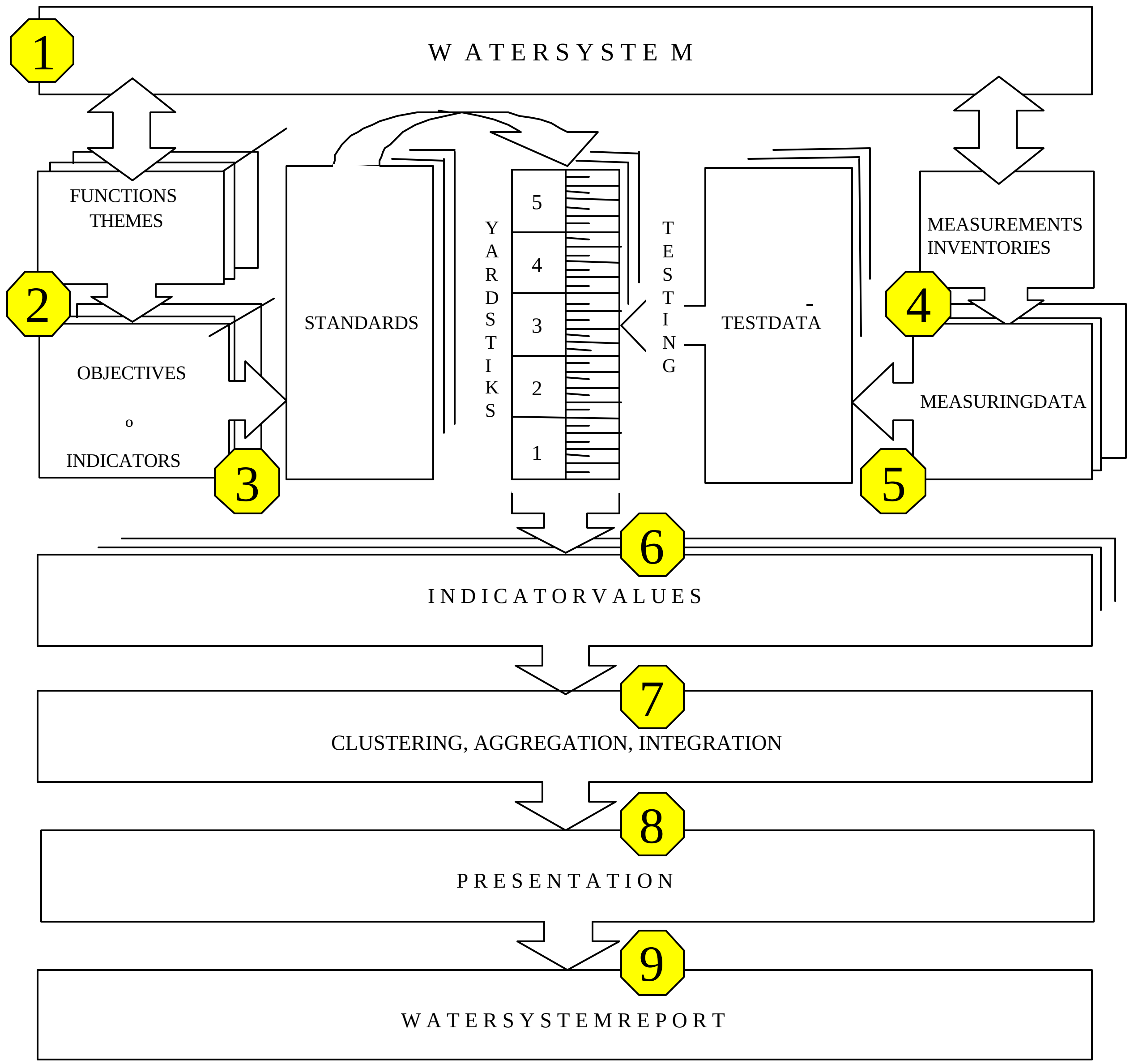


Figure 14 Graphical representation of the Regional WaterSystem Report methodology. Numbers refer to the steps in the text.

Since the watersystem report is presented here as an example of the explorative paradigm (discussed in chapter 3.5 and illustrated in Figure 10), the rest of this chapter will have the structure of this paradigm. In Figure 15 the sequence of steps of the Regional Water System Report is structured according to the explorative paradigm.

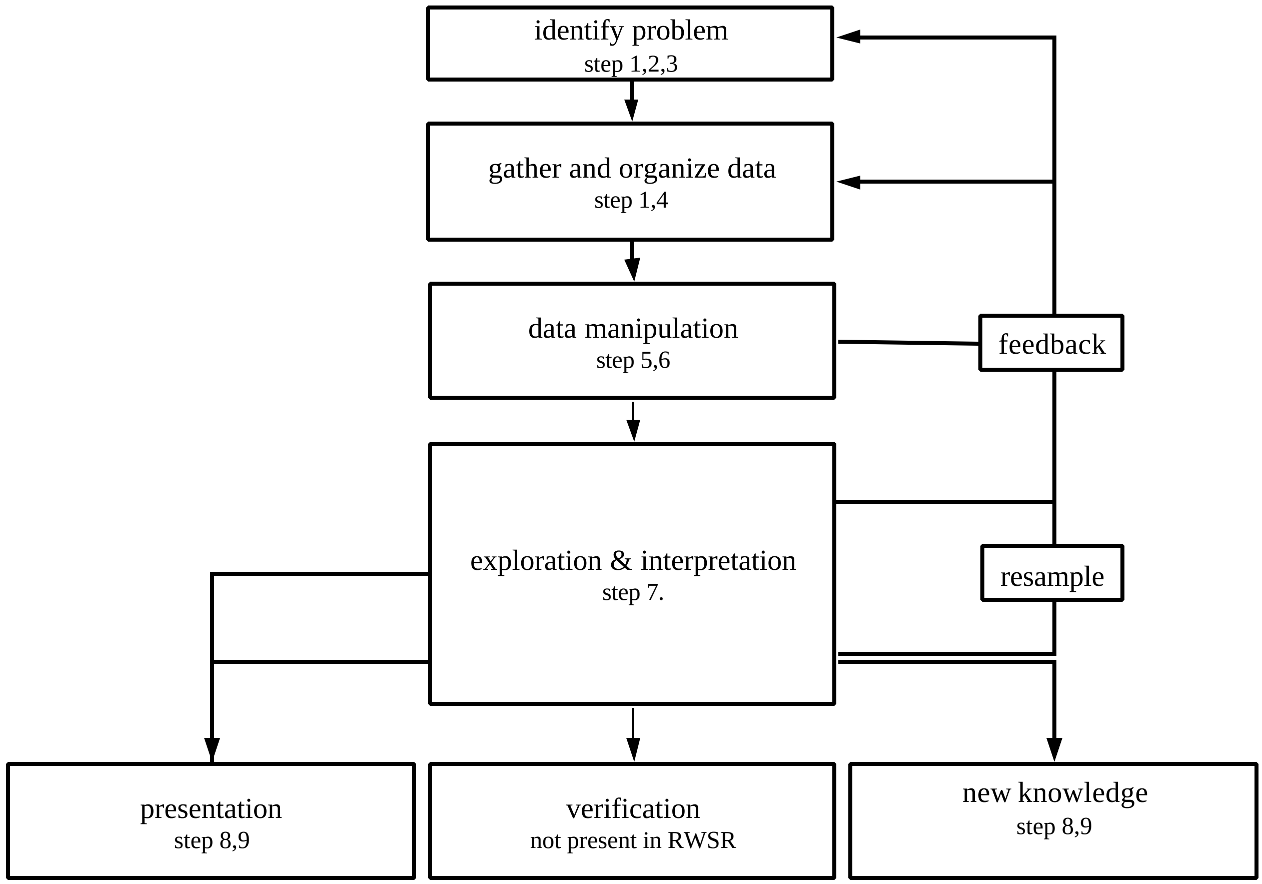


Figure 15 Regional WaterSystem Report (RWSR) as an example of the explorative paradigm.

5.3.2. Tools for exploration

In the Aquatic Outlook project an information system was introduced called "The WATER DIALOGUE" (Latour *et al.*, 1996). This information system was developed to retrieve the great diversification of data required for integral water management in an uncomplicated manner. The application provides processing facilities and a simple presentation of data relating to the condition and use of water systems in the Netherlands. In addition to information regarding the physical, chemical and biological quality of the waters, WATER DIALOGUE also provides input indicators, the costs involved in the water policy and information with respect to the functional aspects. The WATER DIALOGUE has been implemented as an interactive PC application, which enables the user to present results of analysis using a number of graphical methods (e.g. Radialplots, Waterindexplots, WaterMondriaan). The WATER DIALOGUE lacks geographical presentation beyond the highly abstract WATER-MONDRIAAN (see Figure 16). Within the Regional WaterSystem Report a "regional watersystem picture" has been introduced as a method to present the results (Van der Straten et al., 1998).

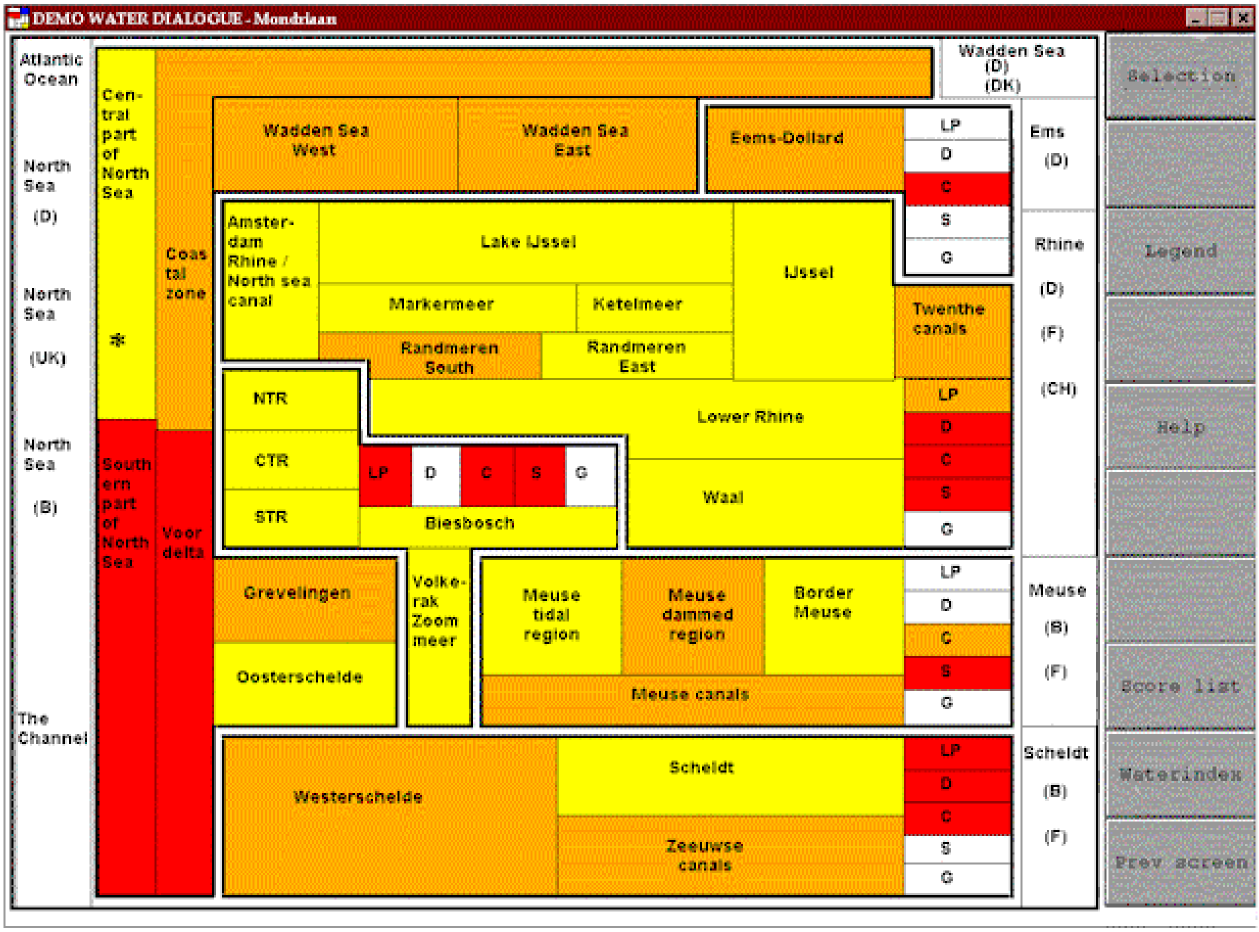


Figure 16 *Water-Mondriaan. In the Water-Dialogue a schematic representation of the map of the Netherland is present: The Water-Mondriaan. This graphical presentation shows the most important watersystems and watertypes in each box. The colour of each box represents the state of the watersystem or watertypes under consideration.*

This new "picture" is a good way of presenting the results for one watersystem but lacks the possibilities to compare watersystem in a view and interact with the information. Because RWSR uses lot of spatial information (e.g. watersystems, functions, measuring point) from the start of the project it was the intention to use GIS (NITG-TNO, 1997a, 1997b). The GIS-module (of the RWSR informationsystem) should offer functionality for geographical selection of data/indicatorvalues and clustering, aggregation, integration facilities to present information on a map (IPO, 1999). The province of Gelderland has produced an ArcView-project for their pilot-project RWSR (Provincie Gelderland, 1999). This ArcView-project is the prototype for IPO-RWSR to realise an GIS-module (ESRI Nederland, 1999). This Gelderland prototype is used in this case-study to demonstrate the use of maps in Regional WaterSystem Reports (see Figure 18 for an overview).

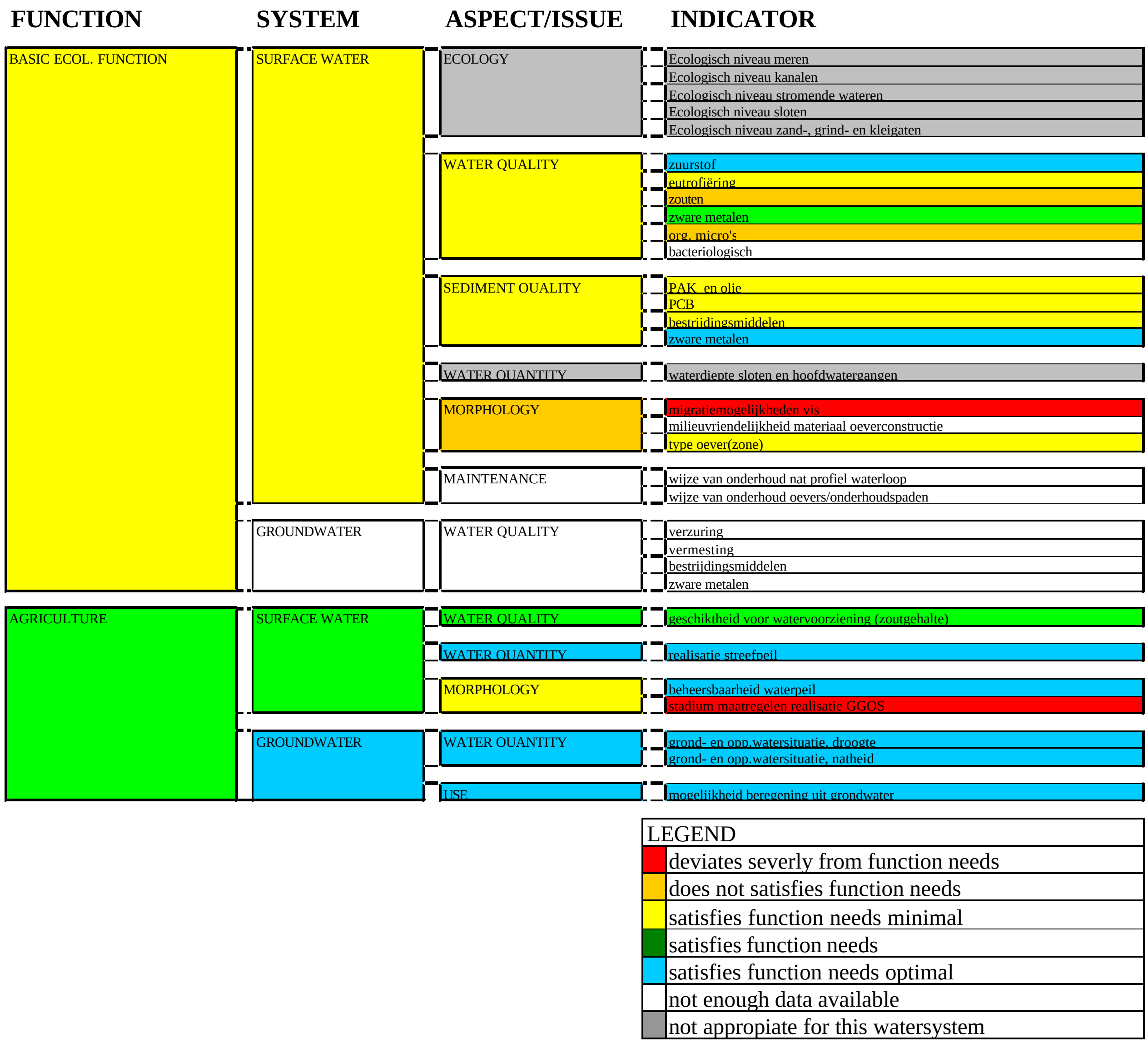


Figure 17 Regional Water Picture, the proposed presentation format in RWSR (Van der Straten et al., 1998). This graphical presentation gives the results of the results of a pilot watersystem report for Roptazijl. This area is north of the Marne-Middelsee watersystems. Note: the indicators are in dutch and 'zouten' means salination.

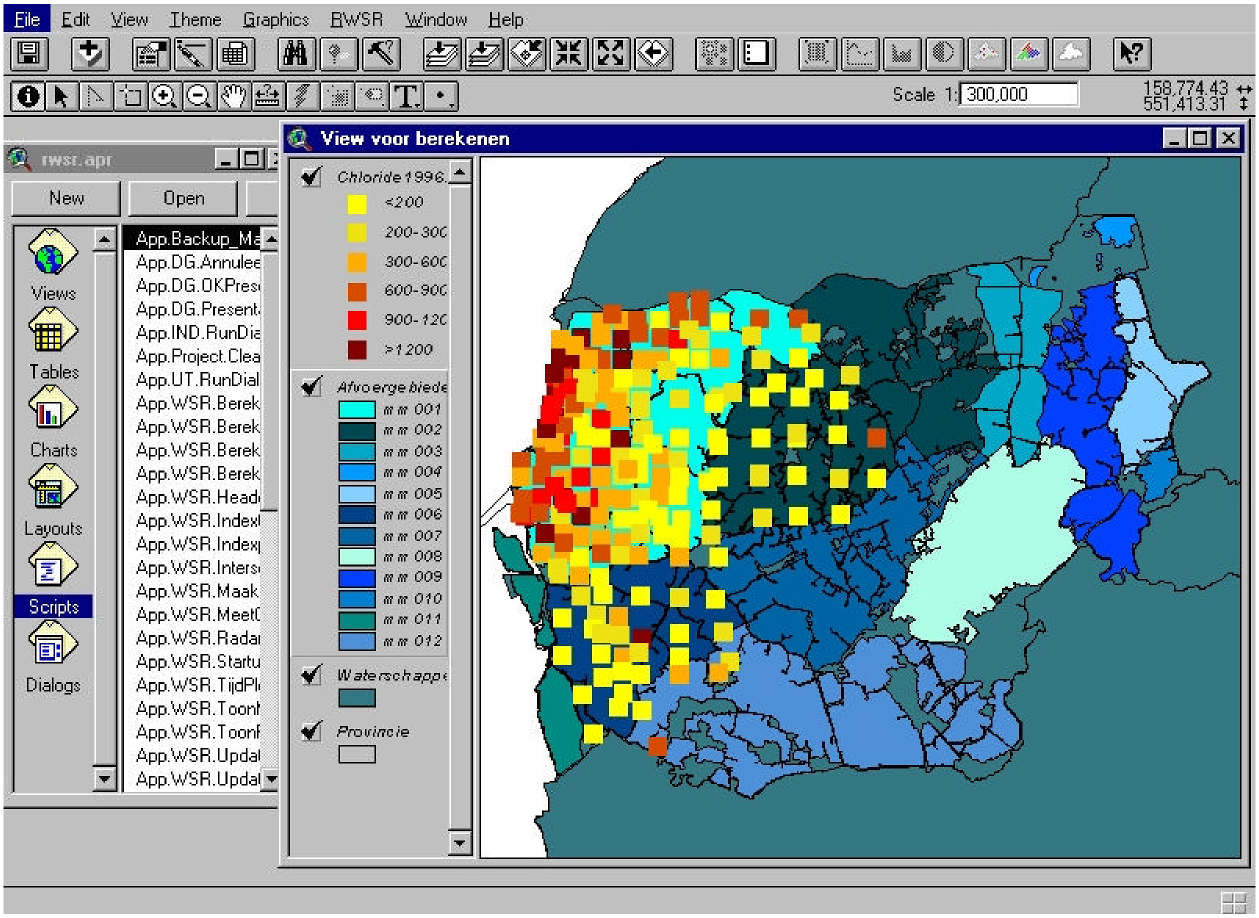


Figure 18 A look at ArcView *- combining the menubar (top), project window (left) and the view window. Here the chloride concentrations in week 13 of 1996 are shown as points on the watersystemclusters of* Marne-Middelsee*. The view window is an explorative environment as it does not contain a north arrow or proper legend.*

ArcView is currently one of the most popular desktop GIS software environments (ESRI, 1996), primarily geared to the manipulation of vector data. ArcView however lacks the possibilities of dynamic and interactive linking between maps and diagrams or figures. For this purpose CDV (Cartographic Data Visualiser)(Dykes, 1996, 1997a, 1997b) is used in this case (see Figure 19 for an overview).

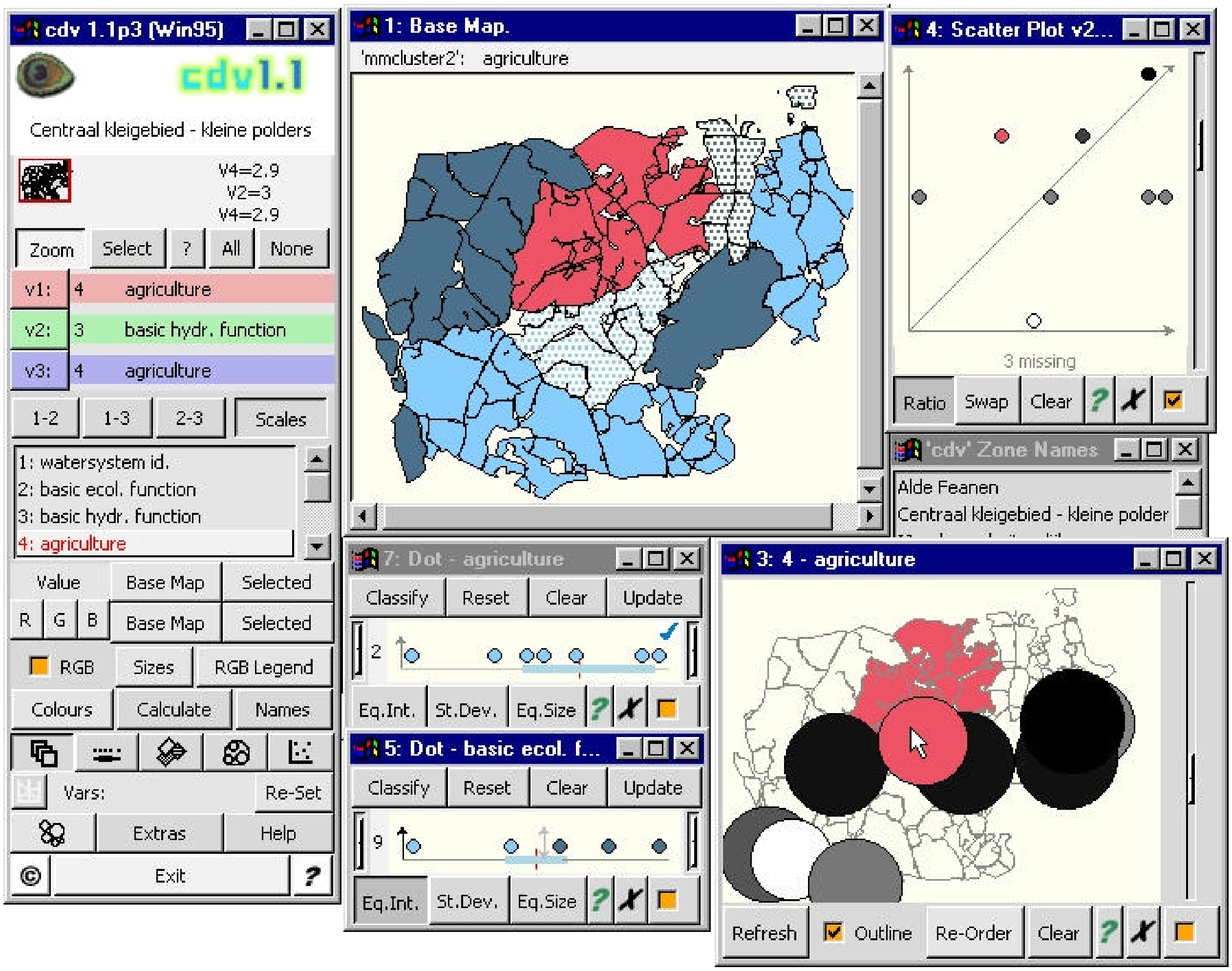


Figure 19 *A look at CDV - linked geographical and statistical views. Here the RWSR-score for agriculture is shown in the watersystemsclusters of* Marne-Middelsee *as a chloropleth (base map) and a circle map. Dot plots and scatterplots showing the functions (e.g. agriculture and basic ecological) can be used to identify watersystemclusters with extreme value in the other linked views. Note that by pointing (cursor) in the circle map the corresponding watersystemcluster in the scatterplot and base-map is highlighted.*

CDV produces a 'View Control Panel' (the left part of Figure 19) and a 'Base Map'. The control panel allows users to create different views of the same data. It contains a scrollable list of variables loaded . The lower part of the control panel allows the user to control the colour symbolism and the choice of view. By using the dotplot, polygon map, circle map, scatter plot and cartogram buttons on the control panel the user can produce alternative views of the variables.

Re-scaling, zooming and panning views, and user-defined data classification, colour schemes and symbol scaling are supported. All views are dynamically linked, meaning that when a symbol is selected with the cursor, or when a group of symbols are selected with a lasso, symbols corresponding to those chosen in one view are highlighted in all others. In a CDV session, the emphasis is on transience and with all these options the user is encouraged whenever possible to adapt the maps and diagrams to highlight specific information (Dykes and Unwin, 1998)

5.4. Identification of the problem

A decision maker, water manager has just a few (simple) questions (Hofstra, 1995; Witmer, 1995):

What is the condition of watersystems (surface waters, ground waters, banks, ..)?
Where, how, and why are water-quality conditions changing over time?
Where are problems related to water quality, and what is causing the problem?
Are programs to prevent or remediate the problems working effectively?
Are we meeting water quality goals and standards?

There may be other questions, but this gives some idea of the concerns of water management. In viewing these questions there is a clear need for generating information in a systematic method to explore, analyse and present the conditions of a watersystem. This is the aim of Regional Water System Report. It is obvious that such an assessment can only be given if the monitoring data is adequate and the objectives (as far as the reference situation) are given (IPO, 1998).

RWSR follows the structure to define watersystems, functions in this watersystems, issues/aspects of interest and indicators for these issues to identify the problems. For this case-study it is of no importance to do this for all watersystems, all functions, all issues and all (potential) indicators. Figure 20 gives an overview of the focuspoints for this case study.

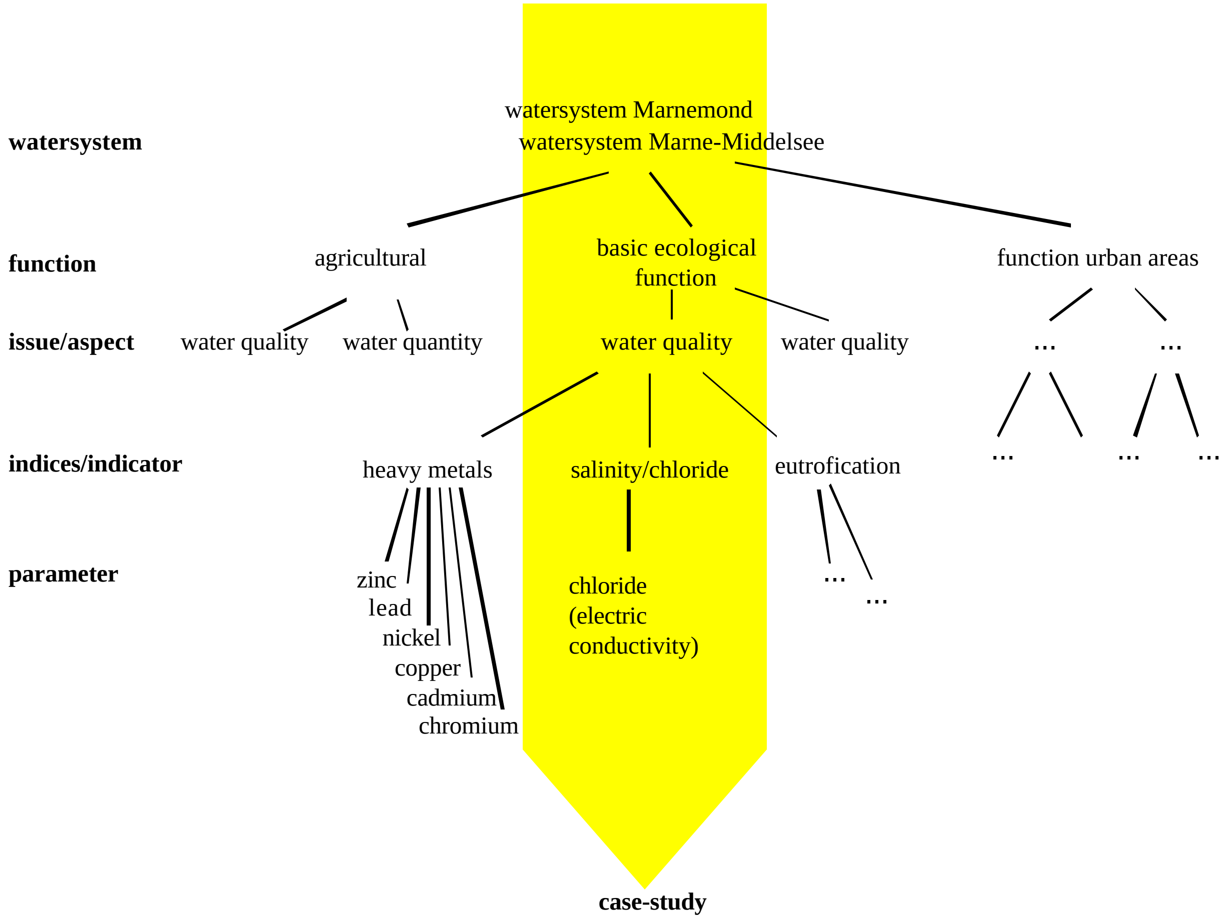


Figure 20 Structure of Regional WaterSystem Report and the focuspoints for this case-study.

5.4.1. Definition of watersystems

In 1997 the Provincie Friesland finished a study about a systematic division of watersystems in Friesland (Provincie Friesland, 1997a). This study has resulted in maps with different aggregation levels divisions for hydrological (connected) units. This method has been adopted by the CIW (1998), and is now the national guideline for defining watersystems.

The division of watersystems (or clustered watersystems) is used as the basis for the next waterpolicy document for water in the province Friesland (Provincie Friesland, 1999). For all levels digital maps are available and this thematic map will be the basis for many presentations and analysis in the waterboards the next year(s). Figure 18 gives as an example a map of the study area in Waterboard Marne-Middelsee of clustered watersystems with an overlay of chloride concentrations on measuring points. In the Waterboard Marne-Middelsee there are more then 75 watersystems, these watersystems are clustered into 11 watersystemclusters. Aggregation of watersystems can be done on different levels. From the perspective of the European Water Framework Directive (1999) the watersystems in Friesland belong to the watersystem of the Rhine.

5.4.2. Identification of functions

The next important step, after the definition of watersystems, is to define the relevant 'functions' for the water systems to be studied. The definition of 'functions' is depending on the aims and phrasing made in policy plans. Drinking water, agriculture, cooling and heat disposal, fisheries, recreation (e.g. swimming water), transport (not just goods and passengers, but also ice and sediments) and ecology and landscape are examples of 'functions' of the water. Integrated water management is the approach to meet the demands of all the different functions. This involves defining the functions and the tasks set by this function (and compromise). Defining functions for the watersystems are one of the important aspects of the planning structure in the Netherlands of both the Regional spatial plan ("streekplan") and the provincial policy document for water ("waterhuishoudingsplan"; Provincie Friesland, 1999).

5.4.3. Identification of issues

In the watersystems of Marne-Middelsee the agriculture wants salinity drops in the water but this leads to degradation of the brackish ecological potentials. By accepting both a higher agricultural (economic) value and a less perfect ecological (habitat) value, it may be possible to maximise the total (economic, ecological and socio-cultural) value of the watersystem.

In the study area the agricultural is the most important function (more then 90% of the surface). In the area there are also some brackish nature reserves. This brackish ecosystems are diminishing (Claassen and Van Straaten, 1984) and are mostly deteriorated. In the Marne-Middelsee area the brackish ecosystems are important feeding grounds for the Spoonbill (*Platalea leucorodia*).

5.4.4. Identification of indicators

Salination has been one of the first waterquality problems. The interest in this aspect of waterquality is related to the drinking-water, agricultural and ecological function of water. All of these aspects have different standards and indicators for salinity (Claassen and Van Straaten, 1984).

Salinity is important to understanding seepage of seawater in freshwatersystems. Salinity also defines suitable habitats for plants and animals (ecological function) and usability of water for agricultural use. So measuring salinity can be important when there are conflicting interest from the agricultural and ecological function.

Salinity is defined as the total amount (grams) of solid material dissolved in a kilogram of water when all carbonate has been converted to oxide, all bromide and iodine replaced by chlorine, and all organic matter completely oxidised. It is expressed as g/kg or parts per thousand (ppt or $^{o}/_{oo}$).

From this definition it is already clear that measuring salinity is not straightforward. There are four major methods to choose from when testing for salinity chloride titration, hydrometer, refractometer and conductivity. The most used testing methods are electrical conductivity (E.C) and chloride-concentration (Cl^{-}). Conductivity is the ability to conduct electricity and is the inverse of electrical resistivity. The more salts dissolved in a water sample, and the warmer the temperature, the better the water sample conducts electricity. The flow of electricity between the electrodes in the meter probe is measured and read as "micromhos per centimetre" (mmhos/cm) or "microSiemens/cm" (µS/cm). The relation between conductivity and salinity (or chloride) can be determined empirical and varies between 0,55 and 0,9 (Golterman *et al.*, 1978). The most straightforward way to express salinity is by measuring the chloride concentrations, also called chlorinity or halinity. Salinity can be determined from chlorinity by the formula (chlorinity (ppt)*1,80655 (+0,03) = salinity (ppt) (Wildberger, 1993, Claassen and van Straten, 1984).

Figure 21 gives also an overview of standards, references and classifications based on chloride-concentrations (after Claassen and van Straten, 1984). This overview of standards makes clear that the human mind percieves water quality as being good if desired water uses are satisfied and not good if they are not. This means that a particulair concentration of salinity may reflect either good or bad quality water, and the word "desired" carries the implecation that water management plans should be formulated in the light of the public's desires for water use.

The Venice-system (1959) is a universal classification based on the salinity of waters. In the second national waterquality policyplan (1981) there is a standard for shellfish based on salinity ($^{o}/_{oo}$). The international standards for drinkwater (75/440/EEG), irrigation (ECE,1993) and ecology (UN,1987) are 200 mg/l, 150 mg/l and 150 mg/l for chloride. The dutch standard for basic ecology functions has been set on 200 mg/l chloride (Ministry of Transport, Public Works and Water Management, 1999). This will be used here are the reference value (yardstick) for conditions and use of the water systems from the perspective of salinity.

		AGRICULTURAL REFERENCE				ECOLOGICAL REFERENCE	
chloride (mg Cl/l)	halinity ($^0/_{00}$)	Schaeffer (1975)	CUWVO-V (1982) standard	CUWVO-V (1982) absolute	CUWVO-V (1982) class	Redeke (1922)	Venice (1959)
					I		
50	0.05		greenhouse		(greenhouse)		
100	0.1				II	freshwater	estuarien
		sensitive crops	bulbs				
200	0.2	greenhouse	shrub	greenhouse	III		
				bulbs	IV		
				shrub			
500	0.5		horticulture		(horticulture)		
		potatoe			V	oligohalien	oligohalinicum
1000	1	cattle drinking	potatoe				
2000	2		cattle drinking	potatoe			mesohalinicum
						mesohalien	
5000	5			cattle drinking			
10000	10						
						polyhalien	polyhalinicum
20000	20					euhalien	saltwater

Figure 21 Standards of salination for agricultural and ecological function of surface waters (after Claassen and Van Straaten, 1984).

5.5. Gather and organise data

In this case there are two important sources of data needed: a map with the watersystem and chlorideconcentrations on measuring points.

The watersystem map is digitised by the province of Friesland (Hainje, 1997). The watersystem map was the digital result of the study of a systematic division of watersystems in Friesland (Provincie Friesland, 1997a). This map is based on the TOP10VECTOR of the Topographical Service of the Netherlands (TDN). This topographical map will serve as the uniform geometric reference map for the Friesian waterboards (Hainje, 1999). The Friesian waterboards have agreed to maintain this map in their GISsystem INTWIS, but as long this is not fully operational the province of Friesland will maintain and update this map. For this case the aggregation level of watersystems and watersystemclusters are used (Provincie Friesland, 1997b).

In 1996 the waterboards Marne-Middelsee, de Waadkant and Friesland have combined their monitoring activities for salinity. These waterboards (and their predecessors) have been measuring salination ever since the 1970's, however their methods where not comparable and the frequency not harmonised. In 1996 the combined monitoringprogram started with a high density of measuring points (1 locations every 4 km^2 to 1 locations every km^2) for the whole north and north-west of the province of Friesland. The measuringpoint were spread over the main watercourse (both polder- and basinwatercanals) and smaller watercourses (e.g. ditches)). In 1996 the measurements were done in week 13, 15, 17, 19, 23, 28 and 33. From 1997 the frequency during the summer (may till September) is every 2 weeks and in all other months once every 4 weeks.

In this case-study only the 208 locations in the Marne-Middelsee area are used (see Figure 18 for an impression of the density of points). From 1997 the waterboards have downsized the number of measuringpoints drastic (in the Marne-Middelsee area just 27 remained).

The waterboards have measured chloride and electrical conductivity. Electrical conductivity was measured in the field using an pocket conductivity meter and chloride concentrations were measured using an chlorocounter. On 40 locations watersamples were collected and chloride and conductivity were analysed in these samples by the laboratory of Waterboard Friesland. This was done to calibrate and verify the fieldmethods, however this verification has never taken place.

The results of the monitoringprogram were stored in several spreadsheets by the three different organisation separately. For this case-study all the different spreadsheets (and different data formats and coding within these spreadsheets) are converted and imported into an the Microsoft database of BEVER (HKVLIJN IN WATER, 1997). The structure of the database used in BEVER is based on Adventus, the national data standard for waterboards (Unie van Waterschappen, 1998).

The data stored in the BEVER database is intensively checked for errors. For instance the measuringpoints were digitised in the three successive years (1996,1997,1998) in 8 different shape-files. Locating the correct position of the measuringpoints was already a piece of investigative research. Since a number of locations was on the border of watersystems it was important to find the correct location otherwise the point would be allocated to the wrong watersystem in the RWSR-analysis.

Since the standards for salination are based on the chlorideconcentrations (see Figure 21) these parameter is of interest for the RWSR. Conductivity is far more easy to measure then chloride and therefore the waterboard where particularly interested to see whether this parameter could substitute chloride-measurements. Mid 1996 the regression between conductivity and chlorideconcentration (both in situ measurements) was calculated (personal communication G. Talsma, Waterboard Friesland). The empirical formula chloride [mg/l] = (E.C [µS/cm] * 0.35) -180 was found and used to convert E.C into chloride-concentrations. This relation was never verified and is still being used by the waterboards in presentation of salination (IWBP, 1998).

A small explorative analysis on this relationship makes one wonder of this empirical formula is the best estimator for chlorideconcentrations. In the case-study here presented these calculated chlorideconcentrations (based on the EC field measurements) are used as the measuring data in the RWSR report.

5.6. Process the data

5.6.1. Data manipulation (calculating the testdata, testratio and indicatorvalue)

The data manipulation step in RWSR consists of 2 activities: calculating the testvalue from the measuring data and calculating the indicator value. Basically this is the standardprocedure for testing for compliance (Niederländer *et al*., 1996).

In the first step the measured value from a calendar year for a location in a water system is converted to a test value (see Figure 22 for an graphical presentation of this step) . This is the value which is to be compared to the standard. For the variables, various procedures are used to determine the test value. For example for phosphorus (as an eutrophication indicator) the test value calculated is the half-yearly summer average. In the case of chloride 90-percentile of the measured values were used as test value. This calculation step is performed using NOTOVE4.0 (an sub-application of BEVER) (HKVLIJN IN WATER, 1999). The results are entered in the BEVER database (as a single annual figure per variable on a measuringpoint). All these calculation steps are performed outside the RWSR-ArcView prototype. The following steps are part of the RWSR-ArcView prototype.

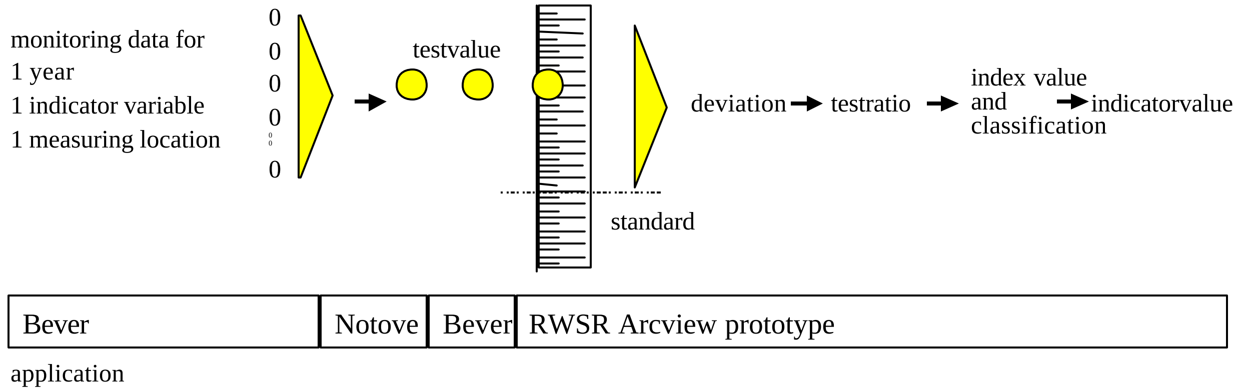


Figure 22 Overview of calculation of the testvalue and indicatorvalue, including used applications.

The next step is to determine how far the test value deviates from the standard. To achieve this, the test value has been divided by the corresponding standard, resulting in a dimensionless value, indicating the distance between the test value and the standard. In RWSR this figure is called the indicatorvalue (IPO, 1998). Within Aquatic Outlook this is called TTO ("Times-to-objective") (Stutterheim and van Buuren, 1994), later renamed in TR ("Testratio") (Stutterheim, 1997).

After this point the testratio is converted into an individual (water-)index. This is a two-step approach. The testratio is limited between 1 and 11. A test ratio less than 1 is equated to 1, while values greater than 11 are equated to 11. This testratio is converted to a water index somewhere in the range between 0 (major deviation from the standard) and 100 (no problem). The formula (waterindex = 110 - (10* testratio) shows that this step is a linear transformation.

When a indicator(value) is based on more variables the individual waterindexes are clustered to one waterindex. This waterindex is classified into the indicatorvalue. In this case-study the indicatorvalue for salination is only based on chloride, but table 10 illustrates an example for heavy metals with three variables.

	Lead	Cadmium	Copper	Zinc	Chloride
testvalue	0.20	0.82	18.9	197	1820
Standard	0.40	0.05	3	9	200
deviation	0.20/0.40= 0.50	0.82/0.05=16.4	18.9/3=6.3	197/9=21.89	1820/200=9.1
delimit	1	11	6.3	11	9.1
Individual index value	110-10*1=100	110-10*11=0	110-10*6.3=47	110-10*11=0	110-10*9.1=19
(Clustering to) waterindex				Heavy metals 37 (average)	Salination 19

Table 10 Example showing how to calculate the water index for heavy metals from four variables and for salination from chloride

The calculated waterindex is classified into a indicatorscore. Table 11 shows a number of different classificationsystems used in several application. This indicatorscore (on one location) could be transferred into a colour (see Table 11), so the individual indicatorscore could be visualised.

	Waterdialogue/RWSR		Bever		Gelderland prototype		
indicatorvalue	range	Deviation of the standard	range	Deviation of the standard	range	Deviation of the standard	Legend
1	<53	>5.7x	<61	>5x	<60	>6x	Red
2	≥53-<80	3-5.7x	≥61-<86	3-5x	≥60-<80	3-6x	Orange
3	≥80-<93	1.7-3x	≥86-<94	2-3x	≥80-<90	2-3x	Yellow
4	≥93-<100	<1.7x	≥94-<100	<2	≥90-<100	<	Green
5	100	null	100	null	100	Null	Blue

Table 11 Different classification sytems used to classify (water-)indexes. Note that the classification used in the prototype differentiates of the RWSR guidelines (Witteveen en Bos, 1999).

5.6.2. Aggregation, integration

The previous step ended with an indicatorvalue on one point in a watersystem. The next important steps are aggregation and integration. Aggregation is defined (in RWSR) as joining, in time or space, the values of one indicatorvalue to a new value for that period or that area. The usual timeframe for a watersystemreport is one year and indicatorvalues are for the same timeperiod so temporal aggregation is of no significance. There are different forms of spatial aggregation possible each with a different outcome (Witteveen en Bos, 1999; Stutterheim and Duijts, 1997). In the prototype spatial aggregation is done by joining all locations within an selected area and calculating the average indicatorvalue. In the prototype an average of all individual indexscores was used. This average was classified (see Table 11) to get a score.

Integration is defined (in RWSR) as joining the values of different indicatorvalues into a new value for the combination of indicators. The value of the separate indicatorvalues and the combined value are representative for the same timeperiod and same spatial area. For instance the indicatorvalues for salinity, heavy metals and eutrofication are integrated into a new value for waterquality (see also Figure 20). Indices of waterquality and waterquantity are integrated to one value for the basic ecological function. There also several options for integration (IPO, 1999). In the case study integration is of no importance because it illustrates just one indicator, one aspect and one system. Figure 17 illustrates however an important negative aspect of step-wise integration and averaging in every step. Note that going from indicator to function with each step the worst cases dissolve and at the function level no bad situation is present.

5.7. Reach and apply generalisations

5.7.1. Presentation

In watersystemreport there is a number of graphical methods that can be used. The most used in pilotstudy is the "regional watersystem picture" (e.g. Witteveen en Bos, 1998), but especially Gelderland (1999) has also used maps. These approaches are static (paper maps, graphics) and there are no initiatives to present results in a more interactive way.

During presentation the calculated testvalues, indicatorvalues, indexvalues are expressed as a colour (see Table 11). The usual sequence of colours is from good to bad: blue, green, yellow, orange and red (Stutterheim and van Buuren, 1994). However the points at which the colour changes are very different.

In past WATER DIALOGUE was an useful presentation form where an user could explore in a quick and easy way different types of information. WATER DIALOGUE has never been updated with new information. Since 1985 Rijkswaterstaat offer Presentator (formerly Digital Yearbook) an PC-program to explore statistical data (year-, summer- and winter- averages, maximum, minimum or medians) of their monitoring programs in a graphical way (Rijkswaterstaat, 1999). The program offers the possibility to consult the data of one of the parameters on a specific location. This data can both be consulted in tabular as in graphical form. Besides this it is possible to export data of a selection of data to Excel-files (.xls) or ASCII-files (comma separated value .csv). Data in the Presentation database are (statistical) aggregation of the measured data.

In this case study CDV demonstrates the possibilities of combining maps and graphics to switch quickly and easily to and from between the various modes of presentation. A form which encourages personal discovery of the results instead of seeing the results in one prefixed format.

5.7.2. Watersystem report

In the province of Friesland six waterboards are responsible for the water management. These waterboards are currently preparing one integrated water management plan (in dutch "Integraal Water Beheerplan; acronym IWBP). Regional WaterSystem Report has been used in the Frysian planning process to portray the current situation. The current situation of water systems is the result of area-specific characteristics and the management pursued. Measurement data and their interpretation create the basis for a description of the status of the water systems elements: water, sediments, banks and their related biota. Because not all elements of the water systems could be described by measurement data and the time evaluable within the planning process the methodology was implemented to a "Quick Scan RWSR/IWBP" (IWBP, 1998a). The results of the Quick Scan are used here as an example data set to illustrate the possibilities of dynamic exploration in CDV.

5.8. Re-evaluate

This case of watersystemreports has illustrated that the map could play an important role in each step of the explorative paradigm. An important aspect in the explorative project design is to re-evaluate each step in the analysis, look for alternative procedures and feedback. Table 12 gives an overview of the steps of this case and questions to evaluate.

Step	Example from case study	Evaluation
Gather and organise data	Measuring of electric conductivity (E.C.)	How do E.C. field measurements compare with laboratory measurements?
	Chloride (and E.C) concentrations entered in database	Are there typo's, errors, missing values in database?. Are there outliers or other curious measurements?
	Transferring E.C. to chloride concentrations	Is the emperical formula appropriate? (chloride [mg/l] = (E.C [µS/cm] * 0.35) - 180)
	Measuring points are digitised	Is the location of the measuring point right?
	Watersystems are digitised	Are watersystem properly digitised? Is the relation between measuring point topological accurate?
Data manipulation	Calculating the 90 percentile testvalue	What if the average or median value would be used?
	Comparing with the standard of 200 mg/l	What if the international standard of 150 mg/l was used?
	Clustering more variables into one index (not necessary in this case)	What if E.C. was used as variable second variable. What kind of integration would be used (averaging, logical operators (e.g. worst case) or otherwise?
	Classification of the index	What if the WaterDialogue classification would be used?
Aggregation and integration	Aggregation of the individual index on the locations the one	What if we used one representative location or the median value of the

	(area-)score	locations or average the classified scores?
	Integration of the different indicator to a new combined indicator (not in this case)	What is the effect of stepwise integration, what is the effect of weighting factors?
Presentation	Reporting the results on a map	What is the best colour legend for the different scores?

Table 12 Evaluation of the Regional Water System procedure for salination.

In this case study ArcView and CDV were used to support this evaluation process. ArcView provide functions to capture, edit, query and display spatial information. It provides an "anchor" to select, do simple manipulation of the data and provide sophisticated maps of the spatial domain. This is important for maintaining the frame of reference to the data - a map provides context for the data, which can otherwise be lost tables or graphics. CDV provided an environment in which different maps and other graphics (e.g. scatterplot) were linked. Brushing the scatterplot instantaneously update the map view and brushing the map instantaneously update the scatterplot. This kind of interactive tools make maps more expressive.

This new kind of visually exploring spatial data could be extended with statistical analysis systems (e.g. S-plus, SAS, Xplore, XlispStat, DataDesk) (Cook, 1999). It may be that such graphical analysis needs to be supported by statistics designed to help the eye/brain detect areas of "interest' on the map. A number of these statistics have been suggested (Anselin, 1999). In an integrated system it should be possible to calculate these local statistics and make plots of these linked to the map.

6. Discussion, what has this thesis illustrated?

This final chapter contains a discussion of the findings of the theoretical (part 1) and practical part (part2).

The objectives of this thesis are to define:

1. the *developments and changes* in water management and *water monitoring systems* which are taking place.

2. the *possibilities offered by new exploratory concepts* and *technologies*, which are "data-driven" and related with "learning from data".

3. the *role maps are getting* in spatial data handling, i.e. the map as a tool for exploration and not just a way of presenting results.

4. to *demonstrate* how *the new exploratory techniques*, and the map as integral part of it, can be used *in water monitoring*.

This conclusive chapter will present some (personal) observations and conclusions related to the objectives of this thesis.

6.1. Development and changes in water monitoring systems

Water management is getting increasingly complex. Monitoring programs must provide information for water management. Monitoring programs are dealing with watersystems where uncontrolled variation of driving forces such as weather, temperature, soils, geology play an role. Add to this the effects caused by human behaviour, which is if anything less predictable. As a result, monitoring programs must study watersystems of great complexity and research must overcome this complexity in order to understand the signals presented in the data. In such a situation it is increasingly hard too find the scientific sound answer (if ever there was one), where water management (and water pollution specifically) calls for action.

In the past monitoring professionals, in order to portray the complexity correct were often been seen as "always wanting more data" or "produce statistically inconclusive results" to policy, decision makers and public. Resulting in a gulf in culture between monitoring programs and decision makers, policy and general public. It is important is to produce results even when these are not "proven". On the basis of this "working" knowledge it is either possible to evaluate whether action is needed or additional knowledge required. Policy makers and water managers prefer to be informed "wrongly" (though not too often) above being informed too late or not at all.

Policy makers (in water management) also deal with the complex systems, yet are required to make tough "yes/no" choices. They are rarely trained in the disciplines of science and so require results in non-technical jargon to understand the "bottom line". In order to make the transfer of information from a monitoring program to policy and water management, monitoring professionals must make information generation an explicit objective of the work they do. They must "speak the same language" as the people who would use their information. In order to present information in an accessible way, one can choose mathematical and statistical compilations. Often it is wise to use graphic tools and visual techniques to improve the impact of the message. In many instances the information has to be short and snappy.

Only 20% of an oral presentation will be remembered, 30% of a visual presentation but 70% of what has been presented by oral and visual presentation simultaneously. The assumption behind explorative analysis is that what is discovered personally will remain the longest. From this perception the interest in interactive graphic tools grew. This interest is strenghented by the fact that water management (and monitoring) must place emphasis on public participation. This requires tools that can easily be used by a large group of non-professionals. In the discussion with colleagues there is general mistrust in exploratory analysis as "it might lead to multiple views and opinions" or that it is like beauty, only "in the eye of the beholder". It's their believe that the desired information can better be professionally analysed and interpreted and finally, be presented and distributed. On the other hand there seems to be a growing distrust of the public in governmental actions (due to many incidents like the BSE-crisis in the U.K. or dioxin-affair in Belgium). In water management there is an increasing notion that the public, broadly defined to include individuals, companies, or organisations, should participate in the decision-making process for a variety of issues. Monitoring is a good medium that should be used to reinforce "knowledge" about waterissues and therefore contribute to more involvement in water management.

6.2. Possibilities offered by new exploratory technologies

It is interesting to perceive that Ward, Loftis and McBride, as the authors of the "Data-rich but information-poor Syndrome" (1986) in water quality monitoring, defined their work from the frustration of the failure of technological innovation to extract meaning information from monitoring programs. This thesis advocates that the explorative paradigm as a result of technological innovation might improve the lack-of-information problem in a revolutionary way. It is based on the assumption that information technology, including mapping and GIS, will play an growing role in monitoring, water management and public participation. Especially the provision of WWW-based mapping, GIS functionality (largely ignored in this thesis) could play a vital role in delivery, analysis and presentation of monitoring results in the near future. Data and information should be provided to the public more quickly and easily. Internet could provide an important medium, by providing data and conclusions on-line. This provides the possibility for others to check the process of generating those conclusions against their own by interpreting the data themselves.

Where Geo-IT (technology that allow users to capture, edit, store and present geographic data (Roodzand, 1999a)) evolves to spatial information technology, there is still no general agreement on what kind of spatial analysis should be added. This in contrast to the general agreement that GIS and an automated mapping system differ in the ability to perform analysis (Burrough, 1986; Goodchild, 1987). There seems to be an notion that exploratory data analysis was the most likely candidate to be added as spatial analysis in the GIS world (see Fisher *et al.*, 1996 or Scholten and LoCascio, 1997).

In the literature there is much discussion how to integrate ESDA and GIS. In view of the technological changes towards open, distributed, heterogeneous systems it should not be the main issue whether this should be a tight or loose coupling (Fisher *et al.*, 1996) or close coupling, encompassing or modular (Anselin, 1998). In fact the more important challenge is now to demonstrate the accessibility, usability and effectiveness (Perdigao, 1998) of exploratory data analysis in many fields of interest (see for instance Unwin and Fisher, 1998).

In line with the earlier discussion about public participation and transfer of information between monitoring and water management, this thesis focus on techniques usable to non-technical and non-statistical end-users. It would be naïve to assume that usability is limited to those tools that exclude statistically and mathematically complex methods. However this thesis has focused on the developments where the map and simple statistical graphics are included as a series of dynamically linked visual displays. For it is felled that these developments are interesting for a large numbers users, who like to be (mis-)lead by it's own action rather than to rely on the computer-generated statistic, data-mining or neural network modelling result. This is no rejection of the work of people like Anselin, Openshaw or Fisher, but in most cases it simply goes beyond the comprehension of most potentially interested users.

6.3. Role maps are getting

Technology has changed the role of the map from a static presentation medium to a dynamic and explorative medium. This implies that maps could be used more generally in monitoring systems which contain a lot of spatial information. The same advances in technology also require more knowledge of the cognitive aspects of visualisation. The effectiveness of interactive, dynamic visual exploration is hard to improve without proper knowledge of spatial cognition and perception of visual displays. While there is a solid base of knowledge about perception and cognition as it relates to static graphics or maps, there is hardly any knowledge about the cognitive and perceptual issues associated with dynamic and interactive displays (MacEachren and Kraak, 1997). Assessing the effectiveness of visual tools is therefore largely an extension of existing work (see for instance Wise *et al.*, 1998; Unwin *et al.*, 1998).

Dyson (1997) notes that there are two basic kinds of scientific revolutions, those driven by new concepts and those driven by new tools. While concept-driven revolutions that results in explaining old things in new ways have received the majority of methodological attention, tool-driven revolutions are comparatively more commonly encountered. The developments in GIS, Scientific Visualisation and Exploratory Data Analysis creates a new tool-driven revolution that are likely to dominate research in those disciplines concerned with the spatial/temporal structure of our world. Good tools do not just extend or amplify existing skills; they change the nature of the task itself, and with it the understanding of what it is we are doing. (Norman, 1991)

Unwin *et al.* (1998) define their ideal tool combining features of several tools. Their ideal would include:

- the direct linking with GIS offered by SAGE and Xgobi (Haining *et al.*, 1996; Cook *et al.*, 1996)
- the spatial data structure and statistics of Spacestat (Anselin and Bao, 1997)
- the interactive graphics of MANET (Unwin *et al.*, 1996)
- the graphics prototyping capability of CDV (Dykes, 1996)
- the flexible statistical programming of LispStat and S+ (Brunsdon and Charlton, 1996; Diggle and Rowlingson, 1993)
- the interactive statistical interface of Data Desk (Velleman, 1995)

In addition, any such software would have to be able to handle large data sets, to produce excellent graphics and (automatic) maps, to be fast, easy to use and accessible. It is obvious that there is much left to be desired in the tool-driven revolution, but we're on the way to a highly visual and graphics, map-oriented technology, that makes it possible to explore and exploit data to arrive at new insights.

6.4. Demonstrate new exploratory techniques in water monitoring

Traditional the map has been treated as an output, thus the endpoint of a query or analysis process. In the Regional Watersystem Report (of the IPO, 1999) there is still this tendency to put the emphasis on the presentation role of the map. The new view in this thesis is that the map has a potential role at all stages of the RWSR process (from pre-processing and error identification through selection of data to be used, to aggregating, clustering and interpretation of the results of RWSR).

Dividing the spatial-data-user in "doers", "users" and "viewers" (Roodzand, 1999a), the Regional Water System Report case is a typical example for the "users". Monitoring professionals can benefit from access to spatial data, They should not spent the bulk of their time to creating, maitaining and managing this data, but on generating information from this data. For the larger group of "viewers" (e.g. policy makers, water managers) ArcView and CDV are to still too complex. They want to assemble data as quickly (and easy) as possible. Based on the structure of Regional Watersystem Report and the intensive use of indicators an OLAP application could be a possibility. With the growing popularity of internet it is obvious that providing tools for "viewers" more and more focus on this environment.

EDA consists of collection of descriptive and graphical tools intended to discover patterns in data and suggest hypotheses by imposing as little structure as possible (Tukey, 1977). This is supposed to lead to "potentially explicable patterns" (Good, 1983; p 290.) is qualitatively distinct from simple descriptive statistics. Modern EDA methods emphasise the interaction between human cognition and computation in the form of dynamic (statistical) graphics and maps that allow the user to directly manipulate various "views" of the data. The case-study has illustrated that much in the philosophy of EDA these methods could also be beneficial too aggregated and integrated data.

Whereas problem solving involves previous knowledge of a defined task, data exploration often involves multiple, changing or ill-defined tasks. In the process of data exploration, experiments are done for the sake of experimenting, not hypothesis testing. With computers, data can be arranged and rearranged, data representations can be edited and updated immediately and interactively. Simple curiosity rather than rigid defined tasks (as in Regional Water System Reports) can lead to new insight or discovery.

7. References

- Adriaanse, M.J., J. van der Kraats, P.G. Stoks and R.C. Ward (eds.), 1995. *Proceedings of the international workshop Monitoring Tailor-Made I*, 1994, Beekbergen, the Netherlands, 356pp.
- Adriaanse, M., H.A.G. Niederländer and P.B.M. Stortelder, 1995. *Monitoring Water Quality in the Future, Volume 1: Chemical Monitoring.* Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad. ISBN 90-802637-1-0.
- Adriaanse, M.J., 1997. Tailor-made guidelines: a contradiction in terms? *European Water Pollution Control Volume 7, number 4, 1997, pp. 11-16.*
- Andrienko, G.L. and N.V. Andrienko, 1998. *Interactive Maps for Visual Data Exploration.* ICA Commision on Visualization Warsaw, may 1998. http://allanon.gmd.de/and/icavis [last accessed 16-11-1998].
- Andrienko, G. and N. Andrienko, 1999. Knowledge-Based Visualization to Support Spatial Data Mining. In: Hand D.J., J.N. Kok, M.R. Berthold (eds.) *IDA'99, LNCS 1642*, Springer-Verlag Berlin Heidelberg, pp. 149-160
- Andrienko, G.L. and N.V. Andrienko, 1999. *Making a GIS Intelligent: CommonGIS Project View*. ***AGILE'99*** *Conference*, Rome, April 15-17, 1999, pp.19-24, On: ftp://ais.gmd.de/pub/SET/publications/released/1999/pdf/Andrienko99.1.pdf. [last accessed 05-05-2000]
- Anselin, L., 1990. What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis. In: D.A. Griffith (ed.), *Spatial Statistics, Past, Present and Future*, Ann Arbor, MI:Institute of Mathematical Geography, pp. 63–77.
- Anselin, L., 1994. Exploratory Spatial Data Analysis and Geographic Information Systems. In: M. Painho (ed.), *New Tools for Spatial Analysis*, Luxembourg: Eurostat, pp. 45–54.
- Anselin, L., 1995. Local indicators of spatial association LISA. *Geographical Analysis* 27: p.93-115.
- Anselin, L., 1999. Interactive techniques and exploratory spatial data analysis. Chapter 17. In: P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.) *Geographical information systems: principles, techniques, management and applications*. New York, Wiley, p. 253-266.
- Anselin, L. and S. Bao, 1997. Exploratory spatial data analysis linking Spacestat and ArcView. In: M. Fischer and A. Getis (eds.) *Recent developments in Spatial Analysis: Spatial statistics, behavioural modelling and neuro-computing*. Berlin, Springer-Verlag, pp. 35-39.
- Anselin, L. and A. Getis, 1992. Spatial Statistical Analysis and Geographic Information Systems. *The Annals of Regional Science* 26, p. 19–33.
- Atkins, N.C, 1993. *A Framework for development of data analysis protocols for ground water quality monitoring.* Colorado Water Resources Institute Technical Report No. 60. Colorado State University, Fort Collins.
- Bailey, T.C. and A.C. Gatrell, 1995. *Interactive spatial data analysis*. Harlow: Longman Scientific and Technical.
- Batty, M. and Y. Xie, 1994. Modelling Inside GIS: Part I. Model Structures, Exploratory Spatial Data Analysis and Aggregation*, International Journal of Geographical Information Systems* 8, p.291–307,
- Becker, R.A., W.S. Cleveland and R.A. Wilks, 1987. Dynamic Graphics for Data Analysis. *Statistical Science 2*, p. 355-395.
- Blind M.W. and P.J. van der Wiele, 1998. *Methodiek voor de evaluatie en optimalisatie van routine waterkwaliteitsmeetnetten. Deel 1: hoofdrapport*. Stichting Toegepast Onderzoek Waterbeheer 98-15, 53p. (in Dutch)

- Blok, C. and B. Köbben, 1998. A Web Cartography Forum: an evaluation site for visualization tools. Working paper for meeting of the ICA Commission on Visualisation in Warsaw, may 1998. http://www.itc.nl/~carto/webcartoforum/paper.html [last accessed 13/06/99]
- Board, Chr., 1990. Report of the working group on cartographic definitions. *Cartographic Journal*, 29, p.65-69.
- Brundtland, G.H., 1987. *Our common future*. Oxford University Press, New York, 400pp.
- Brunsdon, C. and M.E. Charlton, 1996. Developing an exploratory spatial analysis system in Xlisp-Stat. In: D. Parker (ed.) *Innovation in GIS 3*, London, Taylor and Francis, pp. 133-145.
- Buja, A., D. Cook and D.F. Swayne, 1996. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics* 5: p. 78-99.
- Burrough, P.A., 1996. *Principle of Geographical Information Systems for land resources assesment.* Oxford: Oxford University Press.
- Burrough, P.A., 1990. Methods of spatial analysis in GIS. *International Journal of Geographical Information Systems*, vol 4, p. 221-223.
- Chambers, J.M., 1993. Greater or Lesser Statistics: A choice for Future Research. *Statistics and Computing*, 3:4 (1993), pp. 182-184. On http://cm.bell-labs.com/stat/doc/greater.ps [last accessed 25-04-2000]
- CIW, 1998. *Leidraad begrenzing watersystemen*. Den Haag. (in Dutch)
- Clarke, K., 1999. Visual Reasoning: The link Between User Interfaces and Map Visualisation for Geographic Information Systems. In: M. Craglia and H. Onsrud (eds*.). Geographic Information Research, Trans-Atlantic Perspectives*. London, Taylor & Francis, pp. 535-548.
- Claassen, T.H.L. and H. van Straten, 1984. Kwellend zout in Friesland, een oriënterend onderzoek. *Cultuurtechnisch tijdschrift*, jrg.23, nr.6, p. 311-323.
- Claessen, F.A.M., 1997. Comparing monitoring of surface and ground water systems*. European Water pollution Control*, volume 7, number 4, 1997, p. 27-35.
- Cleveland, W.S., 1993. *Visualizing Data*, Summit, NJ: Hobart Press.
- Cleveland, W.S. and M.E. McGill, 1988. *Dynamic Graphics for Statistics*, Pacific Grove, CA: Wadsworth.
- Cook, D., J.J. Majure, J. Symanzik and N. Cressie, 1996. Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software, *Computational Statistics* 11, 467–480.
- Cook, D., 1999. Visualizing Multivariate Spatial Data. On: http://www.ncgia.ucsb.edu/sa_workshop/papers/cook.html [last accessed 24-03-2000]
- Couclesis, H., 1991. Visions of quality: visualising product quality in GIS. *NCGIA, Initiative 7*, Santa Barbara.
- Cressie, N.A.C, 1993. *Statistics for spatial data*, revised edition. New York, John Wiley & Sons Inc.
- Csinger, A., 1992. *The Psychology of Visualisation*. Technical Report Series, Department of Computer Science, University of British Columbia.
- DiBiase, D., 1990. *Visualization in earth sciences*. Earth and Mineral Scienses, Bulletin of the College of Earth and Mineral Sciences. The Pennsylvania State University, 59(2), 13-18. Reprint on: http://www.geovista.psu.edu/publications/dibiase90/swoopy.html [last accessed 19-10-1999]
- De Jong, J., P.T.J.C van Rooy and S.H. Hosper, 1995. Living with water: at the cross-roads of change. In: S.H. Hosper, R.D Gulati, L van Liere and R.M.M Rooijackers (eds.). Integrated water resources management. *Water Science & Technology*, volume 31(8), pp.393-399.
- De Vree, L.G and M.W. Blind, 1998. *Methodiek voor de evaluatie en optimalisatie van routine waterkwaliteitsmeetnetten. Deel 2: overzicht van technieken en methoden*. Stichting Toegepast Onderzoek Waterbeheer 98-16, 129p. (in Dutch)
- Demayo, A. and S. Whitlow, 1993. Graphical presentation of water quality data. In: Lerner, D. (ed.), *Monitoring to detect changes in water quality series*. IAHS Publications No. 157, p. 13-27.

- Diggle, P.J. and B.S. Rowlingson, 1993. SPLANCS: spatial point pattern analysis code in S-plus. *Computers & Geosciences*, 19, p. 627-655.
- Dykes, J., 1996. Dynamic maps for spatial science: a unified approach to cartographic visualization. In: D. Parker (ed.), *Innovation in GIS 3*, London, Taylor and Francis, 177-187.
- Dykes, J., 1997a. Exploring Spatial Data Representation with Dynamic Graphics. *Computers and Geosciences* 23(4), p. 345-370. On: http://www/elsevier.nl/homepage/misc/cageo/dykes/stills/abstract.html [last accessed on 22-01-1999]
- Dykes, J., 1997b. *cdv: A Flexible approach to ESDA with Free Demonstration Software*. On: http://www.geog.le.ac.uk/jad7/BSC/paper.html [last accessed 22-01-1999]
- Dykes, J. and D. Unwin, 1998. Maps of the Census: a rough guide. Unwin. D. & Fisher P (eds.), *Case Studies of Visualization in the Social Sciences*. On: http://www.agocg.ac.uk/sosci/casestudies/dykes/abstract.html [last accessed 29/03/2000]
- Dyson, F.J., 1997. *Imagined Worlds: The Jerusalem-Harvard Lectures*. Cambridge: Harvard University Press.
- ECE, 1993. Protection of Water Resources and Aquatic Ecosystems, *Water Series No 1*. ECE/NVWA/31, United Nations.
- EEC, 1975. *Richtlijn van de Raad betreffende de vereiste kwaliteit van het oppervlaktewater dat is bestemd voor produktie van drinkwater in de Lid-staten*. (in Dutch.)
- Egenhofer, M.J. and W.Kuhn, 1999. Interacting with GIS, chapter 28. In: P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.) *Geographical information systems: principles, techniques, management and applications.* New York, Wiley, p. 401-412.
- ESRI, 1996. *ArcView GIS* , Redlands, CA: Environmental Systems Research Institute.
- ESRI Nederland, 1999. *Ontwerp RWSR GIS module*. ESRI Nederland by order of IPO-RWSR. December 1999.
- Fayyad, U. , G. Piatetsky-Shapiro and P. Smyth, 1996. The KDD Process of Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39, 27-34.
- Fisher, M., H.J. Scholten and D. Unwin, 1996. Geographic information Systems, spatial data analysis : an introduction. In: M.M. Fischer, H.J. Scholten and D. Unwin (eds.) *Spatial Analytical Perspective on GIS, GISDATA IV*. LondonTaylor & Francis, pp 3 – 19.
- Fisher, M.M., 1999. Spatial analysis: retrospect and prospect, Chapter 19. In: In: P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.) *Geographical information systems: principles, techniques, management and applications*. New York, Wiley, p283-292.
- Friedhoff, R.M. and W. Benzon, 1989. *Visualization: The Second Computer Revolution*. New York: Harry N. Abrams.
- Freitag, U., 1993. Map functions. In: T. Kanakubo (ed.), *The Selected Main Theoretical Issues Facing Cartography,* Cologne: International Cartographic Association, pp. 9-19.
- Gahegan, M., 1998. Four Barriers to the Development of effective Exploratory Visualisation Tools for the Geosciences. On: http://www.cs.curtin.edu.au/~mark/visworkshop/visproblems.html [last accessed 8-1-1999]
- Gahegan, M., 1999. *Visualisation Strategies for Exploratory Spatial Analysis*. On: http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/gahegan_n [last accessed 18-02-1999]
- Getis, A., 1993. GIS and modeling prerequisites. In: Frank, A.U. and I. Campari (eds.), *Spatial information theory: a theoretical basis for GIS*. Berlin, Springer, 322-340.
- Getis, A., 1999. Spatial statistics, chapter 16. In: P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.), *Geographical information systems: principles, techniques, management and applications*. New York, Wiley, p239-251.
- Golterman, H.L., R.S. Clymno and M.A.M. Ohnstad, 1978. *Methods for Physical & Chemical Analysis for Fresh Waters*. Second Edition: IBP Handbook no.8.

- Good, I.J., 1983.The Philosophy of Exploratory Data Analysis, *Philosophy of Science* 50, p.283–295.
- Goodchild, M.F., 1987. A spatial analytical perspective on geographical information systems. *International Journal of Geographical Information Systems*, vol 1, p.327-334.
- Goodchild, M.F. and P.A. Longley, 1991.
- Goodchild, M.F., R.P. Haining and S. Wise *et al.*, 1992. Integrating GIS and spatial analysis: problems and possibilities. *International Journal of Geographical Information Systems* 6: p. 407-423.
- Goodchild, M.F. and P.A. Longley, 1999. The future of GIS and spatial analysis, chapter 40. In: P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.), *Geographical information systems: principles, techniques, management and applications*. New York, Wiley, p. 567-580.
- Goodchild, M.F., 1999. Multiple Roles for GIS in Global Change Research. In: M. Craglia and H. Onsrud (eds.). *Geographic Information Research, Trans-Atlantic Perspectives*. London, Taylor & Francis, pp. 277-296.
- Gore, A., 1998. The Digital Earth: Understanding our planet in the 21st century. Given at the California Science Center, Los Angeles, California, on January 31, 1998. On: http://digitalearth.gsfc.nasa.gov/VP19980131.html [last accessed 05-05-2000]
- Haining, R.P., S.M. Wise and J. Ma, 1996. The design of a software system for the interactive spatial statistical analysis linked to a GIS. *Computational Statistics* 11, p. 449-466.
- Hainje, H.A., 1997. UNIGIS assignment module 5, Data aquisition.
- Hainje, H.A., 1999. An assessment of the problems of planning and implementing GIS in a (complex) multi-partner organisation. *MSc thesis, University of Salford*.
- Haslett, J., G. Wills and A. Unwin, 1990. SPIDER - An Interactive Statistical Tool for the Analysis of Spatially Distributed Data, *International Journal of Geographical Information Systems* 4, p. 285–296.
- HKVLIJN IN WATER, 1997. *Description systemconcept Bever*, versie 4.0 (in Dutch), Lelystad.
- HKVLIJN IN WATER, 1999a. *User Manual Bever*, versie 2.1. (in Dutch). Lelystad. See also www.minvenw.nl/rws/riza/bcm/projecten/bever [last accessed 05-05-2000]
- HKVLIJN IN WATER, 1999b. *User Manual Notove*, versie 4.0. (in Dutch). Lelystad. See also www.minvenw.nl/rws/riza/bcm/projecten/bever [last accessed 05-05-2000]
- Hogeweg, M, 2000. Spatio-temporal visualisation and analysis. *MSc thesis, University of Salford*
- Hofstra, M.A., 1995. Information is vital for the natianol decision maker. In: Adriaanse M.J., J. van der Kraats, P.G. Stoks and R.C. Ward (eds.), *Proceedings of the international workshop Monitoring Tailor-Made I*, 1994, Beekbergen, the Netherlands, p. 43-54.
- Hosper S.H., R.D Gulati, L. van Liere and R.M.M Rooijackers (eds.). Integrated water resources management. *Water Science & Technology,* volume 31(8).
- Hotto, H.P, T.G. Sanders and R.C. Ward, 1997. Performance evaluation of water quality information systems. In: J.J Ottens, F.A.M. Claessen, P.G. Stoks, J.G. Timmerman, R.C Ward (eds.) *Proceedings of the international workshop Monitoring Tailor-Made II*, 1996, Nunspeet, the Netherlands, p. 277-286.
- IPO, 1996. *Regionale Watersysteem Rapportage, projectprogramma* (in Dutch).
- IPO, 1998. *Manual Regional Watersystem Report* (in Dutch). Oktobre 1998. See also http://www.waterland.net/iporwsr/ [last accessed 05-05-2000]
- IPO, 1999. *Projectplan RWSR automatisering* (in Dutch). Oktobre 1999.
- ITMF, (Intergovernmental Task Force on Monitoring Water Quality), 1995. *The strategy for improving water-Quality Monitoring in the United States*. Final Report of the USGS, Office of Water Data Coordination, 417 National Center, Reston, VA 22092, USA. On :http://water.usgs.gov/public/wicp/lopez.main.html [last accessed 05-05-2000]

- IWBP, 1998. *Themanotitie verzilting*, november 1998. Werkgroep Verzilting, Integraal Waterbeheersplan Friesland (in Dutch).
- John, G.H., 1997. Enhancements to the Data Mining Process. *PhD dissertation, Stanford University.* On: http://robotics.stanford.edu/~gjohn/ [last accessed 05-05-2000]
- Kodratoff, Y., 1997. *From the art of KDD to the science of KDD*. Research report 1096. Universite de Paris-sud.
- Kraak, M.J., 1998. Exploratory Cartography: maps as tools for discovery. On: http://www.itc.nl/~kraak/address [last accessed 12-09-1998]
- Kraak, M.J., 1999. Visualising spatial distribution. In: P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.), *Geographical information systems: principles, techniques, management and applications.* New York, Wiley, p. 157-173.
- Kraak, M.J. and F.J. Ormeling, 1996. *Cartography, visualization of spatial data*. Addison Wesley Longman Limited.
- Laane, R.W.P.M. and B.J.E ten Brink, 1990. Data rich, information poor. The modern monitoring syndrome? *Land Water Int, 68.*
- Langran, G. and N.R. Chrisman, 1988. A Framework for Temporal Geographic Information. *Cartographica* 25,3, p.1-14.
- Langran, G., 1992. *Time in Geographic Information System*s. London: Taylor & Francis.
- Latour, P.J.M., E. Stutterheim and A.J. Schafer, 1996. From data to information: the WATER DIALOGUE. *H_2O*(29), p. 693-696 (in Dutch, abstract in English).
- Lijckema, L.L., 1995. Water Quality Standards: Sense and nonsense. In: Hosper S.H., R.D Gulati, L van Liere and R.M.M Rooijackers (eds.). Integrated water resources management. *Water Science & Technology,* volume 31(8), p. 321-328.
- Literathy, P., 1997. Water Quality Monitoring in the Danube River Basin. In: Ottens J.J, F.A.M. Claessen, P.G. Stoks, J.G. Timmerman, R.C Ward (eds.) *Proceedings of the international workshop Monitoring Tailor-Made II*, 1996, Nunspeet, the Netherlands, p. 213-220.
- Longley, P., 1999. Position Statement on Spatial analysis. On: http://www.ncgia.ucsb.edu/conf/sa_workshop/papers/longley.html [last accessed 18-10-1999].
- Luiten, J.P.A., 1995. The water system explorations – A new dutch project (The Aquatic Outlook) for combining monitoring, research and policy analysis for integrated water management. In: Hosper S.H., R.D Gulati, L van Liere and R.M.M Rooijackers (eds.). Integrated water resources management. *Water Science & Technology,* volume 31(8), p. 329-344. See also http://www.waterland.net/rikz/wsv/intro.html [last accessed 05-05-2000].
- Maasdam, R. and T.H.L. Claassen, 1997. The Frisian lakes: Trends in water quality. *Verh. Internat Verein Limnol.,* Vol 26., pp 736-739.
- Maasdam, R. and T.H.L. Claassen, 1998. Trends in water qualitu and algal growth in shallow Frisian Lakes, The Netherlands. *Water and Science Technology*, vol. 37, no3. , pp. 177-184.
- Maasdam, R. and D.G. Smith, 1994. New Zealand's National River Water Quality Network 2. Relationships between physico-chemical data and environmental factors. *New Zealand Journal of Marine and Freshwater Research*, Vol. 28, pp 37-54.
- MacEachren, A.M. and M. Monmonier, 1992. Geographic Visualization: Introduction. *Cartography and Geographic Information Systems,* 19(4): p. 197-200.
- MacEachren, A.M., 1994. Visualization in modern cartography; setting the agenda. In: D.R.F. Taylor and A.M. MacEachren (eds.) *Visualization in modern cartography*. London. Pergamon Press.
- MacEachern, A.M., 1995. *How maps work: representation, visualization and design*. The Guilford Press.

- MacEachern, A.M. and M.J. Kraak, 1997. Exploratory Cartographic Visualization: Advancing the agenda. *Computer & Geosciences,* 23, p335-344. Speciaal Issue- ICA: Commission on Visualization. http://www.geog.psu.edu/ica/icavis/m&kintro.html [last accessed 15-01-1999].
- MacEachern, A.M., 1998. *Visualization – Cartography for the 21st century*. International Cartographic Association Commission on Visualization meeting, Warsaw, Poland, may 1998. On: http://www.geog.psu.edu/ica/icavis/poland1.html [last accessed 16-11-1998].
- Mar, B.W., R.R. Horner, J.S. Richley, R.N. Palmer and D.P. Lettenmaier, 1986. Data acquisition: cost-effective methods for obtaining data on water quality. *Environmental Science and Technology,* 20(6), p. 545-551.
- Marchak, F.M. and L.C. Marchak, 1990. Dynamic graphics in the exploratory analysis of multivariate data. *Behavior Research Methods, Instruments, & Computers*, 22 , p. 176-178.
- Marchak, F.M. and L.C. Marchak, 1991. Interactive versus passive dynamics and the exploratory analysis of multivartiate data. *Behavior Research Methods, Instruments, & Computers*, 23, p. 296-360.
- MathSoft, 1996a. *S+Gislink*, Seattle: MathSoft, Inc..
- MathSoft, 1996b. *S+Spatialstats User's Manual, Version 1.0*, Seattle: MathSoft, Inc.
- McCormick, B.H., T.A. DeFanti, and M.D. Brown (eds.), 1987. Visualization in Scientific Computing: *Computer Graphics*, v. 21, no. 6.
- Miller, H.J. and J. Han (eds.), 1999. *Discovering Geographic Knowledge in Data-Rich Enviro*nments. Report of Specialist meeting of the Varenius Project, Kirkland, Washington 18-20 March 1999. On: http://www.spatial.maine.edu/~max/varenius/KDreport.pdf [last accessed 25-04-2000]
- Ministery of Transport, Public Works and Water Management, 1981. *Indicatief meerjaren Programma Water 1980-1984*. SDU, The Hague. (in Dutch)
- Ministery of Transport, Public Works and Water Management, 1985. *Omgaan met Water*. SDU, The Hague. (in Dutch).
- Ministery of Transport, Public Works and Water Management, 1989. *Derde nota waterhuishouding*. SDU,The Hague. (in Dutch)
- Ministery of Transport, Public Works and Water Management, 1999. Waterkader, Vierde nota waterhuishouding (in Dutch). On: http://www.waterland.net/ as Framework for water [last accessed 05-05-2000]
- Misseyer, M.P, 1999. Time, Area, Substance and Human activity referred emission inventory. *Ph.D. Thesis, Vrije Universiteit*, Amsterdam.
- Monmonier, M. ,1991. *How to lie with maps*. Chicago, University of Chicago Press.
- Nagel, M., 1996. Exploratory Analysis of Medical Mass Data. *Computational Statistics*, 11(4), p. 429-448.
- Nyerges, T.L. and R.G. Golledge, 1997. Unit 007, Asking Geographic Questions. *NCGIA Core Curriculum in GIS*. National Center for Geographic Information and Analysis, University of California, Santa Barbara. On: http://www.ncgia.ucsb.edu/giscc/units/u007/u007.html [last accessed 19-02-1999]
- Niederländer, H.A.G., J. Dogterom, P.H.L. Buijs, R. Hupkes and M. Adriaanse (1996). UN/ECE Task Force on Monitoring and Assessment. Working programme 1994/1995. *Volume 5. State of the Art in Monitoring and Assessment of Rivers*. International Centre of Water Studies, Amsterdam, commissioned by (RIZA). ISBN 9036945968.
- NITG-TNO, 1997a. *RWSR Informatie Systeem Definitie, deelrapport I* (in Dutch). NITG-TNO by order of IPO-RWSR deelproject 5. September 1997.
- NITG-TNO 1997b. *RWSR Verkenning Implementatie scenario's ontwikkelvarianten, deelrapport II* (in Dutch). NITG-TNO by order of IPO-RWSR deelproject 5. September 1997.

- Norman, D.A. 1991. Cognitive artefacts in Caroll J.M. (eds) Designing Interaction; the Psychology at the Human-Computer Interface New York: Cambridge University Press, P 17-38.
- Ongley, E.D., 1995. The Global water quality programme. In: M.J. Adriaanse, J. van der Kraats, P.G. Stoks and R.C. Ward (eds.), 1994. *Proceedings of the international workshop Monitoring Tailor-Made I*, Beekbergen, the Netherlands, p. 25-33.
- Ongley, E.D., 1997. Matching water quality programs to management needs in developing countries: the challenge of program modernization. *European Water Pollution Control,* Volume 7, number 4, 1997, p. 43-48.
- Ongley, E., 1998. Modernisation of water quality programmes in developing countries: issues of relevancy and cost efficiency. *Water Quality International*, p. 37-42.
- Openshaw, S., 1991. Developing appropiate spatial analysis methods for GIS. In: D. Maguire, M.F. Goodchild, D. Rhind (eds*.) Gis principles and applications volume 1*, Longman, London, p. 389-402.
- Openshaw, S. and G. Clarke, 1996. Developing spatial analysis functions relevant to GIS environment. In: M.M. Fischer, H.J.Scholten and D. Unwin (eds.) *Spatial Analytical Perspective on GIS, GISDATA IV*. LondonTaylor & Francis, pp. 21–37.
- Openshaw, S. and S. Alvanides, 1999. Applying geocomputation to the analysis of spatial distributions, chapter 18. In P.A. Longley, M.F. Goodchild, D.J Maguire, D.W. Rhind (eds.) *Geographical information systems: principles, techniques, management and applications*. New York, Wiley, pp. 267-282.
- Ottens, J.J, F.A.M. Claessen, P.G. Stoks, J.G. Timmerman, R.C Ward (eds.), 1997*. Proceedings of the international workshop Monitoring Tailor-Made II*, 1996, Nunspeet, the Netherlands, 492pp.
- Payne, F.E. and J. Ford, 1988. *The concept of TIME – Temporally Integrated Monitoring of Ecosystems –Supplement.* Internal report, U.S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, Oregon.
- Perdigão, J., 1998. *CommonGIS: Gis for everyone, from everywhere*, 4th EC-GIS Workshop Budapest, Hungary, 24-26 june 1998. On: http://commongis.jrc.it/CommonGIS/ [last accessed 05-05-2000]
- Petch, J., 1996. *Visualisation, course notes.* International Distance Learning GIS Diploma Programme, Manchester.
- Provincie Friesland, 1997a. *Watersysteemindeling Fryslân*. Rapportage. IWACO in opdracht van Provincie Friesland. (in Dutch)
- Provincie Friesland, 1997b. *Watersysteembeschrijving provinsje Fryslân.* IWACO in opdracht van Provincie Friesland. (in Dutch)
- Provincie Friesland, 1999. *Tweede Waterhuishoudingsplan*, concept. (in Dutch)
- Provincie Gelderland, 1999. *Monitoring Gelders Omgevingsbeleid*, Regionale Watersysteemrapportage Gelderland (in Dutch). Arnhem.
- Rhyne, T.M., 1996. Collaborative Computing and Integrated Decision Support Tools for Scientific Visualization. *Section III of Siggraph 96 Course #16Notes for Visualizing Scientific Data and Information*. On: http://www.education.siggraph.org/materials/hypervis/misc/rhyne1.htm [last accessed 06-01-1999]
- Robinson, A.H., J.L. Morrison, P.C. Muhrcke, A.J. Kimerling and S.C. Guptill (1995). *Elements of cartography*, 6th edition. New York, Wiley.
- Roodzand, J., 1999a. The new approach to GEO-IT. *Geoinformatics,* vol2, june 1999, p. 12-15.
- Roodzand, J., 1999b. The next generation of spatial information systems. *Geoinformatics* ,vol2, september 1999, 10-13.
- Rijkswaterstaat, 1999. *Jaarboek Monitoring Rijkswaterstaat 1998*. On: http://www.waterland.net/jmr [last accessed on 05-05-2000]

- Sanders, T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian and V. Yevjevich, 1987. *Design of Networks for Monitoring Water Quality* (2nd edition). Littleton CO: Water Resources Publications.
- Saeijs, H.L.F. and M.J. van Berkel, 1995. Global water crisis: the major issue of the 21st century, a growing and explosive problem. *European Water Pollution Control*, vol5., no.4., p 26-40.
- Schilperoort, T. and S. Groot, 1983*. Design and optimatization of water quality monitoring networks*. Publications No. 286. Delft Hydraulics Laboratory, Delft, Netherlands.
- Scholten, H. and A. LoCascio, 1997. Gis application Research: History, Trends and Developments. On: http://www.shef.ac.uk/uni/academic/D-H/gis/key3.html [last accessed 27-10-1998].
- Shoham, Y. and Goyal, N. 1988. Temporal reasoning in artificial intelligence. In: Shrob, H.E. (eds.) *Exploring Artificial Intelligence: Surveys Talks from National Conferences on Artificial Intelligence*. Morgan Kaufman Publishers, 419-438.
- Silk, D.J., 1991. *Planning IT: Creating an Information Strategy*. Oxford, Butterworth-Heinmann.
- Slater, F., 1982. *Learning through Geography*. Heineman Educational Books, Ltd, London, UK.
- Smith, D.G. and R. Maasdam, 1994. New Zealand's National River Water Quality Network 1. Design and physico-chemical characterisation. *New Zealand Journal of Marine and Freshwater Research*, Vol. 28, pp 19-35.
- Smith, D.G., G.B. McBride, G.G. Bryers, R.J. Davies-Colley, J.M. Quinn and W.N Vant, 1989. *Design for a National Water Quality Monitoring Network for New Zealand*. Consultancy Report 8016/1, Water Quality Centre, Hamilton, New Zealand.
- Somlyódy, L., 1995. Water quality management; Can we improve integration to face future problems. In: Hosper S.H., R.D Gulati, L van Liere and R.M.M Rooijackers (eds.). Integrated water resources management. *Water Science & Technology,* volume 31(8), p249-260.
- Stutterheim, E. and J.T. van Buuren, 1994. *Aggregation of data in the Water-Dialogue, a management decision support system for water quality*. RIKZ work document RIKZ/AB-94.152x.
- Stutterheim, E., 1997. *Ins en outs van de Waterdialoog, achtergronden en techniek*. Rapport nummer RIKZ-97.044 (in Dutch).
- Symanzik, J., J. Majure, D. Cook and N. Cressie, 1994. Dynamic Graphics in a GIS: A Link between Arc/Info and XGobi, *Computing Science and Statistics,* 26, p. 431–435.
- Symanzik, J., J.J. Majure and D. Cook, 1996.Dynamic Graphics in a GIS; A Bidirectional Link between ArcView 2.0 and XGobi, *Computing Science and Statistics,* 27, p. 299–303.
- Taylor, R. and U. Hänni, 1998. Putting the spin on spatial, *GIS Europe*, may 1998, p. 22-23.
- Timmerman, J.G., M.J. Gardner and J.E. Ravenscroft, 1996. *UN/ECE Task Force on Monitoring and Assessment. Working programme 1994/1995. Volume 4: Quality Assurance*. Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad, and Water Research Centre, Medmenham, England, commissioned by RIZA. ISBN 9036945860
- Timmerman, J.G. and J. Hendriksma, 1997. Informatie op maat: een raamwerk voor waterbeheer. *H_2O*(30), 1997:17, p. 528-530. (In Dutch, English abstract).
- Tobler, W., 1979. Cellular geography. In: Gale, S., G. Olsson (eds.) *Philosophy in geography*. Reidel, Dordrecht, pp. 379-386.
- Tonkes, M., C. van de Guchte, J. Botterweg, D. de Zwart and M. Hof (1995). *Monitoring Water Quality in the Future, Volume 4: Monitoring Strategies for Complex Mixtures*. AquaSense Consultants, Amsterdam, and Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad. ISBN 90-802637-4-5.
- Trodd, N., 1996. *Data Quality, course notes, module 7.* International Distance Learning GIS Diploma Programme, Manchester.
- Tufte, E., 1985. *The visual display of Quantitive Information*. Connecticut: Graphic Press.
- Tufte, E., 1990. *Envisioning information*. Connecticut: Graphic Press.
- Tukey, J.W., 1977. *Exploratory Data Analysis,* Reading MA: Addison-Wesley.

- Tukey, J.W., 1980. We need both exploratory and confirmatory. *The American Statistician* 34:1, p. 23-25.
- UN/ECE Task Force on Monitoring and Assessment, 1996.*Working programme 1994/1995. Guidelines on Water-quality Monitoring and Assessment of Transboundary Rivers* (volumes 1 to 5) Working programme 1994/1995 Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad.
- United Nations Statistical Commission and Economic Commission For Europe, 19 . Conference of European Statisticians, statistical standards and studies - No 39, Environment Statistics in Europe and North America. An experimental compendium. Part one: Time series data and indicators.
- Unie van Waterschappen, 1998. *Gegevensstandaard Water 1998.1*, november 1998, eindrapport (in Dutch). See also http://www.adventus.nl/adventus/index.htm [last accessed 05-05-2000]
- Unwin, A., 1994. REGARDing Geographic Data. In: P. Dirschedl and R. Osterman (eds.), *Computational Statistics*, Heidelberg: Physica Verlag, pp. 345–354.
- Unwin, A. and P. Fisher, 1998. *Case Studies of Visualization in the Social Sciences: An introduction*. On: http://www.agocg.ac.uk/sosci/casestudies/contents.html [last accessed 06-05-2000]
- Unwin, A., D. Unwin and P. Fisher, 1998. Exploratory Spatial Data Analysis with Local Statistics. ***The Statistician*****,** 47 (3), p.415-421. On: http://www1.math.uni-augsburg.de/~unwin/UnwinLocalStats/Unwin&UnwinLocalStats.html [last accessed 20-09-1999].
- Unwin, A., G. Hawkins, H. Hofman, B. Siegl, 1996. Interactive graphics for data sets with missing values - MANET. *Journal Computational and Graphical Statistics*, 5, p. 113-122. See also http://www1.math.uni-augsburg.de/Manet/ [last accessed 05-05-2000]
- Van der Molen, D., 1999. *The role of eutrophication models in water management*. Thesis Agricultural University Wageningen. RIZA report 99.020, ISBN 9058080439.
- Van der Straten, J.W.H., S. Semmekrot, R. Maasdam and H. de Haan, 1998. Regional WaterSystem Report in the picture (in Dutch). H_2O, (15), p. 17-18.
- Van Loon, W.M.G.M. and J.L.M. Hermens, 1995. *Monitoring Water Quality in the Future, Volume 2: Mixture Toxicity Parameters*. Research Institute of Toxicology (RITOX), Utrecht. ISBN 90-802637-2-9.
- Van Rooy, P.T.J.C., D. Anderson and P.J.T. Verstraelen, 1993. Integrated water management considers the whole water system. *Water Environment and Technology*, 5(4), p. 38-40.
- Van Rooy, P.T.J.C., 1995. Towards comprehensive water management in The Netherlands (2) bottlenecks. *European Water Pollution Control*, volume 5, number 6, p.33-40.
- Van Rooy, P.T.J.C and J. de Jong, 1995. Towards comprehensive water management in The Netherlands (1) developments. *European Water Pollution Control, volume 5, number 4, 59-66*.
- Van Rooy, P.T.J.C, 1997. *Interactieve planvorming voor waterbeheer* (in Dutch.). Proefschrift Technische Universiteit Delft.
- Vasiliev, I.R., 1997. Mapping time. *Cartographica*, vol 34, no2, pp 1-51.
- Velleman, P.F., 1995. *Data Desk 5.0, Data Description.* Ithaca, New York.
- Venice System, 1959. The Venice System for the classification of marine waters according to salinity. *Arch. Oceanogr. Limnol.,* 11, p. 243-245.
- Visvalingham, M., *1994. Visualization in GIS, cartography, and ViSC. In: Hearnshaw, H. M. and Unwin, D.J. (eds.). Visualisation in Geographical Information Systems*. New York, John Wiley & Sons, p. 18-25.
- Villars, M.T. 1995. *Monitoring Water Quality in the Future, Executive Summary*. Delft Hydraulics, ISBN 90-802637-6-1, 31p.
- Ward, R.C., 1979. Regulatory water quality monitoring: a system's perspective. *Water Resources Bulletin,* 15(2), p. 369-380.

- Ward, R.C., 1986. *Framework for designing water quality information systems*. Proceeding of the 2nd Scientific Assembly entitled "Monitoring to detect changes in water quality time series". Publication No. 157, International Association of Hydrological Scienses, Wallingford, Oxfordshire, UK, pages 89-98.
- Ward, R.C., J.C Loftis and G.B. McBride, 1986. The "data-rich but information-poor" syndrome in water quality monitoring. *Environmental Management,* 10(3), p. 291-297.
- Ward, R.C., J.C Loftis and G.B. McBride, 1990. *Design of Water Quality Monitoring Systems*. Van Nostrand Reinhold, New York, ISBN 0-442-00156-8.
- *Webster's Dictionary of the American Language*, 2nd College Edition, William Collins Publishers, INC.
- Wickens, C.D., D.H. Merwin, and E.L. Lin, 1994. Implications of graphics enhancements for the visualization of scientific data: Dimensional integrity, stereopsis, motion, and mesh. *Human Factors*, 36, p. 44-61.
- Wildberger, S., 1993. Salinity Testing Methods. *The Volunteer Monitor*, Vol .5, No.1, Spring 1993. On: http://www.epa.gov/OWOW/monitoring/volunteer/spring93/school27.htm [last accessed 05-03-2000]
- Wise, S., R. Haining and P. Signoretta, 1998. The Visualization of Area-based Spatial data, in the exploratory spatial data analysis of area-based data. On: http://ww.shef.ac.uk/~scgisa/vizrpt/vizrpt.htm [last accessed on 13-09-1999]
- Witmer, M.C.H., 1995. Information needs for policy evaluation. In: M.J. Adriaanse, J. van der Kraats, P.G. Stoks and R.C. Ward (eds.), 1994. *Proceedings of the international workshop Monitoring Tailor-Made I*, Beekbergen, the Netherlands, p. 55-61.
- Witmer, M.C.H., 1997. Monitoring for policy evaluation on the regional scale. In: Ottens J.J, F.A.M. Claessen, P.G. Stoks, J.G. Timmerman, R.C Ward (eds*.), Proceedings of the international workshop Monitoring Tailor-Made II*, 1996, Nunspeet, the Netherlands, p. 471-472.
- Witteveen en Bos, 1998. *Proefproject Regionale WaterSysteem Rapportage (RWSR) Friesland*. Witteveen en Bos by order of Stuurgroep RWSV Friesland, April 1998.
- Witteveen en Bos, 1999. *RWSR proefproject tweede fase* (in Dutch). Witteveen en Bos by order of IPO-RWSR, December 1999.
- Yu, C.H. and J.T. Behrens, 1995. Paper presented at the Annual meeting of Society for Computer in Psychology, Los Angeles, CA. On: http://seamonkey.ed.asu.edu/~alex/alignment/alignment.html [last accessed 08-04-2000]
- Zwart, D de., 1995. *Monitoring Water Quality in the Future, Volume 3: Biomonitoring*. National Institute of Public Health and Environmental Protection (RIVM), Bilthoven, The Netherlands. ISBN 90-802637-3-7.

A. Appendix

This appendix will contain a presentation of the case study.