Continuing Lines

Application of heuristic optimisation techniques for spatial environmental problems with multiple objectives

CONTINUING LINES

Application of heuristic optimisation techniques for spatial environmental problems with multiple objectives

Toepassing van heuristische optimalisatietechnieken voor ruimtelijke milieuvraagstukken met meervoudige doelstellingen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. W.H. Gispen, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 16 juni 2006 des ochtends te 10.30 uur

door

Cornelis Vink

geboren op 25 april 1960 te Putten

Promotor: Prof. dr. ir. M. F. P. Bierkens

Co-promotor: Dr. P.P. Schot

This doctoral study was carried out at the Faculty of Geosciences, research group Environmental Sciences, Utrecht University, the Netherlands. The European Union LIFE program and Delft Cluster funded parts of the research. RIVM-MNP, the Dutch national research institute for public health, environment and nature, Kiwa Water Research, the drinking water companies Waterleiding Maatschappij Limburg, Hydron Midden Nederland and the provincial authorities of Limburg are acknowledged for facilitating this research.

ISBN 10:	907383883 5
ISBN 13:	798907383883 3
Cover painting:	Darja Vos
Cover design:	Jeroen Vos
Printed by:	F&N Boekservice, Amsterdam

Contents

List of figures

List of tables

1. Introduction

1.1 Problem definition

Resource management in densely populated and technologically developed countries is faced with changing conditions, tending towards increased complexity. Some examples of these trends are:

- increasing scarcity of resources and hence increased need for efficient production;
- increasing number of objectives, such as those referring to ecological and environmental values and the quality of the landscape;
- increasing number of actors in the decision process;
- increasing rate of change of technological innovations.

The tendency towards increased complexity is clearly present in spatial planning in the Netherlands. Urbanisation and economic development have led to a reduced quality of the ecological status and landscape and to an intensified agricultural landuse. As a result, conflicts of interest between socio-economic, environmental, ecological and agricultural values have become more pronounced. Public authorities and stakeholders try to protect and enforce a wide range of values and interests that are present in different types of land use, often seeking opportunities to avoid conflicts with other objectives.

Groundwater protection zones are maintained by provincial authorities to ensure safe drinking water production, but are an obstacle for further urbanisation, development of industry, infrastructure and agriculture. Agriculture is limited by many regulations, in particular if urban or nature areas are nearby. Space, unpolluted groundwater, natural vegetation and clean air have become scarce and therefore changes of land use have become complex, time consuming operations, where every expected benefit of change may come with costs for a wide range of other values.

The presence of multiple objectives combined with the spatial heterogeneity of landuse, geographical and hydrological conditions result in a highly complex system of relations. Scenarios that correspond to optimal fulfilment of objectives cannot be identified effectively by the 'unarmed' human mind. As a result, the risk of choosing suboptimal scenarios has become more pronounced. Choosing suboptimal solutions to spatial planning problems implies that costs are higher than necessary. These higher costs are not restricted to economic values but concern a wide range of values; environmental, ecological or other.

1.2 Hypothesis

At least a part of the solution of the problem described in the previous section can be found in using better tools and techniques in spatial planning. The use of advanced numerical techniques for processing detailed spatial information is required for efficient spatial planning and resource management. The use of decision support systems that are based on scientific models and apply efficient optimisation techniques can compensate for the increased levels of complexity in present day decision making. More specific, application of heuristic optimisation techniques to complex spatial environmental problems with multiple objectives can improve the identification of (near) Pareto-efficient solutions and thus contribute to better, more efficient decision making.

1.3 Objectives

The main objective of this thesis is to assess whether the use of heuristic optimisation techniques can lead to better decision making in complex environmental problems.

Particular goals of this thesis are:

- to investigate if and how heuristic optimisation techniques, in particular genetic algorithms, can be applied successfully to a suite of complex optimisation problems in the environmental and hydrologic fields;
- to investigate how results of heuristic optimisation techniques can be validated more effectively;

- to investigate how optimisation problems with multiple objectives can be handled effectively.

1.4 Structure of the thesis

This thesis is structured in three parts, A, B and C.

Part A contains an overview of optimisation techniques and decision support systems in the context of environmental problems as an introduction to the subject of this thesis (Chapter 2).

Part B (Chapter 3 to 6) consists of four case studies, each focussed on a specific optimisation problem in the field of environmental assessment and decision making.

Chapter 3 deals with the multiple objective parameter optimisation for calibration of a groundwater model by means of a genetic algorithm. Calibration of numerical models of groundwater flow and transport is a complex optimisation problem, due to the large number of parameters involved and the presence of nonlinear relations within many groundwater models. Identification of a unique best parameter set is a problem as many parameters are correlated. A genetic algorithm is applied for model calibration by formulating the parameter estimation question as a multiple objective optimisation problem. The selected objectives are minimal residuals (1) and minimal differences between initial and final parameter settings (2).

A case study on multiple objective optimisation of regional drinking water production with a genetic algorithm is described in Chapter 4. A genetic algorithm (GA) is applied as to master the combinatorial explosiveness and multiple objectivity of the optimisation problem. The performance of the GA is partially validated by applying it to a hypothetical and simplified type of problem that enables comparison of the results with those of an alternative optimisation procedure.

In Chapter 5, the use of a GA for multiple objective optimisation of land use is demonstrated. A genetic algorithm is applied to identify the optimal spatial allocation of nature and agriculture. Nature suffers from nitrogen deposition from agriculture, while agriculture is limited by regulations if it is located nearby nature areas. The performance of GA is partially validated by applying it to a hypothetical and simplified type of problem. The discussion is focussed on the required properties of optimisation problems that make it suitable to be solved with GA's. The application of more than two objectives is investigated. Various ways to validate optimisation results are discussed.

In Chapter 6 a new general framework for the prioritisation of groundwater quality prediction studies is presented. Prioritisation is viewed as optimisation over time. Prediction studies are comprised of design and update of groundwater quality monitoring systems and predictive modelling studies. Decision variables that express the uncertainty of predictions are introduced and the performance of various strategies is investigated by means of a sequential game approach. The study illustrates how decision makers can use advanced optimisation techniques to benefit from recent developments in numerical modelling, automated data acquisition and quantification of the uncertainty of modelling results. It is shown that apart from optimal *scenarios*, within an experimental game setup optimal *strategies* can be identified as well. It is shown that effective prioritisation of prediction studies requires techniques that enable assessment of expected uncertainty reduction due to additional research.

Finally, Part C (Chapter 7) contains the synthesis of this study. The main results of the research are summarized and discussed and general conclusions are drawn.

PART A: THEORY

2. Multiple objective optimisation and decision support in spatial environmental problems

2.1 Concepts and terminology

In this section a number of terms and concepts that are used frequently in this thesis and in the literature on optimisation will be described briefly.

2.1.1 Optimisation problems

In mathematics, optimisation is the discipline that is concerned with finding maxima and minima of functions that are usually subject to constraints. There are quite a few descriptions of the concept "problem". The word originates from the Greek expression "Pro Ballo", that can be translated into English as "before/forward throwing". It can be seen as an "inner conflict", an "inner discrepancy" that demands to be solved or eliminated. Mintzberg et al. [1976] provides a description of problem solving with three major phases of decision-making: identification (of the problem), development (of alternative solutions) and selection (of the best solution). Within this framework, the contents of this thesis is focussed particularly on development and selection.

Spatial optimisation problems are a subdivision of optimisation problems, with at least one property that represents a degree of freedom in the spatial dimension, i.e. a choice of location.

2.1.2 Spaces

Quite a number of concepts in optimisation theory are coupled to the word "space": *decision space*, *search space*, *impact space* and *solution space*. The term "space" derives from the fact that each variable is seen as a vector in one dimension. All variables/vectors span a n-dimensional space in which all possible combinations of variables ly. These spaces refer to choices and consequences; choices reside at the domain side, consequences of choices on the range side. *Decision space*, which is also called *search space* or *parameter space* refers to the domain side of a problem. *Impact space*, also called *solution space*, refers to the range side of a problem. The number of dimensions of *parameter space* corresponds to the number of decision variables. The number of dimensions at the range-side of the problem corresponds to the number of *indicators* (see next section).

2.1.3 Objective categories, criteria, indicators and objective functions

Objective categories express the general objectives of an optimisation problem. Each *objective category* may be expressed by one or more *criteria*. The term *'objective category'* is no standard term in optimisation theory. It is used in this thesis to express the difference between a general, usually 'lumped' value and specific, operational instances of such a value. For instance, the objective category "environmental quality" may consist of the *criteria* rarity, biodiversity, presence of pollutants and the aesthetic value of the landscape. For every criterion one or more *indicators* can be determined as to express (semi) quantitatively in which degree a certain objective is fulfilled. An *objective function* describes the relation between one or more properties of solutions and indicator(s). Typically, in a simulation–optimisation approach, indicators are the output of impact (simulation) models. For instance, the criterion rarity could be measured by the number of species in a particular scenario that occurs on the IUCN red list of threatened species as an indicator. Biodiversity as a criterion could be expressed as the average number of species per hectare as an indicator.

2.1.4 Impact model

An impact model is a dose-response model that can be used to simulate scenarios and thus enable an assessment of the consequences of decisions. An impact model may for instance predict changes of species abundance, atmospheric deposition or groundwater quality as a result of particular decisions. The indicator values are the final output of impact models that are used for the optimisation.

2.1.5 Complex optimisation problems

In this section a number of properties of optimisation problems will be described that make them complex, i.e. difficult to solve. There are five basic properties of impact models that determine the complexity in environmental and spatial optimisation problems:

- nonlinearity;
- discontinuity;
- interdependency;
- feed back loops;
- combinatorial explosiveness.

Nonlinearity of dose-response relations plays an important role in many physical, chemical and biological processes. The nonlinear relation between physical or chemical conditions and plant species abundance is an important issue in many environmental optimisation studies. Typically, there is a clock-shaped nonlinear optimum curve, which results in a variable sensitivity of plant species abundance as a function of changes in physical chemical conditions (see Figure 2-1).

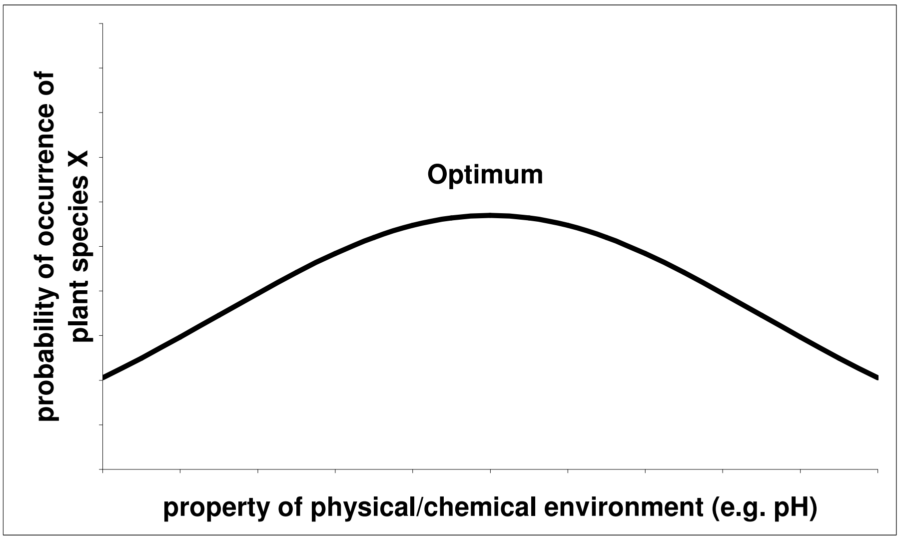

Figure 2-1 Example of a nonlinear dose-response relation optimum curve

Discontinuity is like nonlinearity a property that makes it difficult to solve optimisation problems. It occurs in many processes and consists of sudden leaps in impact functions, or sub-domains where no feasible solution exists. An example of discontinuity in biological processes is mortality of species; in technical processes it may be related to the maximum capacity of a transport pipe or groundwater pump.

Interdependency in optimisation problems refers to the phenomenon that many processes depend on more than one single variable and that the influence of these variables on a particular impact cannot be superposed. The response of one variable depends on the value of another variable. This feature is illustrated in Figure 2-2, where three curves describe the relation between pH of the soil and probability of occurrence of a hypothetical plant species for three different groundwater regimes. It is shown that there is interdependency between the groundwater regime and the pH with respect to the probability of occurrence of plant species. The importance of pH to the probability of occurrence depends on the groundwater regime in a nonlinear way.

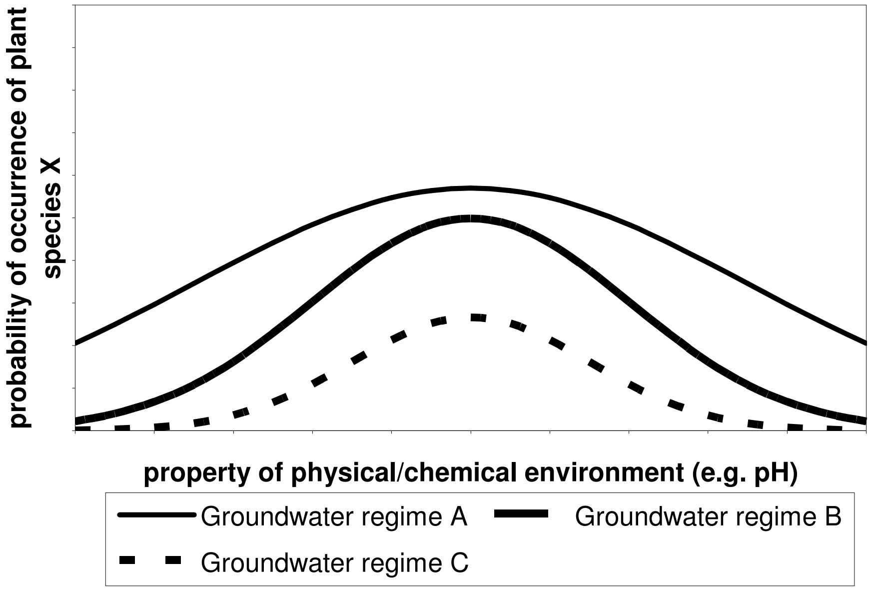

Figure 2-2 Example of interdependency: Conceptual optimum curves of wetland plants as a function of pH and groundwater regime.

Feedback systems [Stear, 1987] occur in many types of processes, and make it difficult to predict system behaviour. Small differences between the initial state of systems eventually may lead to large differences. Similarly, small errors in the description of the feed back relation may result in large errors in the predicted state because deviations accumulate. An example of a feedback system is the interaction between the transport capacity of a road network and the spatial distribution of population density. The transport capacity of a road network has an influence on the places where people decide to live. The resulting spatial distribution of residential areas influences the further development of road networks.

Combinatorial explosiveness is the term that describes the exponential increase of possible scenarios in systems with many degrees of freedom. It occurs for instance in network systems, such as regional drinking water production systems or in other spatially interrelated systems such as landuse allocation problems. The temporal dimension may also invoke combinatorial explosiveness, like in a game of checkers. An example of an optimisation problem with a time-based combinatorial explosiveness is presented in Chapter 6. An example of 'spatially based' combinatorial explosiveness is discussed in Chapter 4. The latter concerns different scenarios to meet the required regional drinking water production. More concretely, in a regional drinking water supply system of N wells of equal capacity, interconnected by a transport system of unlimited capacity, the total number of configurations is given by:

$$R = S^N \tag{2-1}$$

Where:

R	the total number of combinations
N	the number of wells
S	the number of discharge rate steps per well

2.1.6 Multiple objectives

Many real world optimisation problems have conflicting objectives. Indicators of the fulfilment of different objectives generally cannot be converted objectively to a common scale. Ecological quality for instance, cannot be expressed objectively in economic terms. Therefore, it is not possible to identify a single objective optimum. The problem of incommensurable objectives, also known as the 'problem of apples and pears', implies that (inter) subjective *valuation* is necessary to identify the optimum solution of a problem with multiple, conflicting objectives.

2.2 Valuation

Multiple objective optimisation needs to be based on valuation of impact categories. The valuation of impacts can be done by two different approaches:

- explicit approach, in which the impacts are all converted to a common scale;

- implicit approach, in which the valuation of impacts is expressed indirectly by formulation of constraints.

According to the *explicit* approach, impacts are translated into a common, often monetary scale. Conversion of damage to natural vegetation according to an explicit approach could for instance consist of valuations that are based on replacement costs. In that case the economic costs involved in creating a similar natural vegetation elsewhere would be the basis for valuation of the impact. This approach is being criticized for its inability to reflect all relevant aspects in a meaningful way [Nijkamp, 1979]. However, there are objective categories on which stakeholders have managed to agree on a translation of impacts into monetary terms. Explicit valuation on a basis of 'willingness to pay' of individuals offers good possibilities to differentiate between various kinds of environmental capital and services, particularly if the valuers are well informed. Over the last decade a considerable number of authors have applied explicit valuation by assessing 'how changes in the quantity or quality of various types of natural capital and ecosystem services may have an impact on human welfare' [Costanza et al., 1997].

Alternatively, an *implicit* approach for value attribution could result in formulation of constraints of the type: "the maximal drawdown induced by pumping may not exceed X cm in area Y". The latter approach offers a better possibility to formulate absolute boundaries to which solutions should comply than translation into monetary terms as in the explicit approach. On the other hand such a rigid valuation can seriously hamper the finding of compromises. Both explicit and implicit approaches may be combined within optimisation.

Ecological impact models generally lack a comprehensive analysis of values that are represented in particular plant species and vegetation types. Witte et al. [1993] based the value of an ecotype group on its national and international rarity. Many other approaches to valuation of ecological entities have been proposed and/or applied. Valuation of (semi) natural vegetation and other types of environmental resources on a quantitative basis is an issue of continuing development [e.g. Costanza et al., 1997, Lazo et al., 1992, Whitehead et al., 1991]. Progress towards a consistent valuation system is hampered by the complexity of our global ecosystem, that makes it hard to determine the exact importance of particular kinds of environmental capital and services. Besides, the entire range of possible decisions (i.e. compromises) for a problem is often poorly known and the environmental impacts of decisions are frequently insufficiently clear. Examples of inconsistent attribution of values can be observed in the investments that are made for the conservation of nature: in some cases large amounts of money are spend to save relatively few phreatophytes where at a similar project nothing was spent for protecting wetland vegetation and yet the net benefit of these investments could have been much higher.

Optimisation approaches and decision support systems such as presented in this paper can facilitate the identification of possible compromises and make the corresponding impacts better known. The weight or value that is attributed to impacts by decision makers can thus become more transparent. This could contribute to the development of a more consistent valuation system and thus it could help to allocate investments for the conservation of environmental values where they are needed most.

2.3 Pareto fronts

Decision makers often prefer to inspect the total collection of feasible and rational solutions before deciding on valuation of objectives. If decision makers cannot agree a priori on an either explicit or implicit valuation of objectives, then the interdependence between conflicting objectives can be evaluated by means of Pareto fronts (trade-off curves).

Already in the beginning of the 20th century the political economist Pareto studied optimisation problems [Pareto, 1906]. He introduced the terms *inferior* and *noninferior* solutions. A choice between different noninferior solutions is impossible without valuation, but inferior solutions can be rejected without valuation. The term noninferior is equivalent to *Pareto efficient* or *nondominated.* It is a relative property: the solution is noninferior if there is no other solution that performed better for one objective without performing worse for any other objective.

An example of Pareto-efficiency:

If a person needs to buy tea, he can choose to buy expensive tea of relatively good quality, but he can also choose to buy a relatively cheap tea of a lesser quality. A choice between these two options cannot be made on rational grounds, as quality of tea cannot be expressed in financial terms without subjective valuation. However, if there exists a third option, to choose a tea that is even more expensive than the good quality tea but of a lesser quality, then this option can be discarded on rational grounds, without valuation. No rational person would ever choose to buy this expensive tea of relatively poor quality, provided that only price and quality are relevant properties. The first two options are called Pareto-efficient, Pareto-optimal or noninferior solutions, the third option is called 'inferior' or 'dominated', because it can be rejected without valuation.

The complete set of noninferior solutions is called *Pareto front*. The decision-making can be facilitated by discarding all inferior solutions. Only nondominated solutions are to be evaluated by decision makers. If there is an optimisation problem with two conflicting objectives, then the Pareto efficient solutions jointly represent the Pareto front if the indicator values of the solutions are visualised graphically in a chart.

Pareto fronts can be used as reference information for decision makers as they mark how the best solution depends on valuations of incommensurable objectives. The value of the Pareto approach for multiple objective optimisation lies in "providing a set of alternative options for system improvements rather a single prescriptive solution" [Azapagic, 1999]. Pareto fronts can be used to show how economic cost and environmental damage are interrelated for Pareto efficient solutions. If a particular budget X is available for alleviation of environmental damage, then the solution that corresponds to optimal allocation of this budget can be identified at the Pareto front (Figure 2-3). The slope of the curve indicates the marginal reduction of environmental damage as a function of economic investments.

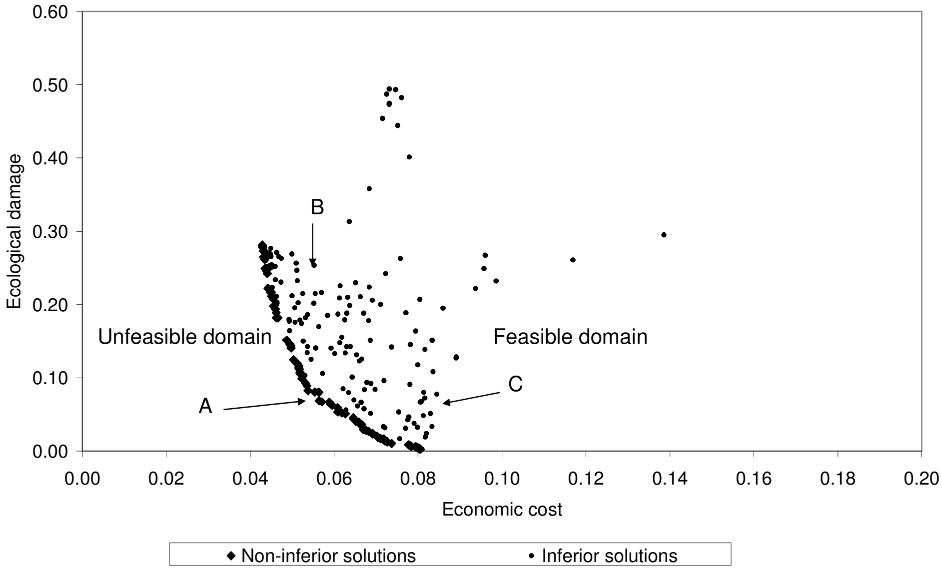

Figure 2-3 Example of a Pareto front of two conflicting objectives

Solution A is Pareto-efficient (noninferior) because there is no other solution with both lower economic costs and less environmental costs. Solution B implies similar economic costs as solution A, but the ecological costs are higher. Solution C implies similar ecological costs as solution A, but the economic costs are much higher. The feasibility of solution A implies that solutions B and C can be rejected on rational grounds.

2.4 Overview of optimisation techniques

The following sections contain a concise description of the principal optimisation techniques that are suitable for solving complex spatial optimisation problems.

2.4.1 Mathematical programming

A mathematical programming problem is an optimisation problem that can be formulated as:

Maximise f(x): x in X, g(x) <= 0, h(x) = 0 (2-2)

Where X indicates a domain that delimits the search space of the variables that are to be optimised. The relations g(x) <= 0 en h(x) = 0 are constraints and f is the objective function. The functions f,g and h need to be explicit mathematical functions.

There are various methods and techniques to solve problems of the type that was formulated are above, of which linear programming (LP) is the most popular. Linear programming is suitable for a specific class of mathematical problems where the objective function and the constraints are linear. There is a wide range of optimisation problems that can be formulated as a linear function, or that can be approximated by a linear function. Generally, linear programming is a fast technique, even if the number of variables is quite large. The method was developed in the 1940s by G.B. Dantzig, who developed the so-called simplex method [G. B. Dantzig, 1963] and by J. von Neumann, who presented it in the same period his theorem of duality [Neumann and Morgenstern, 1944]. In 1984 N. Karmarkar introduced the so-called interior–point method that in many cases could locate optima much faster than the simplex method [Karmarkar, 1984]. This search technique does not follow the corner points like the simplex method, but follows interior points within the search space. Essential to both methods is the solution of systems of linear equations where techniques are applied that originally were developed by Lagrange, Gauss and Cholesky, back in the 18th and 19th century. In 1975 the Russian mathematician L. Kantorovich and the American economist T. Koopmans received the Nobel Prize for economy for their contribution to the "Optimal Resource Allocation Theory", in which linear programming plays a key role. At present, in many enterprises this technique is used as a standard tool, for instance for the optimal allocation of resources. There is a large spectrum of commercial software in which this technique is applied.

Linear programming has proved itself as a very powerful tool for modelling and solving a wide range of real-world problems. However, many optimisation problems cannot be linearised and therefore linear programming is inappropriate. An important group of these problems are intrinsically discrete, so-called mixed integer problems. MILP (mixed integer linear programming) problems occur in many practical situations and are many techniques suitable for solving these problems. The computational costs of solving MILP problems are much higher than of solving ordinary LP problems, particularly if the number of integer variables is high.

Quadratic programming (QP) can be used if the objective function is of a quadratic form. There are many examples of this type of optimisation problems and many techniques to solve them too, provided that the object function has a convex shape. However, solving quadratic optimisation problems is much more demanding in computational terms than solving linear optimisation problems.

Nonlinear programming (NLP) refers to a group of optimisation problems with nonlinear, non-convex objective functions. Some of these problems can be solved by approximative solution techniques. Although there are many examples of these kind of optimisation problems, it is generally uncertain whether they can be solved by NLP, if the number of variables is large.

If mathematical optimisation techniques are unsuitable, heuristic optimisation techniques are appropriate. Two major techniques are discussed here: genetic algorithms and simulated annealing.

2.4.2 Genetic algorithms

Over the last decades genetic algorithms have been applied successfully to optimisations problems [e.g. Cieniawski et al., 1995; Deb & Kalyanmony, 1999]. The development of genetic algorithms was inspired by the genetic processes of biological organisms. The concept of natural selection by survival of the fittest as stated by Charles Darwin in *The Origin of Species* plays a major role. Application of the principles of selection and mutation in computer programs was first proposed by Holland [1975].

Three principles of evolutionary theory are also essential in genetic algorithms:

- survival of the fittest;
- incomplete inheritance of properties;
- variation among individuals.

The first principle results in not only survival, but also reproduction of the fittest. The consequence of the second principle is that children-solutions resemble their parents-solutions but are not identical to them. The third principle enables a continuous exploration of alternative solutions.

GA's work with a 'population' of possible solutions to a problem. The 'fitness' of each member of the population is calculated and the properties (genes) of those who perform best are mixed with other solutions, leading to new members of the population. The 'cross-over' process has been implemented in many different varieties according to the specific character of optimisation problems. The reproduction cycle is repeated until there is convergence, in the sense that no further improvement of solutions occurs. The procedure of the optimisation is presented in Figure 2-4.

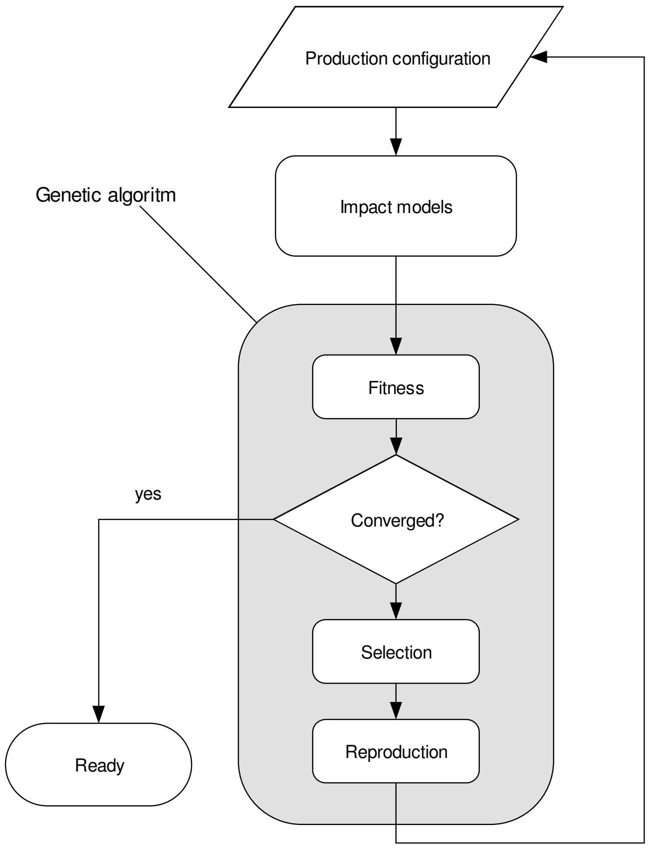

Figure 2-4 Flow chart of an optimisation procedure with a genetic algorithm

Typically two new solutions are reproduced by two existing solutions. There exists a large variety of reproduction techniques, that is the way new solutions are constructed out of existing solutions. *Arithmetic crossover* consists of calculation of a weighted average of numerical values that represent the various properties of the parent solutions. S*ingle point and multi point crossover* consist of random heritage of a particular part of the genes of both parents.

Many different crossover techniques have been applied in GA's. Comparative studies indicate that there is no large difference of final results or speed between them [Beasley et al., 1993]. Eshelman et al. [1989] performed an extensive comparison of different crossover operators and analysed them both theoretically and empirically but no overall winner emerged; speed differences did not exceed 20%.

For the hypothetical case studies in this paper we achieved best results with arithmetical and uniform crossover techniques. Arithmetical crossover was implemented as described in Michalewicz [1996]; genetic information of offspring is then determined according to:

$$U: u[1],..u[n] \qquad (2\text{-}3)$$

$$V: v[1],..v[n] \qquad (2\text{-}4)$$

$$X: x[1],..x[n] \qquad (2\text{-}5)$$

$$Y: y[1],..y[n] \qquad (2\text{-}6)$$

$$x[1],..x[n] = c1*u1+c2*v1,.., c1*u[n]+c2*v[n] \qquad (2\text{-}7)$$

$$y[1],..y[n] = c2*u1+c1*v1,.., c2*u[n]+c1*v[n] \qquad (2\text{-}8)$$

Subject to:

$$c1,c2 >=0 \qquad (2\text{-}9)$$

$$c1+c2=1 \qquad (2\text{-}10)$$

Where:

U,V	chromosomes of existent (parent) solutions
u,v	genes of parent solutions
X,Y	chromosomes of new solutions (offspring)
c1, c2	crossover constants (-)

Uniform crossover results theoretically in the most effective exploration of the search space and combines well with the 'hill climbing' character of arithmetical crossover that is effective in finding local optima. In uniform crossover, offspring are constructed by random inheritance of genes from either parent U or V. A reproduction results in two new solutions, of which solution Y is attributed the inverse genetic information of X, i.e. where X possesses a gene of U, Y is attributed the gene of V and vice versa.

Similar to the biological principles in evolution theory, *mutation* is applied as a supplementary principle. In evolution theory, mutation is a factor that prevents *genetic drift*, the tendency in a population to become genetically homogeneous. Similarly, mutation is applied in genetic algorithms as to avoid premature convergence by assuring sufficient variation within the population of solutions. However, a mutation probability that is too high limits the possibilities to inherit successful properties of parent solutions. Hence, there is a trade-off between variation and inheritance. As yet, there are no generally applicable methods to determine the optimal probability of mutation for specific optimisation problems.

Genetic algorithms are particularly suitable for multiple objective optimisation (MOO) because the presence of a population of solutions implies that there exists a collection of solutions [Goldberg, 1989]. Genetic algorithms have been applied to a wide range of optimisation problems and are viewed as a last resort for particularly difficult optimisation problems.

2.4.3 Simulated Annealing

The development of simulated annealing was inspired by the physical process of annealing, i.e. the cooling of metals. The first component of the optimisation procedure consists of the construction of alternative solutions from existing solutions according to a Metropolis Monte Carlo scheme [Metropolis et al., 1953]. The second component controls the selection of alternative solutions according to a Boltzmann acceptance criterion. Unlike the traditional Monte Carlo simulation, alternative solutions are constructed by modifying existing solutions. The degree of modification is large in the initial stages of the optimisation and is gradually reduced. The temperature of the physical annealing process is analogous to this degree of modification. The ‘cooling scheme’ refers to the gradual reduction of the ‘temperature’ and varies among different simulated annealing approaches. The existing solution is replaced by an alternative solution if the difference between the alternative solution and the current solution is within ranges that are defined by a criterion based on the Boltzman probability factor. The ranges vary along the optimisation as a function of ‘temperature’.

$$P = e^{\left(-\frac{E(current)-E(new)}{kT}\right)} \qquad (2\text{-}11)$$

Where:

P	Boltzman probability factor
E (current)	current 'energy level' (i.e. values of the objective function)
E (new)	alternative 'energy level'
k	Boltzmann constant
T	temperature

The alternative solution is always accepted if it has a better performance than the current solution. If it leads to a worse performance, it will only be accepted if a randomly generated number (between 0 and 1) is less than or equal to the Boltzmann probability factor.

Simulated annealing has been applied successfully to many, generally combinatorial optimisation problems [Azencott, 1992].

2.4.4 Classification of techniques

One of the possible classifications of optimisation techniques divides the available techniques in two groups: heuristic versus 'classical' , or local versus global. Within both groups there are methods that can be applied to complex spatial optimisation problems. The principal global optimisation methods consist of linear and nonlinear programming techniques, whereas genetic algorithms and simulated annealing belong to the most prominent heuristic techniques.

The heuristic methods possess a typical iterative approach where the search consists of a sequential evaluation of alternative solutions. Global optimisation techniques typically consist of formulating and then solving a set of equations that refers to all possible solutions. The specific advantages and disadvantages of the two groups are presented in Table 2-1.

Table 2-1 Differences between local and global optimisation techniques

A. Heuristic (local) techniques	**B. ' Classical' (global) techniques**
No guaranty that the found solution is the global optimum.	Often better guaranties that a found optimum is global.
Tuning[1] required for complex problems.	Requires generally no tuning.
The optimisation requires less preparation than for B.	The optimisation requires generally more preparation than for A.
Operates 'outside' impact models.	Either operates "within" impact models (strictly analytical), or operates on the outside, by extrapolation.
Constraints can easily be incorporated in the procedure.	Constraints cannot always be easily incorporated in the procedure.
Paralellization can be implemented relatively easily.	Paralellization cannot be implemented easily.
Availability of a large collection of possible solutions is inherent to the techniques.	Availability of a large collection of possible solutions is not inherent to the techniques.
Interdependency can be treated relatively well.	Interdependency cannot be treated easily.
Discontinuity can be treated relatively well.	Discontinuity cannot be treated easily.

[1] The term *tuning* refers to adaptation of optimisation codes to case specific needs.

2.5 Game theory and experimental game models

Over the past 50 years, game theory has provided successful techniques for finding optimal strategies in situations where there is uncertainty about the result of decisions. It was pioneered by Von Neumann and Morgenstern who in 1944 published their book "Theory of Games and Economic Behaviour". It has been applied to many subjects in economic and political sciences.

The work of von Neumann and Morgenstern demonstrated that game theory can be applied to decision processes in which the results depend on the strategies of two or more persons with conflicting objectives. Not much later it became accepted that, even if there is no interaction between persons involved, many optimisation problems can be viewed as games too [Wald, 1950]. Uncertainty with regard to physical, chemical or biological processes can be treated similarly, to some extent, as in human processes in economic problems. In this case, the decision maker plays against 'nature' in stead of human players. Nature is then viewed as a player who, unlike in games with human players, is not necessarily focussed at maximisation of profits or of losses for it's opponents, but nevertheless is a player whose actions are uncertain.

A further development in game theory are experimental game models. In this branch of game theory, simulations are carried out with numerical models on computers. The earliest published article on an informal economic game experiment was by Chamberlain [1948]. The use of these games is now generally accepted in many research areas. Experimental games can be a powerful instrument for the structuring of decision problems and for developing creative decision options. The underlying processes are simulated, their uncertainty is quantified in statistical terms by using both hard data and more soft information and thus it becomes possible to estimate the outcomes of various strategies.

2.6 Decision support systems

There are many definitions of Decision support systems (DSS) available. A representative example is the following:

Decision Support Systems (DSS) are a specific class of computerized information systems that supports business and organizational decision-making activities. A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions." [Anonymus, 2005].

The development of decision support systems is closely connected to that of computers. From the 1960's onwards the computational power of computers came available to universities and companies, and since the 1980's personal computers became available to individuals. Parallel to this development, decision support systems were developed and used.

According to Sprague and Watson [1979], around 1970 business journals started to publish articles on management decision systems, strategic planning systems and decision support systems. The first International Conference on Decision Support Systems was held in Atlanta, Georgia in 1981.

If many different objectives and corresponding indicators are involved, it is convenient that some kind of interactive interface between the decision makers and the results of the optimisation calculations is available, where results of alternative solutions can be presented in a coordinated way. If there are different actors involved in a decision process, coordination and interactivity becomes even more useful.

Decision support systems can facilitate the identification of possible compromises and make the corresponding impacts better known. The weight or value that is attributed to impacts by decision makers can thus be made more transparent. This could contribute to the development of a more consistent valuation system and thus it could help to allocate investments for the conservation of environmental values where they are needed most.

2.7 Optimisation of spatial allocation problems

Optimisation of spatial allocation problems has not been studied extensively for a long period of time, but over the last two decades, the number of publications in scientific journals has increased significantly. Many complex spatial allocation problems are related to finding optimal landuse configurations, as occur in timber harvesting, agricultural production [Mayer et al., 2001], grazing management plans [Horton, 1996], crop rotation [Parsons, 1998] and ecological modelling [Haberlandt et al., 2002]. The spatial dimension introduces often a large search space due to the high degree of combinatorial freedom. In some cases the temporal dimension is involved too, such as in crop rotation or in ecological modelling where succession of vegetation types is included.

Mayer et al. [2001] have published on the optimisation of agricultural production on farm level. The authors applied and compared a number of techniques. They conclude that simulated annealing and genetic algorithms are the most appropriate techniques to solve these complex spatial allocation problems with a large search space and nonlinear and discontinuous system properties. Mayer et al. [2001] consider hill climbing, direct search algorithms, the Nelder-Mead simplex methods and the tabu search metastrategy inappropriate for the task of these types of model optimisation. They state that the remaining two general families of optimisation methods, viz. simulated annealing and evolutionary algorithms have proven valuable in this field. Simulated annealing has been used successfully to identify the economic optima of a range of agricultural systems. However, the rates of convergence have been problematic for some problems, which then leads to excessive runtimes for achieving the global optimum. A number of comparisons between optimisation algorithms were conducted on mathematical test functions, with varying results. In general, Mayer et al. [2001] consider modern heuristic optimisation methods like simulated annealing and evolutionary algorithms to belong to the most powerful techniques, although other methods can perform well on specific types of problems. The mathematical test problems are different from the optimisation of management strategies in simulation models, which makes it difficult to analyse the effectiveness of the various techniques. Complex simulation models form one of the more difficult classes of optimisation problems They are typically of higher dimensionality, as each individual management option to be optimised contributes an extra dimension to the search-space. The impact space (response surface) that is generated by simulation models can be nonlinear and include discontinuities. In the cases that were investigated by Mayer et al. these nonlinear and discontinuous properties occur when the agricultural systems are pushed 'too far' (e.g. overstocking) and biologically (near) unfeasible parts of the search space are entered. The complexity of ecological and biological system models usually implies the presence of nonlinear, discontinuous and interdependent system properties, whereas classical engineering optimisation problems typically can be described with a limited number of differential equations that can be solved by linear and nonlinear programming. However, this complexity also makes it rather difficult to compare different ecological and biological optimisation problems. Mayer et al., 1991, 1996] showed that the simplex method performed better than gradient methods on a dairy farm model, but found both to be inferior to genetic algorithms and simulated annealing.

Seppelt and Voinov published in Ecological Modelling [2002] about spatial optimisation in environmental modelling. They tried to identify the optimal distribution of land use for the best ecosystem management and developed a genetic algorithm that was coupled with a GIS. Both the mathematical formulation and the techniques that were used vary widely in the aforementioned studies, which makes it difficult to determine general criteria for the choice of the most suitable optimisation technique.

Seppelt [2000] considers two aspects that play a role in the complexity of spatial optimisation problems:

- model complexity;
- spatial complexity.

Model complexity is a function of the type and the number of mainly biological and physical processes involved, whereas the spatial complexity is determined by the size of the study area, the grid cell size and the number of spatially interacting processes. The optimisation problem that Seppelt studied concerns the allocation of 7 different types of land use in a watershed: “Our goal is to find out what is the optimum land use pattern and what should be the strategy of fertilizer application to reduce nutrient outflow out of the watershed and increase yield”. The problem was analysed with a spatially distributed model, simulating various biological, hydrochemical and agronomical processes. The objective function contained both economic and ecological criteria, where weights were applied as to merge these criteria in a single objective function. The complexity of the process models involved proved too high to allow an integrated regional optimisation. Therefore, the authors tried to extend the results of a local optimisation to a regional domain through bivariate correlation analysis. However, absence of strong correlations between local spatial properties and local optimal solutions inhibited successful upscaling of local optimal solutions to a regional scale.

Haberlandt et al. [2002] tried, like to Seppelt and Voinov, to upscale the results of process-based models from a local scale to a regional scale as to reduce the cost of the calculations. They used a numerical model to simulate nitrate leaching from agricultural areas. The results of the process model were taught to a fuzzy rule system per soil class. Simulated annealing was used for the selection of the best system of a fuzzy rules. The authors reported that the upscaling technique worked satisfactory and that the differences between the fuzzy rule system and the original process model were small. Average correlation coefficients between the two models varied between 0.78 and 0.94.

Brookes [2001] applied a genetic algorithm for optimal landuse configuration. He defined the optimisation problem of allocation of spatial functions to an area as an optimal patch design problem. The optimisation problem consisted of the allocation of two land use functions in Nepal, namely agriculture and carpet industry and concerned multiple objectives. He introduced a set of specific spatial concepts such as patch size, shape, connectivity, composition and configuration for the spatial properties of ecological and physical processes such as erosion, sedimentation and runoff. He also developed specific spatial genetic operators such as creep, sum and average. Brookes concludes that the GA technique worked satisfactory but that the use of specific spatial operators was essential.

Zander and Kächele [1999] investigated optimal land use configurations for sustainable development. The authors stress that sustainability is a lumped concept, consisting of multiple objectives that are partially conflicting. Due to the lack of objectivity in the concept of sustainability they conclude that the best way to determine optimal land use configurations requires participation of stakeholders in the optimisation process. Negotiation, interactive game approaches and decision support modelling are viewed by the authors as the appropriate techniques to identify the optimal solutions. Sustainability is interpreted as consisting of basic economic and ecological components. The optimisation problem is solved by application of linear programming techniques. Application of different weights to the components of the objective function enables the construction of a trade off curve. The optimal solution is located at the point of intersection between the trade off curve and an indifference curve, that reflects a stakeholder's preferences. The optimisation problem is approached as a pricing problem in classic economic theory, where the point of contact between the two curves represents market equilibrium. Different stakeholders have different indifference curves and therefore need to negotiate in order to agree on a mutually accepted compromise.

2.8 Optimisation for water management

Many studies have been dedicated to the use of optimisation models to solve the increasing complexity of water management problems. Linear programming [Aguado et al. 1974, Molz and Bell 1977, El Magnouni et al. 1994], nonlinear programming [Gorelick et al. 1979, 1984; McKinney et al. 1992], dynamic programming [Makinde-Odusola and Marino, 1989; Andricevic 1990], goal programming [Rajabi 1999] and genetic algorithms [McKinney et al., 1994; Cieniawski et al. 1995] have been used as optimisation techniques. Linear programming is an efficient technique for problems that can be linearized but becomes inaccurate and inappropriate for pronounced nonlinear and interdependent optimisation problems. Nonlinear programming techniques cannot handle interdependency either and can be unpractical due to the problematic determination of gradients if impact relations are highly nonlinear [McKinney, 1994]. Rajabi et al. [1999] applied a modified goal programming (GP) technique to a multiple criteria water supply planning problem, for the first time with interdependency between and within impact relations. A disadvantage of GP is that it requires valuation of criteria before final solutions can be presented. Decision makers need therefore to decide on valuations while the impact of these valuations on the final solution is unclear. Dynamic programming (DP) was developed for the optimisation of sequential decision processes. This technique for multistage problem solving may be applied to problems that can be described as a nested set of sub-problems. Pronounced interdependency among impact relations combined with a large search space makes DP infeasible. Optimisation by genetic algorithms (GA) is one of the few techniques that are capable of handling the highly nonlinear, interdependent and non-convex type of problems that often occurs in groundwater management. GA's have been applied successfully to a wide range of problems, but applications to water management problems are relatively scarce as yet. McKinney et al.[1994] illustrated how GA could be applied successfully to single-objective optimisation of well field development and aquifer remediation design. Cieniawski et al.[1995] applied GA to a two-objective (independent) groundwater monitoring optimisation problem and compared the GA technique with simulated annealing. Optimal groundwater remediation strategies were determined using this approach [Kuo et al., 1992].

2.9 Conclusions

From the previous sections it can be concluded that heuristic optimisation techniques and particularly genetic algorithms have been used successfully to solve a wide variety of spatial optimisation problems. The complexity of the models that are used follows from the complexity of environmental processes that are simulated. It tends to be further increased by a distributed representation of the spatial and temporal dimensions and thus frequently leads to nonlinear, interdependent and combinatorial explosive optimisation problems. Generally, every objective needs to be assessed by a separate, often complex model. As a result, appropriate optimisation techniques need to be used in a modular way "on the outside" of the models, as it is generally not feasible to reduce this class of optimisation problems to a limited set of partial differential equations. Heuristic optimisation techniques such as simulated annealing and genetic algorithms allow such a modular approach. Therefore, they can be viewed as 'last resort' techniques that may be able to tackle a problem where classical global optimisation techniques fail. In some cases, authors report that GA and SA were used successfully, but there are also quite some studies where tuning of algorithms and use of additional techniques were necessary. Examples of these adaptations are the design of specific crossover procedures [Brookes, 2001] and reduction of the computational costs by gradually increasing the level of detail of impact models [Haberlandt et al., 2002, Seppelt and Voinov 2002].

In spite of the substantial success of heuristic optimisation techniques there are a number of issues that are unsolved and require more attention:

1. It is unclear how the complexity of an optimisation problem can be assessed on beforehand as to determine whether it can be solved with GA's

2. Directly related to the issue mentioned here above and even more important is the unsolved question how to determine whether solutions that are found with heuristic techniques are true global optima

3. A third issue that deserves more attention is related to the valuation of objectives. It is unclear how two and more objectives can be handled effectively and unbiased in a multiple objective setting.

These issues will be investigated in the next chapters of this thesis.

PART B: CASE STUDIES

3. Multiple objective parameter optimisation of a groundwater model by means of a genetic algorithm[2]

3.1 Abstract

The parameter optimisation (calibration) of a groundwater model is approached as a multiple objective optimisation problem with two separate objective functions. The first objective function concerns minimum difference between simulation results and observations, the second objective concerns minimum deviation from initial estimates of the parameter values. A trade-off curve of Pareto-efficient solutions is constructed by computing multiple model runs for a genetic algorithm linked to Modflow. The genetic algorithm provided a stable and flexible technique for the parameter optimisation. Forms of “circumstantial validation” are applied to confirm the hypothesis that the identified solutions are global optima. Inspection of the relation between both objective functions in the trade-off curve enables a better understanding of the search space and a better founded choice for a final calibrated model.

Key words: groundwater; optimisation; parameter estimation; calibration; genetic algorithm; evolutionary programming

3.2 Introduction

Numerical groundwater models are useful instruments for assessing the impact of changes in the hydrological system on groundwater levels, flow patterns and chemical composition of the groundwater. However, the

---

[2] Adapted from: Proceedings of the FEM_Modflow conference of 2004 in Karlovy Vary, Czech Republic.

calibration of groundwater models is a difficult task. Three major problems exist:

Parameter value uncertainty. The extent to which parameter values can be modified from initial estimates to achieve maximum “model fit” is unclear.

Combinatorial explosiveness. The number of combinations of possible geohydrologic model properties or parameter values is enormous. Finding the combinations of parameter values that lead to best model fit is a complex task.

Identification problem. Uncertainties in parameter values and correlations between parameters result in the identification problem: different combinations of parameter values can result in similar model performance. But, the impact of simulated changes in the hydrologic system may differ significantly between different calibrated models with similar residuals and it is unclear which set of calibrated parameters give the best assessment of the impact of changes in the system.

3.2.1 Parameter value uncertainty

As model schematisation is never perfect and reference data used for calibration to some extent are also uncertain, parameter values should not be changed unlimitedly during the calibration process. Geostatistical techniques such as *kriging* have been developed in order to estimate probability density distributions of spatially distributed model variables and offer a theoretical solution to the question which relation exists between probability and distributed parameter value estimates. Uncertainty can thus be estimated and the modeller can determine the likeliness of different sets of parameter values, provided that interdependency among model parameters should be absent or known. However, kriging and many other statistical techniques are not very suitable to take spatial discontinuities of geohydrologic variables into account. Differences among sub-areas that are caused by the variety of geophysical processes that shaped the present geohydrologic properties imply different properties of the presupposed statistical models. The obstacle of parameter uncertainty in model calibration can be reduced by the aforementioned techniques, but remains a serious problem in practice.

3.2.2 Combinatorial explosiveness

Present day numerical groundwater models may contain more than 1 million cells and in many of these cells more than 3 static parameter values may be allocated. If the model is meant to simulate non-stationary processes the search space is even larger. Fortunately the number of likely combinations is strongly reduced by spatial correlation of distributed model parameters, but still it is unfeasible to assess all likely parameter configurations by exhaustive exploration, i.e. running the model for all these realisations. Zoning of spatially distributed parameters is a practical way to reduce the number of combinations and is effectively facilitated in recent releases of Modflow. However, defining zones implies simplification, which may reduce the reliability and accuracy of model results.

3.2.3 Identification problem

The identification problem is that more than one single solution of the parameter optimisation problem yields a best fit. If small differences in "model fit" exist among different configurations of model parameters then these are usually insignificant in relation to the precision of the input data. However, the results of scenario simulations may differ significantly. The best way to reduce the size of the identification problem is the introduction of more reference data and objective functions. For instance, adding water balance criteria to hydraulic head criteria may reduce the problem. However, ending up with sets of parameter values that are equally likely is often inevitable. The degree to which these solutions differ is strongly related to the number and precision of empirical data available.

Parameter optimisation techniques can be used to find optimal solutions, provided that the search space is reduced to a feasible size. Both classical "analytical" and heuristic techniques have been used for "automatic calibration" of groundwater models. The first group consists of linear and nonlinear programming techniques, the latter of Monte Carlo, simulated annealing and evolutionary programming (genetic algorithms).

Many hydrologists have carried out repetitive model runs while applying stochastic variation of parameter values according to uniform or other probability distributions as to quantify the range of possible model results and thus be able to assess model precision. The Monte Carlo approach is a great improvement on manual calibration techniques that were applied in the early days of Modflow, but the disadvantage of the method is that it demands much computer time, particularly for complex problems with many uncertain parameters that are a typical for present-day distributed models. Dettinger and Wilson [1981] applied the first-order second moment method to assess model output uncertainty as a function of input uncertainty and model sensitivity. The uncertainty that is associated with the probability distribution that is assumed for model parameters remains problematic. Doherty [2002] and Hill [1992] have developed very useful software in which nonlinear optimisation techniques are applied for parameter optimisation of Modflow and other models. These nonlinear parameter optimisation techniques are less expensive in terms of CPU time than Monte Carlo techniques. However, they are not particularly suitable for multiple objective optimisation problems and may sometimes become numerically unstable when applied to complex models. To avoid these problems, Zheng [1997] used a genetic algorithm, for single objective parameter estimation of Modflow models. In this paper I extend Zheng's approach to provide a practical technique for applying a genetic algorithm to model calibration where the parameter optimisation is approached as a multiple objective optimisation problem.

3.3 Methods

Over the past 20 years, genetic algorithms have been applied successfully to multi-objective optimisation problems [e.g. Cieniawski et al. 1995]. The development of genetic algorithms was inspired by the genetic processes of biological species. The concept of *natural selection* by survival of the fittest as stated by Charles Darwin in The Origin of Species plays a major role. At present, evolutionary programs have been applied successfully to a wide range of problems. GA's work with a 'population' of possible solutions to a problem. The 'fitness' of each member of the population is calculated and the properties (genes) of those who perform best are mixed with other solutions, leading to new members of the population. The reproduction cycle is repeated until there is convergence, in the sense that no further improvement of solutions occurs.

Here the method has been applied to the calibration of a numerical groundwater model. All parameters were allocated a calibration range, consisting of a parameter-specific minimum and maximum value. These range limits can be chosen on a basis of expert judgement, but also determined with a geostatistical or any other technique. The initial estimates of parameter values are located in the centre of the calibration ranges.

Two objective functions were defined for the calibration, both resulting in a fitness score:

1. Minimum differences among simulated and averages of observed hydraulic heads
2. Minimum deviation from initial estimates of parameter values

$$O_1 = \sqrt{\frac{\sum r^2}{n}} \qquad (3\text{-}1)$$

$$r = (\phi_o - \phi_s) \cdot \sqrt{\frac{\ell}{m}} \qquad (3\text{-}2)$$

$$O_2 = \sqrt{\frac{\sum \delta^2}{k}} \qquad (3\text{-}3)$$

$$\delta = \frac{V - V_i}{V_{\max} - V_{\min}} \qquad (3\text{-}4)$$

Where:

O	objective function
r	weighted difference between observed and simulated head
n	number of reference heads
ℓ	weight factor 1; number of available annual averages of head at a particular location
m	weight factor 2; distance to the focus of interest of the model
δ	the relative deviation of initial estimate
k	number of calibration parameters

ϕ_o	reference head
ϕ_s	simulated head
V	parameter value
V_i	initial estimate of parameter value
$V_{\max}$	maximal allowed parameter value
$V_{\min}$	minimal allowed parameter value

The second objective function is defined in such a way that configurations of parameter values become progressively more undesirable when they differ more from initial estimates, thus reflecting the presupposition that increasing deviation from initial estimates yields progressively increasingly improbable solutions. By taking both objective functions in account the identification problem is reduced in a pragmatic way.

3.4 Results

The optimisation was applied to a Modflow 2000 model of an area of about 100 km^2 near Hilversum, in the centre of The Netherlands. The hydrogeology of the area is strongly influenced during the Saalien glacial period and consists of sand and clay layers with highly heterogeneous hydraulic conductivities. The model consists of 7 layers and about half a million cells. It represents a subsoil of 200 m thickness in total. Zoning of recharge, based on landuse, and of hydraulic conductivity, based on the geological map, resulted in 120 parameters to be optimised. The initial and final stage of the optimisation process are presented in Figure 3-1 and Figure 3-2.

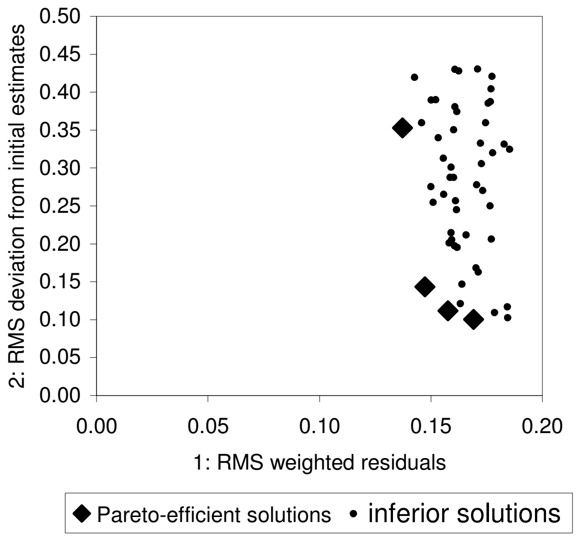

Figure 3-1 Solution fitnesses at the initial state of the optimisation

The initial fitnesses of the solutions with respect to the objective functions show still traces of a random character (Figure 3-1). By the end of the optimisation significantly better solutions are identified and a distinct Pareto front (trade off curve) can be recognised (Figure 3-2). The shape of the Pareto front (Pareto-efficient solutions) shows that both objectives are conflicting in the near-optimal impact space: increased deviation from initial estimates of parameter values (objective function 2) result in a better fitness for the first objective. The upper end of the Pareto front is located near a value of 0.2 for objective function 2. Further deviation of initial estimates of parameter values does not lead to a better model fit (lower value for objective function 1), since the upper side of the Pareto front ends here. Increasing the calibration ranges of parameters will not result in a better performance with respect to objective 1, unless the parameter values of the solution at the upper end of the Pareto front represent the minimum or maximum of the initially estimated range. As this was not the case, it can be assumed that further improvement of the model's "fit" (objective 1) can only be achieved by a reformulation of the model concept or the parameter definitions.

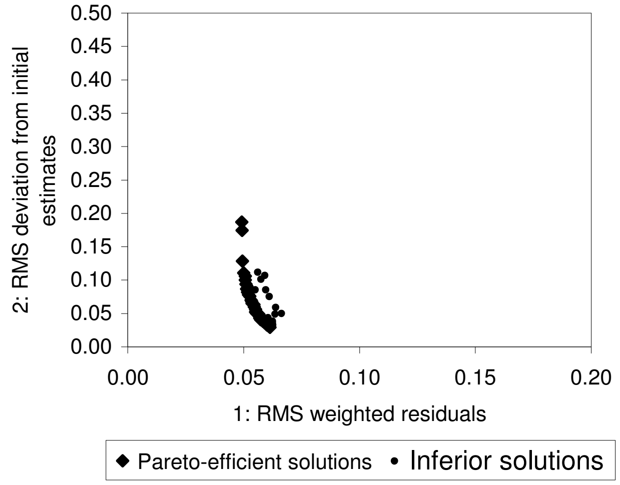

Figure 3-2 Solution fitnesses at the final stage of the optimisation

3.5 Validation

Whether or not better solutions exist remains in principle uncertain. However, inspection of optimised (i.e. calibrated) parameter values offers some degree of validation (circumstantial validation). Three examples are mentioned here:

- Calibrated values of *insensitive* parameters should be equal to the initially expected value V_i in Pareto efficient solutions. Adding a 'dummy', insensitive parameter for this type of circumstantial validation is therefore considered good practice. It enables checking whether the GA has operated correctly.

- The lower extreme end of the Pareto front should have a (near) zero value for objective function 2. It follows directly from the definition of objective function 2 that the global minimum equals 0 if all parameter values V equal V_i.

- Differences between simulated and observed piezometric heads should be consistent with differences between initially expected and optimised parameter values. In the parameter set that corresponds to the Pareto-efficient solution with the best model fit (objective 1), a reduction of the difference between expected and optimised parameter values should not result in an improvement of the result of objective function 1. If this is not the case, the GA is not working as it should or convergence is not yet achieved.

3.5.1 Dummy model test

A fourth form of circumstantial validation consists of 'dummy model tests'. The procedure for the 'dummy model test' consists of four steps:

1. Generate a random realisation of parameter values (D_i). The parameter values are constructed in such a way that initially estimated (parameter specific) ranges are not exceeded.

2. Calculate piezometric heads (DH_i) with the groundwater model realisation D_i, at locations that correspond to the locations where data of observed heads are available.

3. Optimise parameter values with the genetic algorithm where DH_i are used as *reference* heads.

4. Check whether the optimisation results in a (near) zero value for objective function 1 for the solution that corresponds to the upper end of the Pareto front.

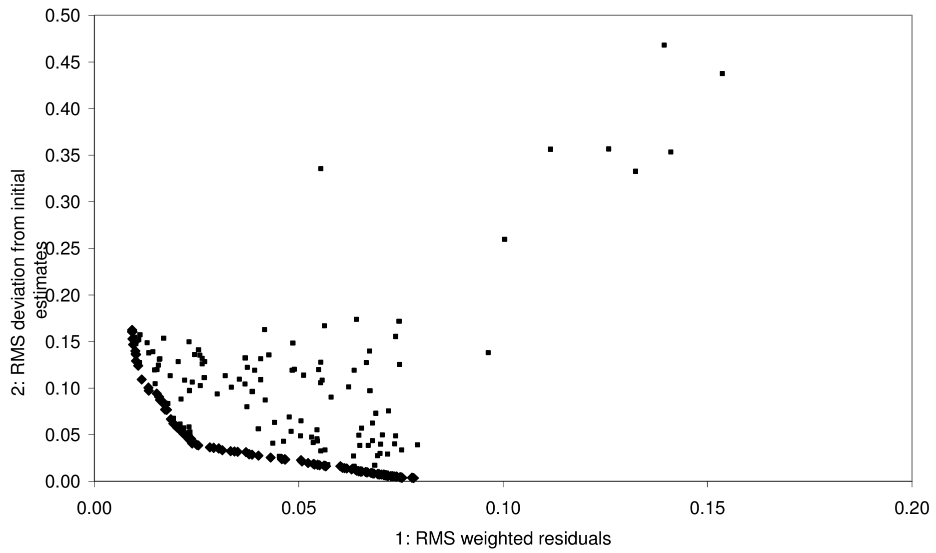

Figure 3-3 Solution fitnesses at the final stage of the dummy model test

The 'dummy model test' results indeed in a near-zero value of objective function 1 for the solution that corresponds to the upper end of the Pareto front (Figure 3-3). It implies that for this example the GA indeed identified a parameter configuration with an almost perfect model fit. The results with respect to objective 1 (model "fit") of the dummy model test (Figure 3-5) are better than those of the non-dummy calculations (Figure 3-4) because the dummy reference heads were *calculated* in stead of based on "real world" observations. Therefore, scale-induced errors, conceptual model errors, and errors in reference data are left out of the dummy model performance. Since the groundwatermodel represents an area with highly heterogeneous geohydrologic properties and in some parts thick unsaturated zones, scale-induced errors and conceptual model errors are expected to be of substantial magnitude. The representativity of the assessment of the magnitude of these errors can be improved by carrying out a sequence of dummy model tests, in stead of a single calculation, as is done in this study. However, the circumstantial validation of the performance of the GA is the principal purpose of the dummy model test in this thesis and is considered sufficiently reliable.

The differences between piezometric heads that were calculated with the optimised parameter values (D_o) and the "true" dummy model parameter values (D_i) for the solution that corresponds with the best result for objective 1 are very small (Figure 3-5). The remaining differences are thought to be mainly a result of the optimisation approach, where solutions with *many* parameters that differ *little* from initially expected values are preferred above *few* parameters that differ *a lot* from expected values.

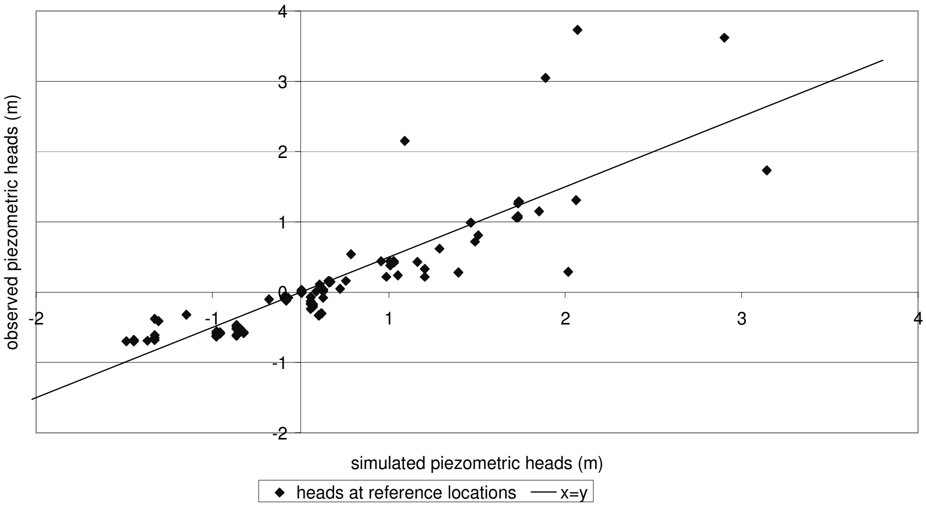

Figure 3-4 Differences between observed and simulated heads of the calibrated model

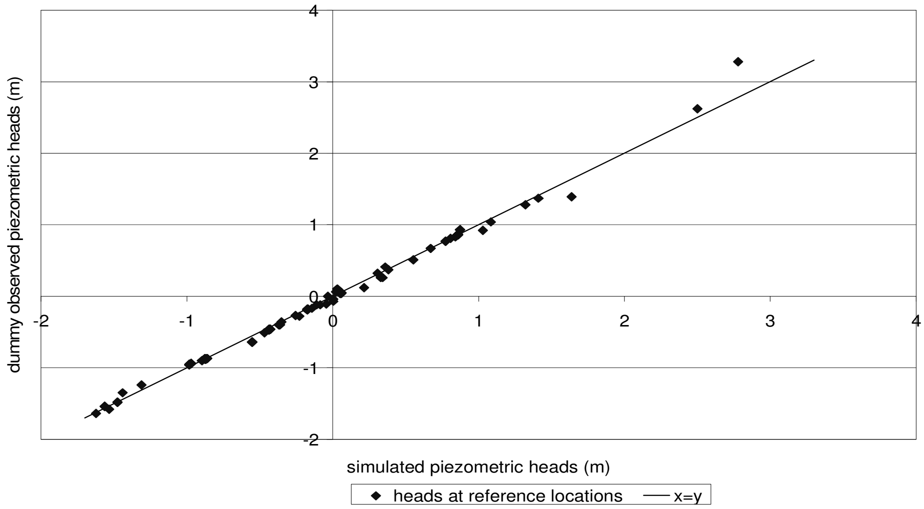

Figure 3-5 Differences between dummy-observed and simulated heads in the dummy model test

The differences between initially constructed dummy parameter values (D_i) and finally optimised values of the dummy model (D_o) are not an indicator of the reliability of the optimisation approach because the D_i dummy parameter values were generated *randomly* within the specified ranges, whereas the optimised values are aimed at minimum differences from initial estimates. Substantial differences between D_i and D_o can therefore be expected and are rather an indicator of the magnitude of the identification problem than of the performance of the optimisation approach. In Figure 3-6 is shown that there are considerable differences in parameter values although the dummy model fit (Figure 3-5) is very good. The differences between parameter values D_i and D_o in Figure 3-6 are calculated with the parameter values represented relative to the calibration ranges, according to equation 3-5:

$$g = \frac{V - V_{\min}}{V_{\max} - V_{\min}} \qquad (3\text{-}5)$$

Where:

g	normalised expression of V, relative to the corresponding calibration range $V_{max} - V_{min}$

A value of 0.5 of g corresponds therefore to the centre of the calibration range (initial estimate). The corresponding average of absolute differences between D_i and D_o in this dummy model test is 26%.

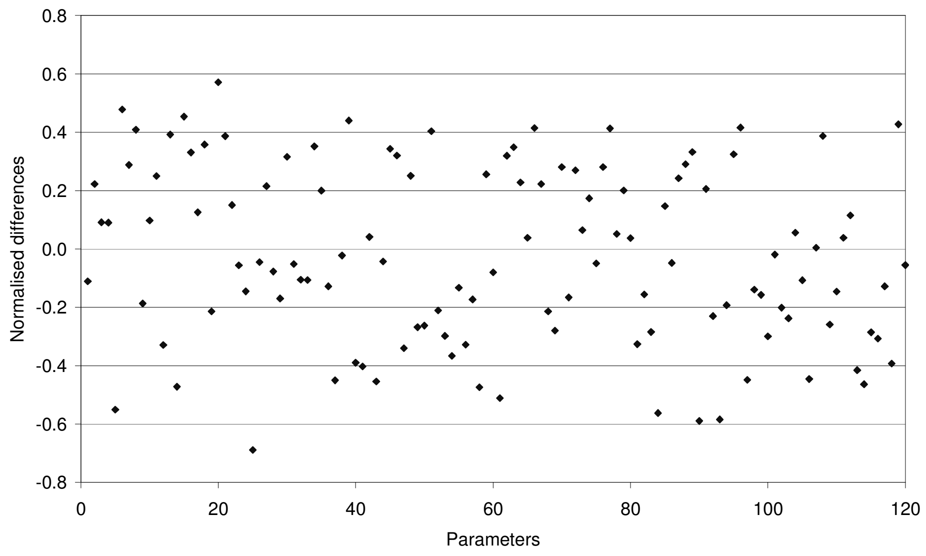

Figure 3-6 Normalised differences between D_i and D_o in the dummy model test

Special dummy model test

A special version of the dummy model test consists of the construction of a parameter set D_i where every parameter value corresponds to the initially expected value V_i. In that case the GA optimisation should eventually result in one single Pareto-efficient solution with (near) zero values for both objective functions (Figure 3-7). The results of the special dummy model test indicates that the GA is working properly. However, no information about the magnitude of scale-related and conceptual model errors and of the identification problem then becomes available.

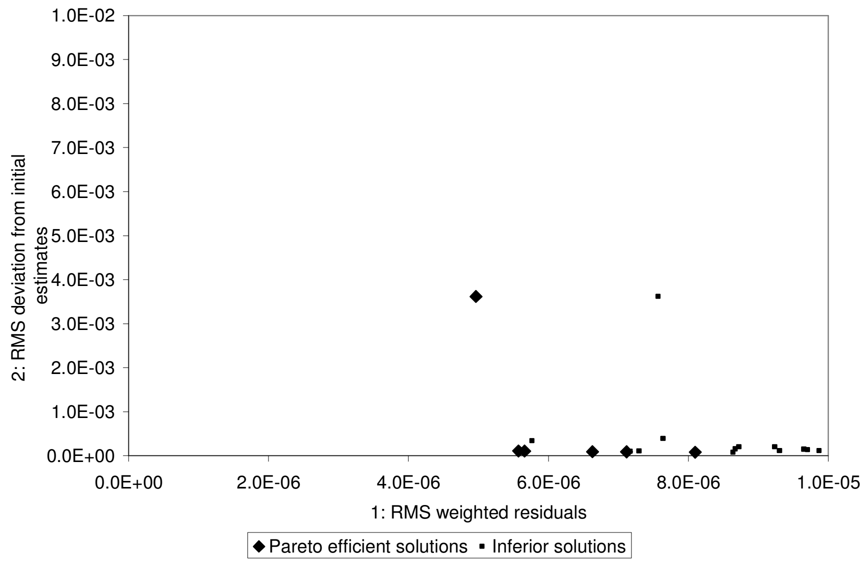

Figure 3-7 Solution fitnesses at the final stage of the special dummy model test

3.6 Conclusions and discussion

The GA demonstrated to be a stable and flexible technique for calibration of the model. 120 model parameters could be optimised in a single optimisation run, resulting in final solutions in which parameter values show a minimum deviation from the corresponding initial estimates, proportional to the estimated uncertainties as expressed in the size of the range between minimum and maximum parameter values.

Analysis of the shape of the Pareto front in relation to the corresponding parameter configurations enables a better understanding of the model and a better founded choice of a particular solution as calibrated model than with a single objective or lumped objective optimisation approach.

Circumstantial validation can be achieved by four different approaches, of which the 'dummy model test' can also produce information about the magnitude of the identification problem and of model errors that stem from scale-related, conceptual and observation errors, particularly if the dummy model test is carried out multiple times, with different D_i 's. The results suggest that the identification problem is of considerable magnitude, probably due to highly correlated model parameters. The magnitude of errors due tot the ensemble of conceptual model errors, scale-induced errors and errors in reference data is substantial and is the principal cause of differences in model fit as displayed in Figure 3-4 and Figure 3-5.

4. Multiple objective optimisation of drinking water production using a genetic algorithm[3]

4.1 Abstract

Finding a production configuration that allows economically efficient drinking water production at minimal environmental cost is often a complex task. A systematic trade off among the costs and benefits of possible solutions is required for determining the optimal production configuration. Such a trade-off involves the handling of interdependent and nonlinear relations for drawdown-related objective categories like damage to wetland vegetation, agricultural yield depression, reduction of river base-flow rates and soil subsidence. We developed a method for multiple objective optimisation of drinking water production by combining Busacker and Gowen's 'Minimum Cost Flow' procedure for optimal use of the transport network with a genetic algorithm (GA) for optimisation of other impacts. The performance of the GA was compared to analytically determined solutions of a series of hypothetical case studies. Pareto-optimality and uniqueness of solutions proved to be effective fitness criteria for identifying trade-off curves with the GA.

4.2 Introduction

Regional drinking water systems in densely populated areas typically consist of a network of transport pipes by which multiple sources and sinks are interconnected. The sources consist of production sites that pump either groundwater or surface water, the sinks are locations where drinking water is used. Generally these systems have spare capacity

---

[3] Adapted from Vink C. and P.P. Schot, 2002 Multiple-objective optimisation of drinking water production strategies using a genetic algorithm, Water Resources Research Vol. 38, NO 9.

available in order to respond to fluctuations of demand and also as an insurance to technical failure of system components. The presence of spare capacity implies the existence of a 'decision space' since there are different combinations of pumping rates by which the total required production within a specific planning period can be met. A need to formulate appropriate strategies for the allocation of production rates to the available production units therefore exists. Different strategies result in different economic and environmental efficiency. Both types of efficiency may vary considerably among strategies due to the sensitivity and spatial variability of the factors that determine impact categories. Optimisation of water management strategies is complex as some impact relations are nonlinear and interdependent. Interdependency and nonlinearity of impact relations often occurs if the impact of a category depends on a lowering of the groundwater level due to groundwater withdrawal. Damage to wetland vegetation, agricultural yield depression, reduction of river base-flow rates and drawdown-induced soil subsidence are examples of these categories. Nonlinearity exists in the relation between pumping rate and drawdown, but also in the way plants and soil react to desiccation. Interdependency of discharge-impact relations occurs when zones of influence of wells overlap. Then, groundwater drawdown at a particular site may be caused by more than just one well.

Before the eighties, defining optimal production configurations was relatively simple, as nonlinear and interdependent impact relations usually play a less important role within the set of objectives that were relevant to planning and management of drinking water production. Strategies were mainly determined by trade-offs between economic costs and benefits of which the major components could be generalized to linear functions of discharge rates [Stoner et al., 1979, 1981]. Then, optimal production configurations could be determined by applying linear programming techniques to economic functions of production and transport of drinking water [e.g. Schaffers, 1984]. From the eighties onward the linearization of impact relations became less appropriate as some environmental impact categories became more important. In the Netherlands, particularly the impact of groundwater withdrawal on the lowering of phreatic groundwater levels in wetland areas and the subsequent decline of groundwater-dependent vegetation (phreatophytes) gained importance [Grootjans, 1985, Braat et al. 1987, Runhaar 1996]. Effective evaluation of drinking water production strategies with respect to both economic and environmental categories requires integrated consideration of all impacts. Intensified exploitation of a remote well may result in a reduction of damage to wetlands, but at the same time it can invoke a steep increase in the use of the (fossil) energy that is needed for transportation of the water. Clearly, it should be avoided that a change of production configuration would indeed reduce the total damage to phreatophytes, but also would cause such a strong increase of other environmental burdens that the net environmental benefit is sub-optimal. Rational production configurations can only be determined if there is a comprehensive valuation of environmental categories. The weight of the relevant objective categories has to be determined by stakeholders and decision makers. Once the impact categories have been valuated, the optimal strategy can be determined. An integrated evaluation of drinking water production strategies based on quantitative valuation of all relevant categories has been rare, in spite of the clear need for it in a rational society.

In this chapter we present a method for the use of GA for multiple objective optimisation of multiple well drinking water production with interdependent and nonlinear impact relations. In the next section the structure of the model is described, followed by a discussion of the optimisation method in section 4. In section 5 the optimisation procedure is applied to a slightly simplified (not interdependent) and hypothetical problem, thus allowing comparison of the results with an alternative solution technique. Section 6 contains the discussion and conclusions.

4.3 Model setup

The optimisation method was implemented in a GIS-based decision support system in order to handle all spatial relations efficiently and to offer decision makers an adequate access to the methodology. The support system represents a model for regional drinking water supply in which the various impacts of strategies are quantified by category (Figure 4-1).

The optimisation problem can be defined according to:

$$\textit{Minimise}\ (\ IC(i=1)..IC(i=m)) \tag{4-1}$$

$$IC(i) = f(i, PC) \tag{4-2}$$

$$PC = Q(j=1)..Q(j=n) \tag{4-3}$$

With these constraints:

$$\Sigma Q(j) = D_t \quad (4\text{-}4)$$

$$Q(j) <= Q_{max}(j) \quad (4\text{-}5)$$

$$IC(i) \text{ subject to } CT(k) \quad (4\text{-}6)$$

Where:

IC(i)	impact of category i (expressed as cost, burden or damage)
m	number of impact categories
n	number of wells
f(i,PC)	impact function of category i
PC	production configuration (a spatial distribution of discharge rates over the available wells)
Q(j)	discharge rate of well j ($L^3.T^{-1}$)
D_t	total demand ($L^3.T^{-1}$)
$Q_{max(j)}$	maximal allowable discharge rate of well *j* ($L^3.T^{-1}$)
CT(k)	constraint *k* (user-defined constraints like drawdown in area X may not exceed Y cm)

The impact categories that we defined are:

- economic costs;
- pumping cost;
- purification cost;
- transport cost;
- agricultural yield reduction due to groundwater drawdown;
- ecological damage to wetland vegetation due to groundwater drawdown;
- energy consumption;
- use of strategic groundwater reserves.

The impacts are quantified by economic and coupled geohydrologic and environmental modelling. The models vary in complexity from simple linear relations such as those for pumping costs, to complex nonlinear impact models, such as the model for ecological damage to wetlands by groundwater drawdown. Production configurations are optimised by applying a genetic algorithm. Efficient use of the transport network does not follow directly from an optimal distribution of pumping rates over the available wells. Optimal use of a directed flow network of limited capacity and pipe-specific transport costs has been studied by many authors [see e.g. Papadimitriou and Steiglitz, 1985]. We solved this secondary optimisation problem by applying a 'min-cost-flow' type algorithm that was based on the work of Busacker and Gowen [1961].

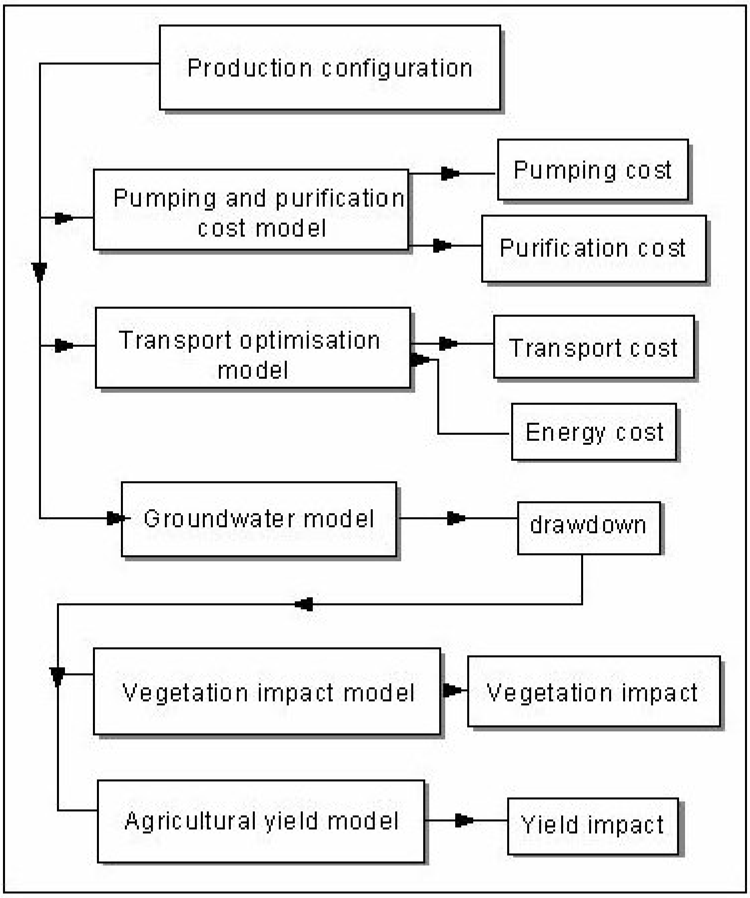

Figure 4-1 Modular structure of the decision support system

4.4 Optimisation methods

Until recently, the nature of the impact relations that are involved in regional drinking water production rendered it in practice unfeasible to optimise both economic and environmental objectives according to the approaches that were sketched in the previous section. Many optimisation techniques are inappropriate because of nonlinearity and interdependency in some impact relations.

In theory, the number of possible strategies and corresponding production configurations is infinite due to the continuous nature of flow. However, if differences in production configurations are sufficiently small, they will not result in significant differences of impacts. Discretization of discharge rates into a suitable step size may therefore be introduced in order to reduce the extent of the search space. Then the problem becomes combinatorial. The theoretical number of possible production configurations that represent feasible combinations of well discharges is defined by a function in which the spatial distribution of the demand, the number of wells, the discretization of discharge rates (step size), the spare capacity of the wells and the capacities of the transport pipes form the principal variables. The total number of possible configurations depends therefore on case-specific constraints and varies widely, due to the exponential explosiveness of the combinatorial problem. For a system of N wells of equal capacity that are interconnected by a transport system of unlimited capacity, the total number of combinations is defined by

$$R = S^N \qquad (4\text{-}7)$$

Where:

R	total number of combinations;
N	number of wells;
S	number of discharge rate steps per well.

The number of feasible combinations is restricted by the constraint:

$$\Sigma Q(i) = D_t \qquad (4\text{-}8)$$

Where:

Q(i) discharge rate of well i

The number of feasible combinations depends on the spare capacity of the regional system. There are already one billion feasible combinations for a hypothetical system of 15 wells with 10 discharge rate steps per well, a spare capacity of 30% and a nonlimiting flow network. As it may take typically 1 minute to calculate the various impacts of a single production configuration, it would require almost 2 million years calculating all feasible combinations. Generally, the number of feasible solutions is by far too large to handle with computer-based 'brute force' techniques such as an exhaustive evaluation of alternatives or a Monte Carlo approach. Often this is even so when capacity constraints from the transport system strongly reduce the number of feasible solutions.

Apart from analytical, exhaustive or Monte Carlo approaches, heuristic approaches are sometimes applied to combinatorial explosive problems with multiple criteria. Purao et al. [1999] discuss how the search space can be explored by generating efficient solutions within local regions by means of heuristic approaches. 'Exchange search' is one of their principal heuristic techniques. It consists of the exchange of one or more of the values of variables in a non-dominated realization by minimal or maximal extremes, as perceived by decision makers. Thus, the boundaries of the local search space can be identified. The nonlinearity and interdependencies that exist in the vegetation impact functions limit the feasibility of this approach for multiple objective optimisation of drinking water supply. As the approach requires valuation a priori, decision makers need to make on beforehand statements about satisfycing levels, or about which relaxations they are prepared to make for the various objective functions. This need for prior information about the preference structure limits the practical use of this approach [White, 1985; Purao et al., 1999]. Rajabi et al. [1999] were successful in handling interdependencies by means of their modified goal programming technique but this method also requires decision makers to identify reasonable goals on each criterion before final solutions can be presented. Decision makers thus miss the great advantage of having access to Pareto fronts and the possible analysis of the 'exchange rates' between objectives before choosing. For the type of application that is discussed in this paper, Pareto fronts can be used to show how economic cost and environmental damage are interrelated for optimal solutions. If a particular budget X is available for alleviation of environmental damage, then the solution that corresponds to optimal allocation of this budget can be identified at the Pareto front (Figure 4-2). The slope of the curve indicates the marginal reduction of environmental damage as a function of economic investments.

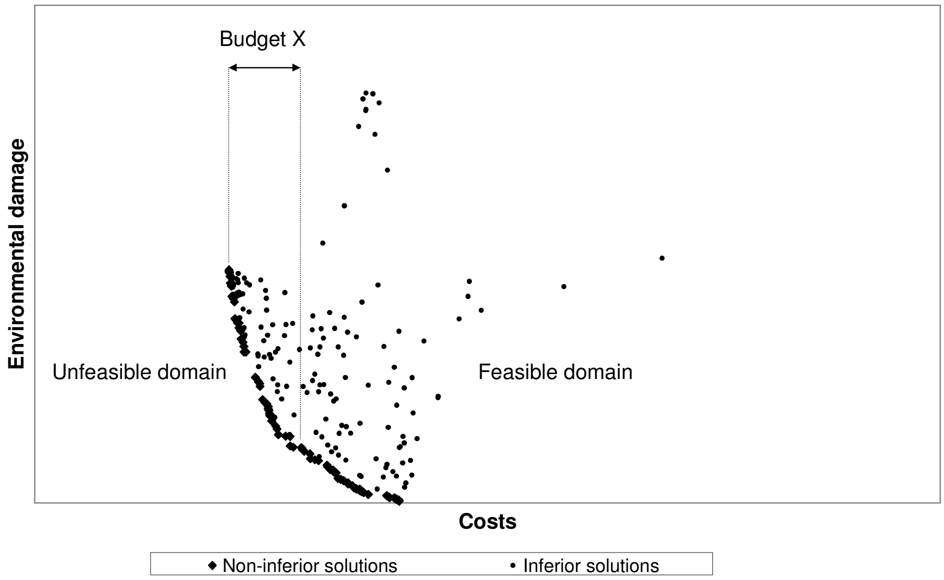

Figure 4-2 Pareto front of two conflicting impact categories

4.5 Genetic algorithm

To overcome the limitations of the approaches that were described here above, we applied a genetic algorithm. This technique is suitable for single objective optimisation, but it is also convenient for multiple objective optimisation by means of Pareto fronts, as many non-dominated solutions can be identified in a single optimisation run.

Over the last few years evolutionary algorithms have been applied successfully to multi-objective optimisations problems [e.g. Cieniawski et al. 1995; Deb & Kalyanmony, 1999]. The development of genetic algorithms was inspired by the genetic processes of biological organisms. The concept of natural selection by survival of the fittest as stated by

Charles Darwin in *The Origin of Species* plays a major role. Application of the principles of selection and mutation in computer programs was first proposed by Holland [1975]. Application of GA's for multi-objective optimisation was outlined in Goldberg [1989]. At present, evolution programs have been applied successfully to a wide range of problems [e.g. Grefenstette 1990]. GA's work with a 'population' of possible solutions to a problem. The 'fitness' of each member of the population is calculated and the properties (genes) of those who perform best are mixed with other solutions, leading to new members of the population. The 'cross-over' process has been implemented in many different varieties according to the specific character of particular optimisation problems. The reproduction cycle is repeated until there is convergence, in the sense that no further improvement of solutions occurs (Figure 4-3).

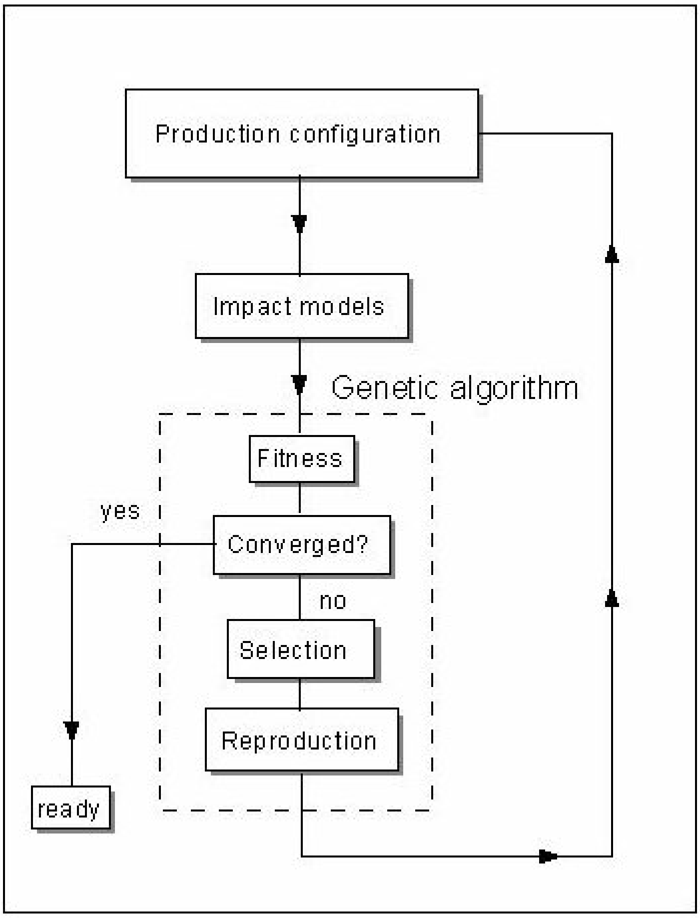

Figure 4-3 Flowchart of optimisation approach

4.5.1 Coding & Genes

Each member of the population is allocated a chromosome that represents the distribution of discharge rates over the available wells. Every gene of the chromosome represents a discharge rate of an available well. The total number of genes is therefore equal to the total number of wells and the ensemble represents a production configuration. The sum of all discharge rates is equal to a user defined demand. Discharge rates of individual wells are being modified during the optimisation, but the sum of all discharges should remain equal to the total demand (equation 4). The number of configurations that comply with this constraint is only a small part of all possible combinations. Therefore we increased the probability of valid reproduction by representing the discharge rate of a well indirectly according to:

$$Q(j) = g(j) \cdot Q_{\max(j)} \cdot \frac{D_t}{\sum [g(j) \cdot Q_{\max(j)}]} \tag{4-9}$$

Where:

$g(j)$	value of gene j (0 <= g <=1) (-)
$Q(j)$	discharge rate of well j ($L^3.T^{-1}$)
Q_{max} (j)	capacity of well j (maximal allowable discharge) ($L^3.T^{-1}$)
D_t	total demand ($L^3.T^{-1}$)

After this transformation it is still possible that unfeasible configurations will be generated, but only by violation of equation 5, the well capacity constraint. The probability of this event is much smaller than violation of equation 4 without transformation. The frequent occurrence of non-valid production configurations thus no longer hampers the optimisation. If a non-valid chromosome has been generated randomly chosen genes are modified with random values between 0 and 1 until a valid configuration is formed.

4.5.2 Reproduction and mutation

Reproduction operators determine the genetic information of new solutions. Offspring are generated by combining genetic information of the 'fittest' 5 solutions with randomly chosen solutions from the population. Two offspring are generated per couple. Each reproduction cycle the newly constructed offspring replace the 10 least fit members of the population. We achieved best results with arithmetical and uniform crossover techniques. The size of the population of possible solutions remains constant as solutions with the lowest fitness are being replaced by new offspring. The reproduction cycle is repeated until there is no further improvement of fitness. The typical size of the populations varies between 50 – 500 solutions. For the current hypothetical case, the optimisation performed best at a mutation probability of p=0.05 per gene.

4.5.3 Fitness

The fitness score of a solution corresponds to the weighted sum of its relevant impacts and expresses the degree in which a solution is desirable. If all impacts can be expressed in a common scale by weighted conversion, the problem is reduced to a single objective optimisation. Otherwise, if there are impact categories of which the 'weight' for conversion into a common scale cannot be agreed on a priori, Pareto fronts of non-dominated solutions have to be determined. In the first case, fitness in the GA corresponds to the lumped impact score. In the second case there are multiple impact classes that should be attributed a fitness score in an unbiased manner for the selection of solutions for 'reproduction'. Then it is undesirable that the optimisation converges to one single solution. Several authors reported difficulties in formulating unbiased and yet effective selection criteria that maintain diversity in the collection of solutions [Beasley et al., 1993]. We achieved good results by selecting reproductive solutions based on two properties:

- Pareto-optimality;
- exclusiveness.

Pareto-optimality is a Boolean property that is true for solution A if there is no other solution that has a higher score for any fitness category, while scores at other fitness categories are not inferior to those of solution A (see 2.3). Pareto-optimality is a relative property; during the calculation the Pareto front gradually moves towards the final shape that divides between feasible and unfeasible solutions at minimal cost. Solutions that are initially non-dominated therefore can become inferior by the creation of other, more efficient solutions. 'Exclusiveness' of solutions is defined as a fitness criterion in order to maintain variation in the population of solutions and thus to prevent premature convergence at one single location of the Pareto-front. The exclusiveness of a solution is expressed by the relative 'distance' of the set of impacts of the solution to other solutions.

‘Distance’ is calculated analogous to geometric distance in a n-dimensional space, where the coordinates are expressed in normalized (relative) units of the various impact categories. Total fitness is calculated by ranking the scores that are calculated according to the two aforementioned components of fitness in such a manner that Pareto-optimal solutions are never lower in rank than dominated solutions, with ‘exclusiveness’ as a secondary sorting criterion. A concise description of the principal notions in genetic algorithms and the specific meaning of these notions for the current application is presented in Table 4-1.

Table 4-1 Concise description of the principal notions in genetic algorithms

Concept	**General description**	**Specific description for the current case**
Population	A set of solutions; the total number of solutions is constant. Solutions that are less desirable than others are being replaced by more competitive ones during the optimisation process (Survival of the fittest) .	Particular distributions of discharge rates over the available wells.
Chromosome	Information that defines the properties of a solution. Each property is represented in a gene.	The number of genes corresponds to the number of wells.
Gene	The value or condition of a particular property of a solution.	Each gene is a variable that refers to a particular well. It contains a scalar value that represents the pumping rate of that well.
Crossover	Sets of rules that define how	50% of the offspring is

Concept	General description	Specific description for the current case
techniques	new solutions (offspring) are constructed by application of the rules to existent solutions.	constructed by arithmetic crossover, 50% is constructed by uniform crossover.
Reproduction	The construction of the genetic information of new solutions (individuals) by using information of existent solutions.	Every reproduction cycle 10 new solutions are constructed from 5 existent solutions.
Offspring	New solutions, constructed by mixing the properties of existent solutions by means of crossover techniques.	Every reproduction cycle 10 least-fit solutions are replaced by new offspring.
Mutation	A stochastic process that may result in modification of the content of genes that are being constructed.	The value of 5% of all newly constructed genes is determined by mutation as a random number $(0 <= g <= 1)$.
Fitness	The degree in which a solution is desirable.	The rank of a solution in the population with respect to pareto optimality and the degree of uniqueness of its impacts.

4.6 Procedure validation

A hypothetical case study was constructed in order to investigate the feasibility of our approach. The case study consists of imaginary supply systems of 4 - 48 interconnected wells. For each well we defined two highly nonlinear impact functions. The impact functions for the wells were constructed so as to imitate ecological impact and lumped economic costs, both as a function of discharge rate. The impact functions were constructed by generating (pseudo) random numbers for each well, according to the following algorithm (in pseudo programming language):

```
Algorithm Generate_impact_functions

For each impact category i
 begin
   For each well w
     begin
         Generate a random factor f
         For each discharge rate step q
           begin
             Generate a random number n (0-10)
             impact(i,w,q) = impact(i,w,q-1) +  n*f
          end
      end
  end
```

The impact functions represent a similar nonlinearity and capriciousness in the relation between discharge rate and impacts as may occur in the real world. We simplified the problem by leaving interdependency of a well's impact functions out of the hypothetical case with the purpose to make it possible to validate the results as generated by the GA technique, by determining the optimal solutions also by means of a discrete 'buildup' approach. The buildup approach consists of stepwise allocation of capacity to wells, using minimal average costs as criterion. The required total discharge rate was defined as 70% of the total capacity. The performance of the GA and Monte Carlo approaches was assessed by comparing the results with the optimal solutions that were identified by the 'buildup' procedure. In Figure 4-4, the performance of GA and Monte Carlo is displayed for single-objective optimisation. All GA versions performed significantly better than the Monte Carlo approach. The number of possible solutions in the hypothetical case of 48 wells, interconnected by a flow network of infinite capacity, is clearly so large that state-of-the-art personal computers lack the 'brute calculation force' for effective optimisation by means of (pseudo) random generation of solutions. The best solution found by Monte Carlo simulations differed 140% with the analytical solution after an equal calculation time. However, the difference of the best solution found by the GA and the analytically determined optimum was only 3% after 5 minutes calculation time on a 200 MHz Pentium computer.

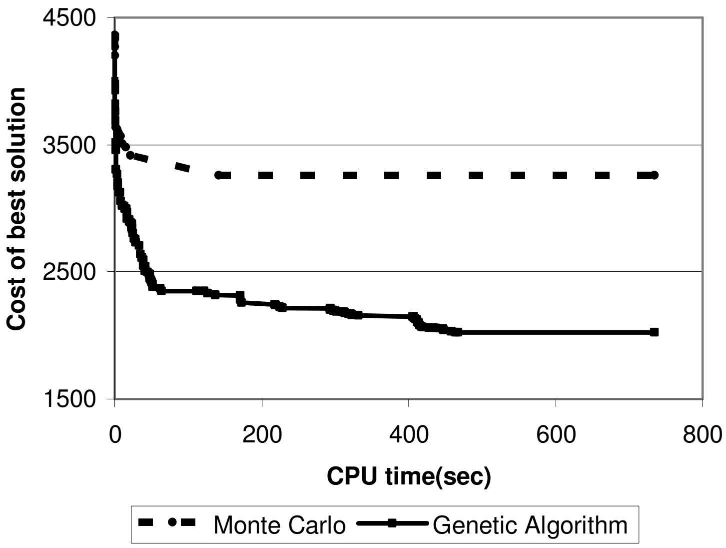

Figure 4-4 Performance of GA and Monte Carlo techniques for a single-objective optimisation of a hypothetical drinking water system with 48 wells.

For multiple objective optimisation, we assumed two impact categories that are not translatable into a common scale. Figure 4-5 shows the performance of the GA approach as compared to Monte Carlo for multiple objective optimisation. The population size was set to a stationary size of 220 solutions. Smaller sizes of the population resulted in Pareto fronts of which sections were insufficiently explored.

Some differences in performance among crossover techniques were observed. However, adding calculation time could compensate for the difference in efficiency of the various crossover techniques. We achieved the fastest approximation of the analytically determined optima with a mixed 'arithmetical' and 'uniform' crossover technique. Both crossover techniques were attributed equal probability for controlling a reproduction. The Pareto fronts that were found by means of GA required 100 - 100,000 generations in order to reach convergence. The number of required generations increased with the number of available wells and hence the size of the search space. Pareto-optimal solutions could be identified efficiently. Results at the ends of the Pareto front differed less than 5% with the corresponding analytical solutions when the maximum number of generations was set to 100,000. In real world situations the search space of comparable systems is often smaller than of the hypothetical cases we investigated, due to constraints from especially the transport network. It is therefore expected that optimisation by means of GA also will be sufficiently efficient for large systems of more than 48 wells.

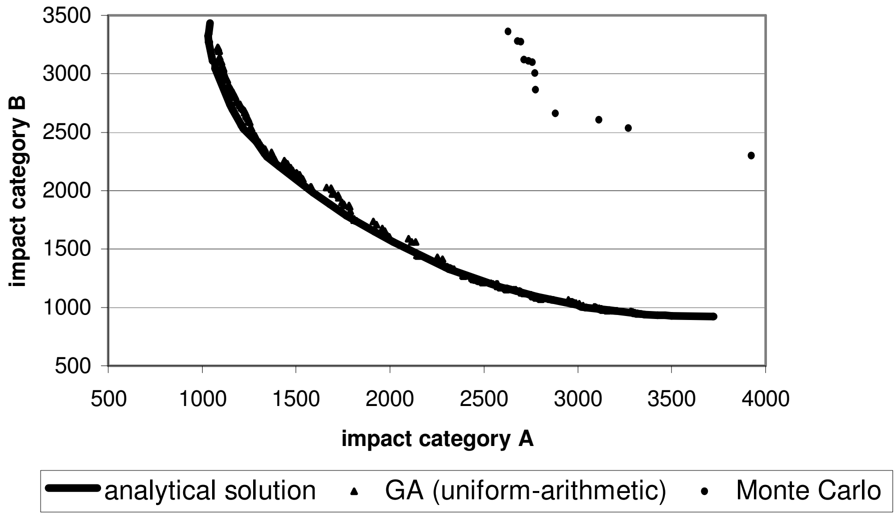

Figure 4-5 GA and Monte Carlo results for two-objective optimisation of a hypothetical drinking water system with 48 wells.

The lack of proof of whether the global optimum or 'true' Pareto front is found is a serious limitation for the applicability of GA in optimisation problems. Different values for parameters like population size and mutation probability may strongly influence the performance of GA's. The evolutionary approach offers in itself no verification for the possibility that essential parts of the search space have not been explored effectively. GA should therefore only be used in cases where at least 'circumstantial' validation is possible. Application of the method to realistic cases showed that the type of optimisation problem that is discussed in this paper generally can be partially validated by inspection of the wells discharge-impact functions. The most efficient wells at maximum capacity with respect to a particular objective category often can be identified by comparing impact scores among wells. Criteria for the partial validation of the results can be based upon the analytically derived conclusion that these wells should be engaged at maximal capacity in the optimal solution that corresponds best to that particular objective category.

4.7 Application example

In the previous section the performance of the GA optimisation method was assessed by applying it to an abstract, two-objective, nonlinear and non-interdependent mathematical optimisation problem. In this section we describe the application of the method to a more practical optimisation problem with more than two objectives. The problem concerns an example based upon an existing area in which many realistic features of regional drinking water are represented. The study illustrates how the GA method can be applied in practice to regional drinking water production issues.

It is assumed that representatives of a drinking water company and provincial authorities jointly have to decide on how discharge rates are to be distributed over the available wells in a particular region. They agree on the models by which the various impact categories are quantified and on the total required production rate. The problem to be solved is the distribution of discharge rates over the available wells in such a manner that the total of weighted adverse impacts and costs is minimal.

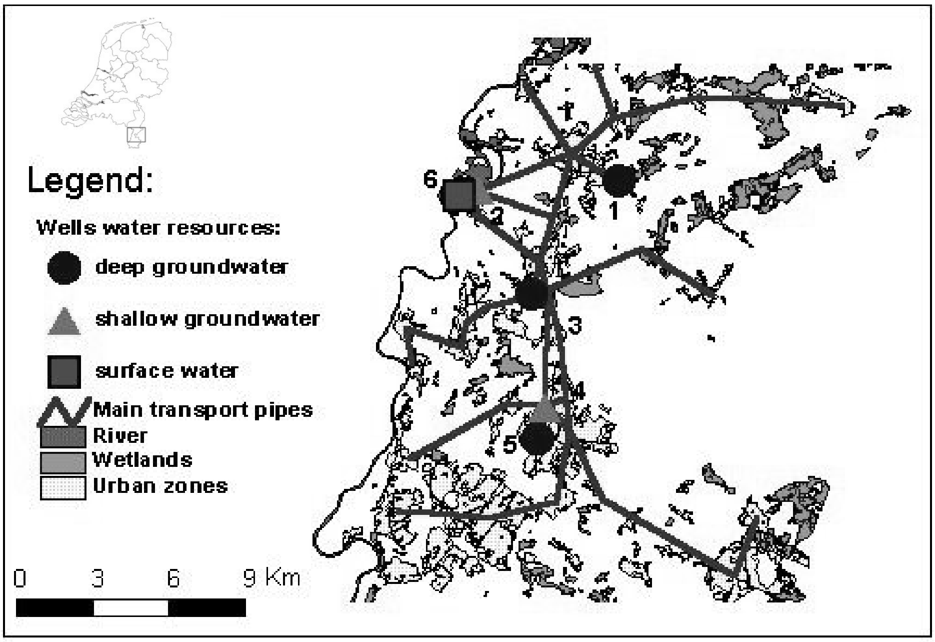

Figure 4-6 Case area

The case area is located in the South of The Netherlands (Figure 4-6) where Drinking Water Company WML is in charge of drinking water supply. The essential characteristics of the area are a plausible representation of drinking water supply systems in The Netherlands. Some data are fictive as not all required information was readily available. The locations of urban zones and areas with (semi) natural vegetations were derived from a digital land use map (courtesy Alterra, Wageningen). The required data on annual demand rates of drinking water was derived from population data within the case area. The unit costs of production, purification and transport are fictitious.

The system consists of 10 production wells, interconnected by a transport network. Some of the wells pump deep groundwater that recharged several hundreds of years ago. Other wells pump water out of relatively shallow aquifers, of which the groundwater is not older than 50 years.

The suitability of the deep old groundwater for drinking water production is excellent, but extensive use would result in depletion of strategic groundwater reserves. The shallow groundwater is of relatively poor quality due to agricultural production and consequently requires more extensive purification than the deep groundwater. One of the wells is located near the shores of the River Meuse and pumps very recently infiltrated surface water. This water requires extensive purification but the pumping invokes very little drawdown and therefore no damage to wetland vegetation is caused. At several locations in the region there are wetlands with valuable groundwater-dependent vegetation. The damage to the wetlands due to groundwater withdrawal varies among the available wells.

4.8 Optimisation objectives

Two groups of objectives were defined. The first group concerns a selection of objectives that are perceived relevant in present day drinking water supply in the Netherlands:

1. minimal total economic costs;
2. minimal damage to groundwater-dependent vegetation.
3. minimal use of deep groundwater

The second group is defined in order to enable additional validation possibilities of the results of the optimisation by analytical inspection:

4. minimal purification costs
5. minimal transport costs

Properties of solutions that comply maximally with these objectives can be formulated on beforehand by inspecting the relevant discharge rate –impact relations:

The production configuration according to the objective of minimal purification costs should operate the wells that pump deep groundwater at maximal capacity because deep groundwater requires only minor purification. The deep wells are therefore very cost- effective with respect to purification costs, as compared to all other sources.

The production configuration with minimal transport cost should operate wells near urban zones at maximum capacity if the demand of that urban centre exceeds its capacity.

The production configuration with minimal use of groundwater should operate the wells that pump surface water at maximum capacity.

If the best solutions for the aforementioned 'objectives for validation' comply with the properties that were defined on beforehand by analytical inspection of the discharge - rate impact relations, it can reasonably be assumed that all solutions that were identified by the GA are (near) optimal. Thus we can obtain strong circumstantial evidence that the results are valid. All solutions therefore have been determined during a single optimisation run in order to maximize the reliability of the validation method. This implies that we carried out the optimisation for 5 objectives simultaneously.

4.8.1 Impact models

The impact models should provide a quantitative criterion for the objectives that were defined previously. There are three essential impact models. The first is the economic cost model, in which pumping cost and purification cost are well-specific and a linear function of discharge rate. Minimal transport costs were determined by applying Busacker and Gowen's MinCostFlow algorithm [1961] to a network with pipe-specific capacities and cost functions. The second is the hydrologic impact model, in which spatial distribution of drawdown is well-specific and a function of discharge rate and aquifer properties. In this study we applied Theis's equation [e.g. Todd, 1980]. The third model is the vegetation impact model, in which the adverse impact of drawdown to vegetation was assessed by a conceptual nonlinear impact model, using a distributed approach of drawdown, and fictitious, location-specific data on vulnerability to drawdown and value of vegetation.

4.9 Validation

The human mind can easily identify optimal solutions that correspond to extreme valuations of a single objective category, but it is not very well equipped to find optimal solutions that correspond to more shaded value schemes. The GA optimisation method is an effective technique to find the best solutions in real-world optimisation problems with conflicting objectives, as compromises are then typically the most satisfying solutions. The solutions with minimum ecological damage, as identified by GA and analytical inspection respectively, consist of the production configuration in Table 4-2. Discharge – impact relations for vegetation damage are displayed in Figure 4-7 and Figure 4-8.

Table 4-2 Optimal production configurations according to GA and analytical inspection for minimal vegetation damage

Wells	Discharge rate GA (m3/a)	Discharge rate analytical (m3/a)
1	0	0
2	4154570	5990116
3	0	0
4	0	0
5	0	0
6	5935101	5307242
7	2546000	2704842
8	980989	0
9	0	0
10	385540	0

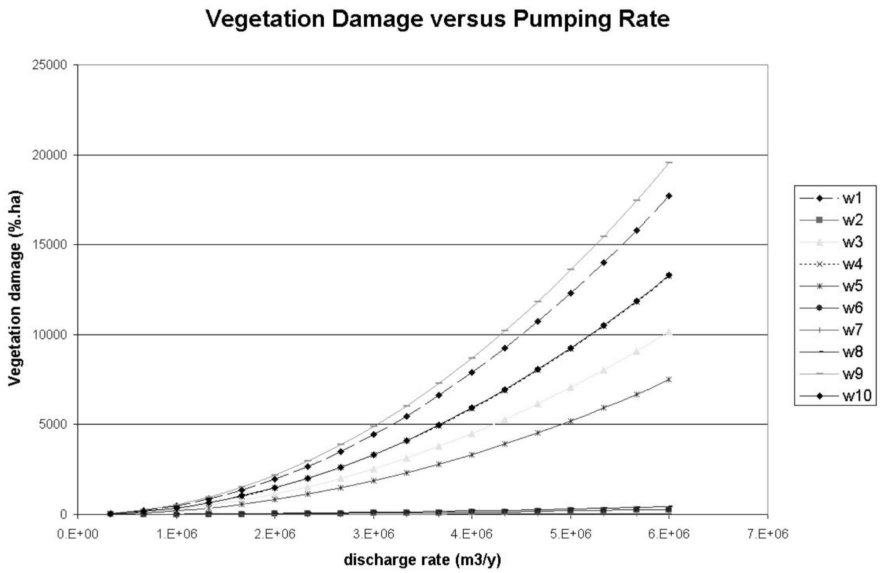

Figure 4-7 Discharge rate - Vegetation damage functions

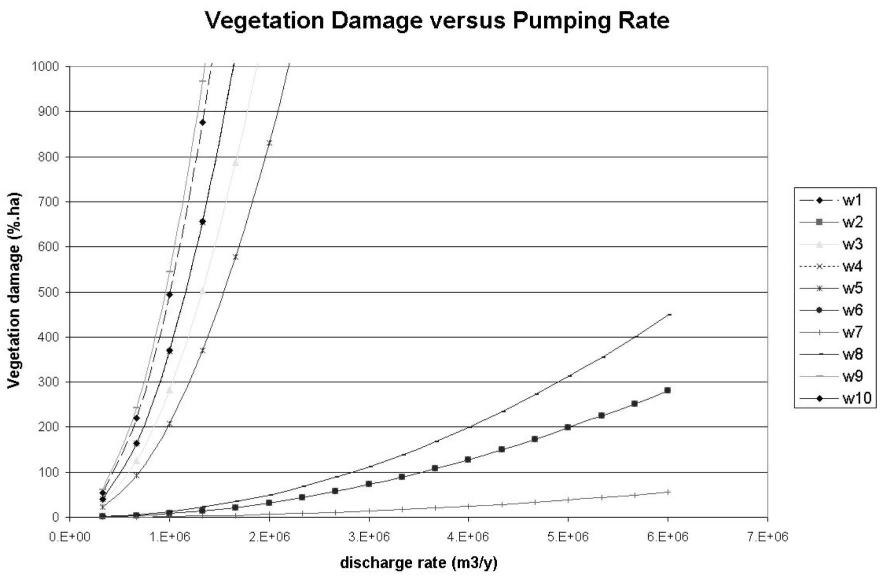

Figure 4-8 Discharge rate - Vegetation damage functions (detail)

The optimal solutions according to GA as compared to analytical inspection show significant differences in terms of discharge rate with respect to all wells, but the difference between the total ecological impact of both solutions is less than 0.05 % of the feasible range of ecological impacts and therefore insignificant. The wells 8 and 10 invoke similar vegetation damage as wells 2, 6 and 7 at low discharge rates (Figure 4-8).

The solutions with minimum use of deep groundwater as identified by GA and analytical inspection respectively consist of a production configuration as specified in Table 4-3:

Table 4-3 Optimal production configurations according to GA and analytical inspection for minimal use of deep groundwater

Wells	Use of deep groundwater	Discharge rate GA	Discharge rate analytical
1	y	0	0
2	n	5690500	any
3	y	0	0
4	n	4732100	any
5	y	0	0
6	n	2515800	any
7	n	162287.4	any
8	n	400238.4	any
9	n	0	any
10	n	0	any

Both solutions show zero discharge rates for the wells that pump deep groundwater and the results as found by means of the GA are consequently validated.

The solutions with minimum purification cost as identified by the GA and analytical inspection respectively consist of a production configuration as stated in Table 4-4.

Table 4-4 Optimal production configurations according to GA and analytical inspection for minimal purification cost

Wells	**Minimal purification cost**	**discharge rate GA**	**discharge rate analytical**
1	y	4222886	any
2	n	0	0
3	y	4111757	any
4	n	0	0
5	y	5667557	any
6	n	0	0
7	n	0	0
8	n	0	0
9	n	0	0
10	n	0	0

Both solutions only show non-zero discharge rates for the wells that do pump deep groundwater as these are lowest in purification requirements. The results as found by means of the GA are consequently “circumstantially” validated.

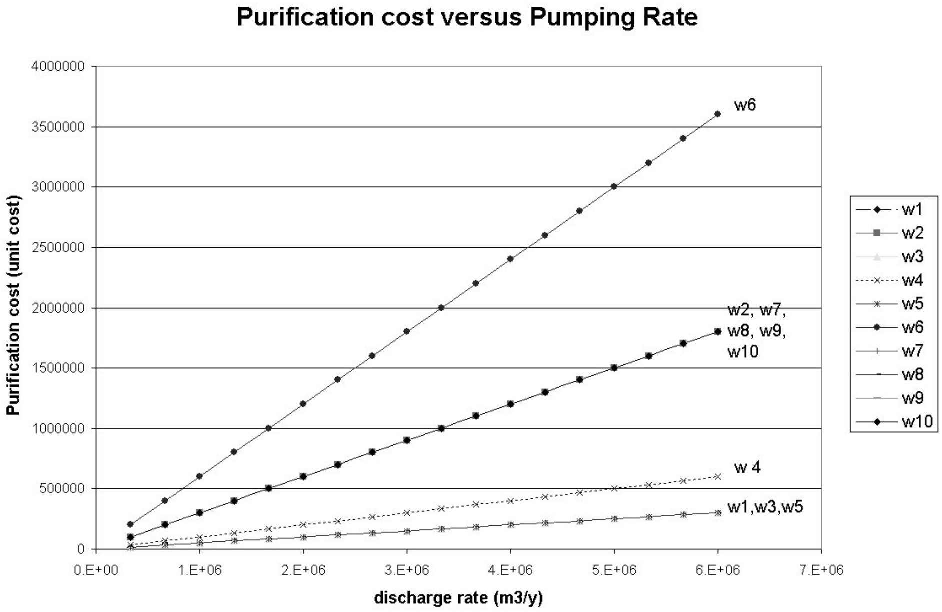

Figure 4-9 Discharge – Purification cost functions

4.10 Pareto-efficient solutions

The Pareto surface of feasible solutions with respect to impact categories 'total cost' and 'vegetation damage' is shown in Figure 4-10.

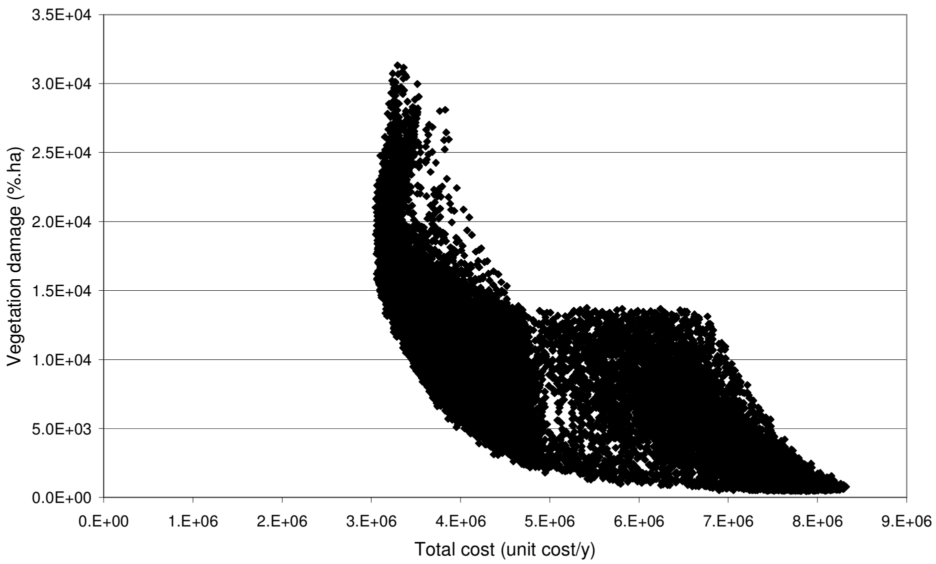

Figure 4-10 Total cost and vegetation damage of Pareto-efficient solutions

The typical asymptotic shape in the Pareto surface indicates that total production costs and damage to vegetation are conflicting objectives in the current case. Pareto-efficient solutions are not located exclusively along a spatial boundary of the solution space as the optimisation was carried out for more than two objectives. The wells that cause less damage to vegetation are to some extent located at a large distance from the urban zones where the water is needed. Pumping large discharges at these sites results therefore in higher transportation costs. Wells that located near the shores of the River Meuse cause negligible damage to vegetation as virtually no drawdown is invoked by pumping, but operational costs of these wells are high due to the purification requirements of the pumped water.

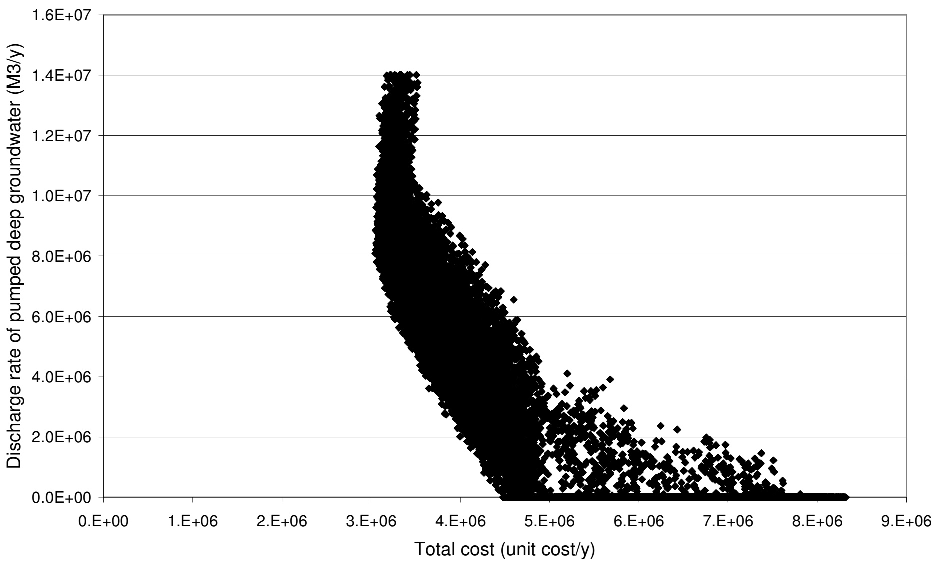

Figure 4-11 Total cost and use of deep groundwater of Pareto-efficient solutions

The objectives minimal total cost and minimal use of deep groundwater are to some extent conflicting as the deep groundwater is cheap because it requires no significant purification. However, solutions with all production allocated to deep wells do not show minimal costs due to the costs of transportation to remote sections of the case area.

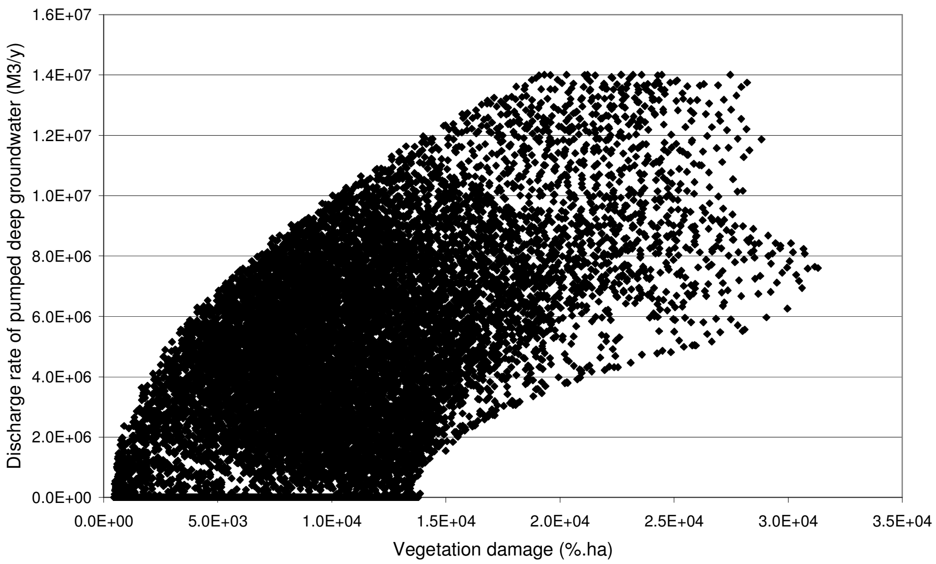

Figure 4-12 Vegetation damage and use of deep groundwater of Pareto-efficient solutions

The objectives minimal vegetation damage and minimal use of deep groundwater are positively correlated because groundwater withdrawal at the deep wells affects the vegetation in wetlands. There are solutions with both minimal vegetation damage and minimal use of deep groundwater. Solutions that use surface wells at full capacity satisfy both objectives. However, the economic costs of these solutions are extremely high.

Decision makers have access to detailed information on promising pareto-efficient solutions. The graphical presentation of Pareto fronts and Pareto surfaces can help to understand the characteristics of the decision space.

4.11 Calculation time requirements

Using GA as an optimisation technique for regional drinking water supply requires calculation times that are for many problems still considerable. It may typically require 1-5 minutes on a 200 MHz PC for a complete impact calculation of a production configuration. Thousands of these calculations may be necessary for the determination of a Pareto front. However, there are good possibilities for use of parallel computers with GA, if computation times become too long. In order to reduce the required calculation time we defined three levels of detail for impact calculations. The 'noisiness' of the impact functions was gradually reduced during the optimisation process. Initially, impact functions were expressed as independent functions of wells discharge rates. According to the principle of superposition, the drawdown at a location caused by the discharge of several wells is equal to the sum of the drawdowns caused by each well individually. Although the principal does not hold in conditions that include the presence of phreatic aquifers, rivers and ditches, it is assumed during the first stage of the impact calculations that drawdown and drawdown-induced impacts of different wells are independent and can be superposed. In reality this is not the case as both the relation between pumping rate and drawdown as well as the relation between drawdown and ecological damage often possesses interdependent, nonlinear and therefore non-superposable characteristics. Once the improvement of the fitness of generated solutions stagnates and near-optimal solutions for the initial fitness function have been found, the objective function is replaced by a less noisy alternative. The superposition principle in the relation between drawdown and damage to agricultural yield or vegetation is then no longer assumed, but for calculation of total drawdown the superposition principle is still assumed to be correct. At the third stage, drawdown at locations that are within the area of influence of more than one well is no longer assumed to be superposable. For each solution a digital drawdown map is generated by a numerical groundwater model. The drawdown map is input to the ecological and agricultural impact models. By applying different levels of precision in the objective function we achieved a significant reduction of the required calculation time to 5 – 100 hours, depending on the number of available wells. Response times that allow immediate inter-activity remain therefore out of reach. We could achieve an acceptable level of flexibility of the decision support system to decision makers by storing the key properties of many non-dominated solutions on disk. Detailed impact calculations and generation of digital maps for a particular solution can be carried out on demand and require typically less than 5 minutes.

4.12 Discussion & conclusions

According to the results of the hypothetical case study we conclude that the GA can be an effective tool for multi-objective optimisation of regional drinking water supply. We achieved best results by applying arithmetical and uniform crossover techniques jointly. Pareto-optimality and uniqueness of solutions proved to be effective fitness criteria for identifying Pareto fronts. These fitness criteria are unbiased and suitable for any number of objective categories. Validation of results that were generated by GA is needed, since not all values for GA parameters like population size and mutation probability yielded adequate results. We constructed a hypothetical case study that does not contain interdependent discharge-impact relations in order to enable validation of the results by another method. The optimisation by means of GA does not require the formulation of an aggregated impact function that is typical for analytical optimisation methods and consequently there seems no ground to suppose that interdependency among discharge-impact relations could impede its effectiveness. "Circumstantial validation" is often feasible in multi-objective optimisation problems because solutions that correspond to maximal or minimal impacts with respect to a single objective category are relatively easy to identify. Circumstantial validation forms a check of the results but it offers no proof.

The proposed optimisation methodology enables a rational and consistent evaluation of production configurations. By using the method implemented in a spatial decision support system "decision makers may become active participants in a regional planning analysis, rather than selectors among a few, preplanned alternatives" [Jones, 1998]. A truly optimal production strategy cannot be defined without taking in account all other potential environmental consequences of decisions made at policy, planning and programmatic levels. In that sense it requires a policy framework where there is a sufficient level of transparency and where all relevant stakeholders are involved in an 'open, participatory process' [Partidário, 1996]. The valuation of non-economic impact categories inevitably requires communication, negotiation and sometimes even confrontation between decision makers and stakeholders. The definition of formal and explicit objectives, valuations and quantification methods with respect to environmental issues is therefore desirable.

5. Multiple objective optimisation of landuse allocation problems with genetic algorithms

5.1 Abstract

Spatial planning in densely populated regions has become increasingly complex over the past decades due to the increasing scarcity of space and the increased number of objectives and preconditions to be taken in account. Simultaneously, there has been a substantial development in information technology. As a result, computers, numerical models and measuring techniques confront decision makers and planners with an overwhelming abundance of data. Yet, spatial planners generally have not been able to benefit fully of these developments due to the lack of suitable optimisation techniques. Consequently, plans and scenarios are generally formulated without the use of optimisation techniques, which makes these plans and scenarios most likely sub-optimal. We developed a genetic algorithm for spatial multiple objective optimisation problems and applied it to case studies on landuse allocation (LUA). The technique enables the identification of unbiased near-optimal solutions by using the concept of Pareto efficiency.

5.2 Introduction

A number of developments have resulted in a substantial increase of the complexity of spatial planning. The use of geographic information systems and spatially distributed models has shown a steep increase over the past years, owing to the substantial technological development both in the field of computers and measuring techniques. If there are high quality data and distributed models for assessing impacts of decisions, somehow the scale and the refinement of policies and scenarios should improve as well in order to benefit fully from these improvements. Classic optimisation techniques such as linear and nonlinear programming can assist decision makers in formulating optimal configurations, but often cannot be applied satisfactorily in spatial planning. Spatial information is typically complex and variable, particularly if the underlying processes are interdependent and nonlinear by nature. Heuristic optimisation techniques such as Genetic

Algorithms and Simulated Annealing are potentially more appropriate in helping to develop optimal configurations for spatial planning. Although genetic algorithms have been applied successfully to a wide range of optimisation problems, application to spatial problems at a regional scale has been relatively rare, so far. We developed a genetic algorithm that is suitable for multi-objective optimisation of complex, spatial problems. We applied the optimisation technique to a case study on the optimal spatial distribution of agricultural production sites as to minimize adverse impacts to natural vegetation.

5.3 Background

The spatial distribution of different land use types in many parts of The Netherlands reflects a long history of agriculture in a poorly drained river delta. Areas used for cattle breeding and dairy farming are intermingled with small wetland patches with valuable natural vegetation. Many of these wetland areas are at present threatened by acidification due to nitrogen emissions from agricultural production [Last and Watling, 1991]. Governmental regulations limit agricultural production in order to protect wetland vegetation. Over the past years, there is a growing insight that improved conditions both for farmers and for nature could be achieved by a redistribution of the land use. National and provincial authorities have dedicated funds for covering the costs if farmers are willing to move to less vulnerable areas or to stop their production entirely. At this point a spatial planning problem emerges: how can the available budget be spent in such a way that wetland vegetation benefits maximally from buying out farmers? This optimisation problem concerns the allocation of different types of land use to lots within a delineated region. We investigated the feasibility of applying heuristic techniques to this optimisation problem.

The research project forms a follow-up of previous investigations carried out by RIVM-MNP, the Dutch national research institute for public health, environment and nature [Heuberger et al., 1997, Erisman et al., 1997]. During these previous studies a source- receptor model was developed to simulate the deposition of nitrogen emissions from agriculture in the Netherlands [de Leeuw and van Jaarsveld, 1992]. The entire Dutch territory was schematised in squared grid cells of 5x5 km and for every grid cell it was analysed how much all other grid cell contributed to the deposition of nitrogen. The resulting source-receptor matrix was used to calculate the optimal configuration of agricultural land use for minimal damage in nature areas by means of linear programming. The maximum acceptable deposition in nature areas was determined and expressed in a so-called critical load. The transport of nitrogen in the atmosphere was thus generalised as a linear process, with a system of linear equations that described the deposition of nitrogen.

In the source-receptor model every single grid cell was allocated an equation of the type:

$$D_i = A_{1,i} \cdot E_1 + \ldots + A_{N,i} \cdot E_N \qquad (5\text{-}1)$$

Where:

D_i	deposition in cell i
$A_{1,i}..A_{N,i}$	Set of N constants for cell i
E_j	emission from cell j
N	total number of cells

Heuberger (1992) and Erisman (1992) applied linear programming to determine the maximum total emission of nitrogen that would not result in an exceedence of a critical load in any cell.

With a grid cell size of 5x5 km there are in total 1684 grid cells, of which about 1000 grid cells represent areas with the vulnerable natural vegetation. Solving the problem of this size would take typically about half an hour CPU time in 1997. A reduction of the grid cell size would make the problem practically unsolvable by linear programming.

Aerts [2002] analysed the problem of optimal land use allocation and compared various methods, viz. simulated annealing (SA), mixed integer linear programming (MILP) and nonlinear integer programming (NLIP). He showed that the latter two methods become rapidly inappropriate if the total number of integer variables exceeds several hundreds. It can therefore be concluded that reducing the grid cell size in the optimisation problem described here above would make it unfeasible to solve the problem by means of linear programming. Furthermore, it is considered desirable to introduce additional objectives, such as the degree of fragmentation of homogeneous land use areas (patch size) and the costs and efforts that would be involved in the implementation of a particular land use allocation scenario. Introducing more objectives would make the optimisation problem much more complex and would therefore require more effective optimisation techniques. As a result of these considerations, a study was carried out to investigate which optimisation techniques could be suitable for these more complex land use allocation problems.

5.4 Problem description

The general optimisation problem that was studied concerns a land use allocation problem. There are two types of land use that can be allocated: nature and agriculture. The objectives are to minimise deposition of nitrogen from agriculture on nature areas and to minimise the degree in which agriculture is limited by regulations due to the presence of nearby nature areas. The optimisation problem is similar to those described in a number of publications by the RIVM-MNP, the Dutch national research institute for public health, environment and nature [Heuberger et al., 1997, Erisman et al., 1997]. The spatial distribution of nitrogen emissions and nature is represented in raster maps. The relation between emission and deposition is determined by means of a source receptor matrix. Three different optimisation problems were studied. A synthetic source receptor matrix was used for the optimisation problems A and B (see 5.5 and 5.6) as to enable verification of the results of the optimisation with the genetic algorithm. The source receptor matrix that was applied for optimisation problem C was calculated by means of a transport model for nitrogen that is available at the RIVM-NMP. In the transport model prevailing wind directions and site-specific properties are taken in account. For every raster cell is determined which relative amount of nitrogen deposition all cells receive from that particular cell. Multiplication of the source receptor matrix with the emission matrix results in the deposition matrix (see 5.3).

5.4.1 Objectives

The principal objectives that were applied in the optimisation sessions are:

1. Minimum exceedence of critical loads[4]
2. Minimum exceedence of maximum emission per cell
3. Minimum exceedence of the maximum surface area nature per cell
4. Minimum total area of nature to be moved
5. Minimum total emission of nitrogen to be moved

The first calculations consisted of single objective optimisations where only objective 1 was applied. Later, other objectives were added, in some calculations apart and in others lumped. Details of the various numerical experiments are describes in the next sections.

5.4.2 Model variables

The model area is represented as a collection of cells of equal size. The model variables are:

- critical load;
- nitrogen emissions from agriculture;
- nitrogen emissions from other functions within the model area;
- nitrogen emissions from outside the model area;
- nitrogen deposition;
- total surface area nature.

The critical load is the maximum deposition that is allowed in a particular cell. The degree in which the critical load is exceeded within the optimisation model for problem A and B is not only a function of the deposition in a cell, but also of the surface area of nature in a cell.

---

[4] Critical load is maximum allowable deposition, variable per cell

At the start of the optimisation the following data are known:

- the total quantity of emissions that can be moved;
- the total quantity of nature area that can be moved;
- the source receptor matrix;
- the critical loads per cell;
- the maximum emissions per cell;
- the maximum area of nature per cell;
- the emissions per cell that can not be moved;
- the depositions that stem from emissions from outside the model area per cell.

5.4.3 Optimisation questions

The questions that should be answered through the optimisation are:

1. Which configuration of nitrogen emissions that originates from agricultural activities from within the model area results in a minimum exceedence of critical loads?
2. Which configuration of nature results in a minimum exceedence of critical loads?
3. Which configuration of nitrogen emissions and nature results in a minimum exceedence of critical loads?

5.5 One-dimensional spatial optimisation – Problem A

A genetic algorithm was developed for the general optimisation problem that was described in paragraph 5.4.

5.5.1 Objective function

The objective function for minimum exceedence of critical loads is stated in the following:

$$O = Min\left(\sum (D[j]-Cl[j])^2 \forall (D[j] > Cl[j])\right) \qquad (5\text{-}2)$$

With constraint[5]:

$$E[j] < Em[j] \qquad (5\text{-}3)$$

Where:

O	objective
D[j]	deposition in cell j ($M/L^2/T$)
Cl[j]	critical load in cell j
E[j]	emission in cell j ($M/L^2/T$)
Em[j]	maximum allowed emission in cell j ($M/L^2/T$)

In equation 5-2, exceedence of the critical loads is squared as to achieve minimum variance of exceedences.

The question that corresponds to the objective function is:

Which configuration of nitrogen emissions that originates from agricultural activities from within the model area results in a minimum exceedence of critical loads?

5.5.2 Selection

Selection of solutions for reproduction is determined as a function of fitness: every generation 10 new solutions are created from five existing solutions with the highest degree of fitness and five solutions that are

[5] In this thesis the difference between an objective and a constraint consists in the following: an objective is something that is aimed to fulfil to a maximum degree whereas a constraint is a condition that is required to be fulfilled for a valid solution.

randomly selected. The 10 new solutions replace 10 existing solutions with the 10 lowest fitness scores.

5.5.3 Reproduction

The reproduction techniques that have been investigated are:

- uniform reproduction;
- arithmetic reproduction;
- multipoint reproduction.

These techniques also have been applied conjunctively. Furthermore a mutation mechanism was applied to ensure a good exploration of the search space and to prevent premature convergence.

5.5.4 Coding

A chromosome consists of a number of genes that equals the number of cells in the model area, multiplied by the number of model variables there are to be optimised. In problem A, there is only one model variable to be optimised, the spatial distribution of nitrogen emissions, i.e. the configuration of agricultural land use. The values of genes represent in this case the quantity of nitrogen emissions that originate from agricultural activities and that can be moved to other cells. The total movable nitrogen emission in the model area is assumed to be a value that is to be determined by decision makers as it is clearly dependent on the budgets that are available to improve conditions for both nature and agriculture. The value of a gene represents the movable emission as a fraction of the maximum emission in a particular cell. Thus, adjustment of the total movable emission within the model area can be carried out without the need to change the coding of the genetic algorithm:

$$E[j] = \frac{g[j] \cdot Em[j] \cdot Et}{\sum \left(g[j] \cdot Em[j] \right)} \qquad (5\text{-}4)$$

Where:

g[j]	value of gene j (0>=g[j]<=1) (-)
Et	total movable emission in the model area (($M/L^2/T$)
Em[j]	maximum allowed movable emission in cell j ($M/L^2/T$)

This type of optimisation problem requires a search for new configurations of emissions in the model area while the sum of emissions that are moved should be the same in all solutions taht are constructed. Therefore, values of genes cannot represent emissions directly, because the sum of emissions would then vary among solutions that are generated during the optimisation run. Correcting deviations from the required total movable emission in the model area by application of some kind of repair function is very inefficient and disturbes the evolutionary optimisation process. Therefore, the representation of emissions in values of genes is indirect (mapped), according to equation 5-4. Application of a 'mapping' to emissions per cell enables the use of classic GA reproduction techniques while maintaining nevertheless a fixed total movable emission of the model area. However, some possible combinations of gene values cause a violation of constraint 5-3. This possible violation has not been compensated by a reparation code as this would result in increased calculation times and because optimum fulfilment of the objective function also prevents a violation of this constraint. The determination of critical loads has been such that a critical load of an individual cell is always less than the maximum emission. If a deposition of a cell is less than the critical load, then the emission of that cell is less than the maximum emission.

5.5.5 Formulation of problem A that enables verification

Evaluation of the results of the test calculations can best be carried out if the properties of the optimum solution are known. Therefore, a limited number of additional model properties have been formulated:

$$\sum Cl[j] = \sum E[j] \qquad (5\text{-}5)$$

$$SRM[j, j] = 1 \qquad (5\text{-}6)$$

$$\forall\, SRM[k,l] = 0 \quad \wedge \quad k \neq l \qquad (5\text{-}7)$$

Where:

SRM[k,l] fraction of the emission in cell *l* that ends as deposition in cell *k*

Equation 5-5 states that the sum of all critical loads equals the sum of all movable emissions; equation 5-6 states that all deposition in a cell originates from emission in that cell. Thus is achieved that the optimisation problem has a unique and known solution and therefore can be verified, whereas the degree of difficulty of the optimisation problem when solved by means of the genetic algorithm, does not differ significantly from a real-world version, were deposition in cells originate from emissions from more than one single cell. The principal difference between the two formulations is that in the current formulation it can be verified if the best solution found is the true global optimum.

As a result of the formulation of the test problem, it holds for the known optimum solution that if the configuration of the emissions equals the configuration of the critical loads. The value of the object function for the global optimum is 0.

At the start of the calculations, a random configuration of critical loads is generated and stored as 'target configuration'. Then, the optimisation run is started and it is tried to find a configuration of emissions that equals the target configuration of critical loads by means of the genetic algorithm. Only one single optimum exists because the sum of the critical loads equals the sum of the emissions.

5.5.6 Premature convergence

Mutation prevents that the population is prematurely converged. Without mutation, a phenomenon which is known as *genetic drift* will cause the population to become genetically identical before the global optimum has been found. However a too high probability of mutation should be avoided because this would prevent that properties of successful individuals can be passed to offspring. The number of genes per individual is typically quite high in spatial optimisation problems. Relatively small probabilities of mutation are in those cases already sufficient to prevent premature convergence. If the probability of mutation is for instance five per thousand, it implies that on average about 200 genes would mutate per crossover if the optimisation problem would consist of 40,000 cells. In the current study a variable probability of mutation was applied as to find a good balance between avoiding pre-mature convergence on one hand and maintain good conditions for inheritance of successful properties on the other hand. The probability of mutation was calculated as a function of the genetic homogeneity within the entire population. This construction improved the performance of the genetic algorithm significantly as compared to calculations with a fixed probability of mutation.

With respect to inheritability it should be noted that as a result of the type of mapping that was applied a change of a limited number of genes (g[j]) leads to a change of the absolute values (V[j]) of many cells. This is a result of the coding of cell properties as described in equation 5-3. The advantage of the mapping is that the sum of emissions and the total area of nature of all solutions can easily be kept constant without the need for expensive correction methods, this at the expense of a mutual dependency of cell values. This interdependency causes a good exploration of the search space but reduces the ease by which successful properties can be inherited by offspring. Particularly in the last phases of an optimisation run it is necessary that only small changes of successful solutions occur. However, adverse impacts of this phenomenon are limited because relative small differences between solutions exist in the final stages of the optimisation process. In many cases there are only minor differences between the genetic properties of parent solutions in the final stages. As a result, the sum of all genes (g[j]) does not change significantly between parents and children and therefore individual cell properties (V[j]) can be inherited quite well.

5.5.7 Results of 1-dimensional spatial optimisation

The calculations have been carried out for two different model sizes, consisting of 10,000 and 40,000 cells respectively. The population size was 200 and the crossover techniques that were applied are: uniform, arithmetic and multipoint.

The relation between CPU time and the error in the best solution found so far is displayed in Figure 5-1. The calculations were carried out with a simplified version of the source receptor matrix model (see equation 5-5 and 5-6). No repair of function was applied for cases where V[j]>Vm[j] (see equation 2) because this condition is implicitly present in the objective function. The true optimal solution has a y-axis value of 0.

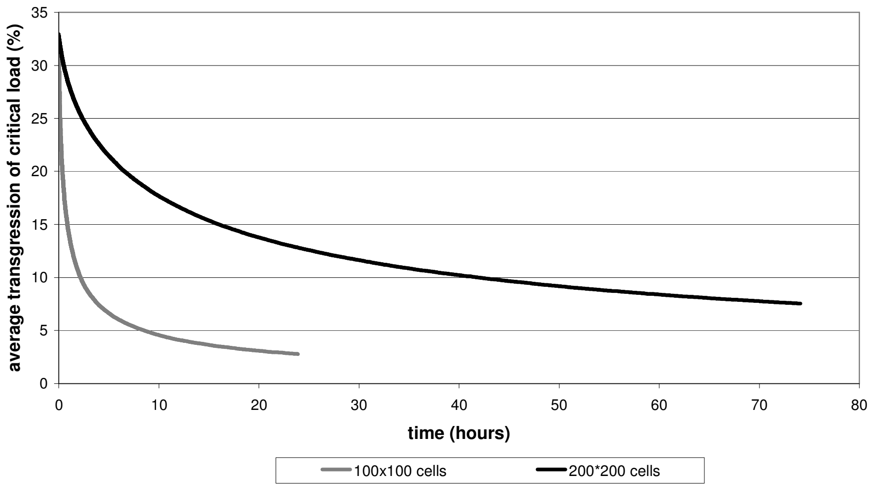

Figure 5-1 Results of 1-dimensional optimisation at model sizes of 10,000 and 40,000 cells.

5.5.8 Conclusions

Convergence to a (near) optimal solution is possible for optimisation problem A. Premature convergence can be avoided by application of an appropriate probability of mutation. Parallel calculation on different PCs is required if model sizes exceed 10,000 cells or if a realistic and therefore more expensive source receptor matrix model is applied.

5.6 Multiple objective optimisation – Problem B

Many real-world optimisation problems have more than one single objective. This feature complicates the optimisation because introduction of more objectives and more constraints make it more difficult to find optimal solutions. A technical solution to this problem can sometimes be found by joining different objective functions into a single lumped objective function. The only way to achieve this is to apply weights to the different components of the lumped objective function. If one wants to determine a pareto front, different combinations of weights have to be applied. Another possibility is that some of the objectives are formulated as constraints. However, such an approach requires a priori choices that are similar to the allocation of weights to different objective functions.

If an optimisation problem with multiple objectives is not translated into a single objective optimisation problem, either by application of weights or by additional constraints, then usually a Pareto front is the result of the optimisation, provided that the objectives are conflicting. Decision-makers can select a final solution from the collection of Pareto efficient solutions, thus implicitly allocating weights to the relevant objectives.

5.6.1 Objectives

The objectives that were defined for problem B are the following:

1. Minimum exceedence of critical loads
2. Minimum exceedence of maximum emissions per cell
3. Minimum exceedence of maximum areas nature per cell
4. Minimum change of nature to other locations
5. Minimum change of nitrogen emissions to other locations

5.6.2 Objective functions

The objective functions that correspond to the objectives stated in 5.6.1 are:

$$O_1 = Min\left(\sum\left(\frac{(D[j]-Cl[j])\cdot 100}{Cl[j]} \right) \forall (D[j] > Cl[j]) \right) \qquad (5\text{-}8)$$

$$O_2 = Min\left(\sum\left(\frac{(E[j]-Em[j])\cdot 100}{Em[j]} \right)^2 \forall (E[j] > Em[j]) + \sum\left(\frac{(N[j]-Nm[j])\cdot 100}{Nm[j]} \right)^2 \forall (N[j] > Nm[j]) \right) \qquad (5\text{-}9)$$

$$O_3 = Min\left(\frac{\sum\left(\frac{(E[j]-E1[j])\cdot 100}{\sum E1[j]}\right)\forall(E[j]>E1[j])+\sum\left(\frac{(N[j]-N1[j])\cdot 100}{\sum Nm[j]}\right)\forall(N[j]>N1[j])}{2}\right) \tag{5-10}$$

The constraints:

$$E[j] < Em[j] \tag{5-11}$$

$$N[j] < Nm[j] \tag{5-12}$$

$$\sum E[j] = C_1 \tag{5-13}$$

$$\sum N[j] = C_2 \tag{5-14}$$

Where:

O	objective function
Cl[j]	critical load in cell j ($M/L^2/T$)
D[j]	deposition in cell j ($M/L^2/T$)
E[j]	emission in cell j ($M/L^2/T$)
Em[j]	maximum allowable emission in cell j ($M/L^2/T$)
N[j]	area nature in cell j (L^2)
Nm[j]	maximum allowable area nature in cell j (L^2)
E1[j]	initial emission in cell j ($M/L^2/T$)
N1[j]	initial area nature in cell j (L^2)
C_1	constant 1; total movable nitrogen emission in the model area
C2	constant 2; total movable area of nature in the model area

Objective function 1 (5-8) corresponds with objective 1, minimum exceedence of critical loads. Objective function 2 (5-9) corresponds with objectives 2 and 3. Objective function 3 (5-10) corresponds with objectives 4 and 5.

5.6.3 Coding

The coding of the cell properties in the genetic algorithm is analogue to that of the one-dimensional optimisation problem. The number of genes of a chromosome has increased since in this test problem not only emissions can be moved, but also nature areas. The number of genes per chromosome equals the product of the number of cell properties multiplied by the number of cells. A model size of 200 x 200 cells and two parameters results thus in 80,000 variables per chromosome. It should be noted that the level of detail in which solutions can be described is higher than the size of a cell because the variables represent for nature the fraction of the cell area that is dedicated to nature and for emission The fraction of the maximum emission of that particular cell. Thus, the coding enables a more detailed description of solutions then a coding where a cell processes either the property “nature” or the property “emission”.

5.6.4 Selection

Selection of promising solutions in an optimisation problem with multiple objectives cannot be based on the results of a single objective function. If the two objectives are conflicting and refer to different, incommensurable values, subjective valuation is necessary to identify the optimum solution. A Pareto front shows optimum solutions for different weightings ('exchange rates') between the two objective functions. There are two ways to calculate the Pareto front. The classic approach consists of carrying out a set of optimisation runs, each with a different weighting of the two objective functions. The approach in this thesis consists of formulation of an unbiased objective function based on two properties: Pareto efficiency (1) and Unicity (2). Pareto efficiency is a relative Boolean property. The solution is Pareto efficient if there is no other solution within the population that has a better score with respect to at least one objective function, without having a worse score with respect to all other objective functions. The unicity of a solution is intermittently defined as:

$$U[j] = \sum_{i=1..M} \left(\frac{\sum_{j=1..N} \frac{I[i,j]-I[i,k]}{Ia[i]-Ib[i]} \quad \wedge j \neq k}{N-1} \right)^2 \qquad (5\text{-}15)$$

and:

$$U[j] = \sum_{i=1..M} \left(Min\left(\frac{I[i,j]-I[i,k]}{Ia[i]-Ib[i]} \quad \wedge j \neq k \right)^2 \right) \qquad (5\text{-}16)$$

Where:

U	unicity (-)
I[i,j]	impact score of cell *j* for objective function *i* (-)
Ia[i]	maximum score of objective function *i*
Ib[i]	minimum score of objective function *i*
N	number of cells
M	number of objective functions

The term *N-1* is not necesarry for the optimisation process but facilitates comparing results of different optimisation problems with different numbers of cells.

Both definitions of unicity have been applied alternatively during the optimisation runs to prevent premature convergence and to improve the exploration of the search space. The first equation (5-15) relates unicity of a solution to the difference with the normalised mean of objective function scores, whereas in the second form (5-16) unicity is related to the minimal normalised difference with other solutions. Both formulations result in a different ranking of total fitness and thus in a different exploration of the search space.

5.6.5 Formulation of a known optimum for optimisation problem B

As in the test for optimisation problem A, a specific formulation was applied as to define a problem with a known solution. Just like in problem A, the total quantity of emission was set equal to the total sum of critical loads. The source receptor matrix is also for problem B defined as a simplified version of the true matrix. This implies that the configuration of emissions in the optimum solution is equal to the configuration of critical loads.

Objective function 1 describes the degree of exceedence of critical loads by depositions. During the test calculations, no weights haven been applied in this objective function for the area of nature in cells. Thus is achieved that only a single solution is possible without exceedence of maximum emissions. The second objective function concerns the degree in which the maximum values of nature area and emissions in cells are being exceeded. This criterion is rather a constraint than an objective, but has been formulated as an objective function as to prevent that many solutions that are being generated during the optimisation process will have to be repaired. The third objective function concerns a minimisation of the quantity of nature area and the emissions that have to be moved to other locations. For the initial situation we chose a configuration of nature and emissions that actually corresponds to the optimal configuration, where the configuration of emissions corresponds with the configuration of critical loads. This was combined with a random configuration of nature areas. Thus it became possible to determine easily what percentage any solution deviates from the true optimum. According to this approach, the configuration of nature areas is only relevant for objective function 3, whereas in a real world application the configuration of critical loads is also partly determined by the configuration of nature areas. The set up of the testing as described here above enables verification of the performance of the genetic algorithm but also implies that the objective functions are not completely conflicting. As a result of the specific formulation of the optimisation problem for testing purposes, all objective functions will display 0 for the true optimum and there will be no other Pareto-efficient solutions than the true optimum.

The values that were applied are displayed in Table 5-1; the initial configuration of model properties was determined randomly.

Table 5-1 Applied values in the testing model for optimisation problem B

	Minimum	**Maximum**	**Mean**
Emission	0	1000	500
Area with nature	0	100	50

5.6.6 Results of multiple objective optimisation – problem B

As was described in the previous section, initially three objective functions were formulated, although objective function 2 is in practical applications rather a constraint than an objective function, considering the great weight that decision-makers attach to this objective. As to investigate the performance of an alternative approach we also carried out experiments where we reformulated objective function 2 as a constraint. A repair function had to be implemented for correction of solutions where the new constraint was not fulfilled. The effectiveness of either of the approaches depends on the severity of the constraint. A repair function will be demanding in terms of CPU time if it concerns a condition that is difficult to fulfil. Besides, the "heuristic path" that is followed by the genetic algorithm is disrupted if a repair function is applied. In those cases an additional objective function in stead of a constraint may be more effective, since no CPU time is required for reparation of invalid solutions and the "heuristic path" is not cut off. Indicative numerical experiments showed that a constraint that can be fulfilled relatively easily and that is therefore not very demanding for reparation is more effective than introduction of an additional objective. Introduction of more objectives has also an adverse impact on the speed that (near) true Pareto-efficient solutions can be identified. The search power of a genetic algorithm becomes diluted if more objectives are introduced because the total surface of the pareto fronts is increased.

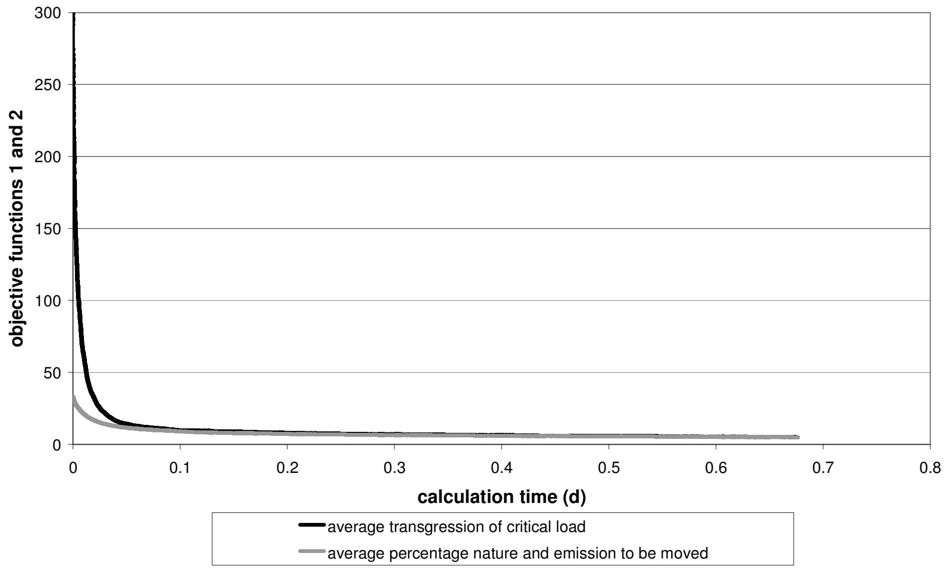

Figure 5-2 Results of 2-dimensional optimisation for a model with 100 x 100 cells

The values for the two objective functions are displayed in Figure 5-2 as a function of calculation time.

1. Average exceedence of the critical loads: $\frac{O_1}{n}$
2. Average percentage of nature and emissions that is to be moved (compared to the total quantities of nature and emissions): O_3 (%)

5.6.7 Conclusions

Comparison of the performance of the genetic algorithm for the different versions of the optimisation problems that were investigated make clear that the required calculation time increases exponentially if the size of the problems increases linearly. An extremely simplified source receptor matrix has been used for the test problems of optimisation problem B, as described here above. More realistic calculations of nitrogen transport will require a more complicated source receptor matrix and the required CPU time will consequently be much higher. Parallel computing can offer a solution to this problem because genetic algorithms are very suitable for splitting up the calculations to different PCs.

The severity of constraints is decisive for the success of reparation techniques. Boundary conditions that are mild and relatively easily to fulfil, such as the constraints regarding the maximum values of emissions and nature areas in problem B, enable more effective application of a repair function than formulation of a constraint as an additional objective. Application of a repair function becomes less effective in the constraint is severe. This would be the case if, for instance, maximum emissions of nitrogen in problem B would differ only a very little from critical loads. In that case, many invalid solutions would be generated during the optimisation process and it would be expensive in terms of CPU time to repair them. In those cases it becomes increasingly interesting to apply alternative mappings or transformations, such as was applied in equation 5-4 in order to keep the total sums of emissions and nature areas constant. If that is not possible, it may be interesting to reformulate the constraint as an additional objective function. Although this would reduce the focus and therefore the search power of the algorithm, it would not disturb the heuristic path towards gradual improvement of solutions.

5.7 Multiple objective optimisation – problem C

Optimisation problem C has been formulated by the provincial authorities of Noord Brabant (Netherlands) and RIVM-MNP, the Dutch national research institute for public health, environment and nature. The definition and the size of the problem are such that it can be solved by linear programming too. The optimisation problem concerns the spatial distribution of nitrogen emissions from agricultural activities in Noord

Brabant Province. The spatial distribution should be such that vulnerable nature areas receive minimal nitrogen deposition from agricultural activities and that agricultural areas are submitted to minimal restrictions due to the proximity of nature areas. Like in the optimisation problems that were described in the previous sections, spatial properties of the problem are represented by raster maps. The raster cells represent either agricultural land use or nature areas. This representation is different from the optimisation problems A and B, where within a single raster cell more than one single type of land use could be allocated. Although the level of detail of spatial configurations is thus reduced as compared to the optimisation studies that were described in the previous sections, it enables comparison of the optimisation results with linear programming techniques.

The model area consists of two subareas. The total emissions of the two different subareas are to remain constant. Export of emissions to areas outside the boundaries of the model should also remain constant. Fulfilment of the latter constraint is implemented by creating an additional ring of raster cells around the model area and by calculating deposition in this additional subarea. As long as total deposition in this subarea is not increased, it is assumed that there is no additional export of emissions outside of the model area. Since this “zero change of export” constraint can not be fulfilled effectively by implementation of a repair function, it was formulated as an additional objective function.

5.7.1 Objectives

The objectives 1, 2 and 3 were formulated as objective functions whereas the objectives 4, 5 and 6 were formulated as constraints. Fulfilment of the constraints was achieved by application of a mapping technique combined with a repair function (see 5.6)

1. Minimum exceedence of critical loads
2. Minimum additional exceedence of critical loads in cells that represent existing nature areas where at present critical loads are already too high
3. Minimum change of total emission that is exported out of the model area
4. Minimum exceedence of maximum emissions per raster cell

5. Minimum falling short of minimum emissions per raster cell
6. Minimum change of the total emission within the model area and also minimum change of the total emission of the two subareas of the model area

5.7.2 Objective functions

The objective functions that correspond to the objectives stated in 5.7.1 are:

$$O_1 = Min\left(\sum(D[j] - Cl[j])\ \forall(D[j] > Cl[j])\right) \tag{5-17}$$

$$O_2 = Min\left(\sum(D[j] - Cl[j])\forall\ (D[j] > Di[j] \wedge Di[j] > Cl[j])\right) \tag{5-18}$$

$$O_3 = Min\left(\sum(D[k] - Cl[k])\forall\ (D[k] > Cl[k])\right) \tag{5-19}$$

The constraints:

1. $E[i] \leq E\max[i]$ (5-20)
2. $E[i] \geq E\min[i]$ (5-21)
3. $\sum E[m] = C_1$ (5-22)
4. $\sum E[n] = C_2$ (5-23)

Where:

i	index of a raster cell
j	index of a raster cell that represents nature
k	index of a raster cell that delimits the model area
m	index of a raster cell from subarea 1

n	index of a raster cell from Subarea 1
O	objective function
$Cl[j]$	critical load in raster cell j ($M/L^2/T$)
$D[j]$	the deposition in cell j ($M/L^2/T$)
$Di[j]$	initial deposition in cell j ($M/L^2/T$)
$E[j]$	emission in cell j ($M/L^2/T$)
$E_{max}[i]$	maximum emission in cell i ($M/L^2/T$)
$E_{min}[i]$	minimum emission in cell i ($M/L^2/T$)
C_1	constant 1; total emission in subarea 1
$C2$	constant 2; total emission in subarea 2

The objectives 1, 2 and 3 (see 5.7.1) correspond with objective functions 1, 2 and 3. The objectives 4, 5 and 6 have been implemented as constraints.

5.7.3 Coding and selection

The coding of the cell properties is analogous to the coding in optimisation problem B. The total emissions per subarea can be kept constant by means of the mapping that was described in equation 5-4. However, incidentally it may occur that the emission in particular cells exceeds the maximum emission. If this occurs, a repair function is called. By means of the repair function the values of genes of that solution are changed randomly until the constraint is fulfilled. Falling short of the minimum emission cannot occur because within the model there is reference to movable emissions only.

Selection of promising solutions is not identical to the selection in optimisation problem B. Pareto efficiency becomes increasingly unspecific if the number of objectives is increased, because the probability that a solution is Pareto inferior becomes less if the number of objectives is increased. A solution is Pareto inferior if there is at least one other solution that performs better with respect to at least one objective criterion while not performing worse with respect to any other objective criterion. If there are more objective functions, then it is more probable that there is in other solutions at least one property where the performance is less. Consequently, Pareto inferior solutions become scarce if there are many different objectives and therefore Pareto efficiency is less effective as a selection criterion.

A solution to this problem was found by introducing the concept of *Pareto score*. The Pareto score of a solution represents the number of pairs of objective functions for which a solution is Pareto efficient. This criterion enables again to select promising solutions effectively. During the calculation sessions it was clearly shown that application of the Pareto score resulted in a faster and better convergence to global optima.

5.7.4 Results

The best results with the genetic algorithm were achieved with a variable probability of mutation. The probability was changed if stagnation of improvement occurred. The criterion for selection consisted of the product of the unicity score and the Pareto efficiency score. The first score indicates the rank of a solution with respect to it's unicity in terms of values of the various objective functions. The Pareto efficiency score indicates the rank of a solution with respect to the number of pairs of objective categories for which the solution is Pareto efficient.

The level of detail of the SRM model was gradually increased during the optimisation as to reduce the computational cost in the initial stages of the process.

The results of the calculations by linear programming, simulated annealing and the genetic algorithm are presented in Table 5-2. The calculations with the linear programming code were carried out at the RIVM by Peter Heuberger [Vink, Heuberger and Bakema, 2002]. The calculations with the simulated annealing code were carried out at Utrecht University by Willem Loonen. The differences between the results of the three techniques for the current optimisation problem are negligible.

Table 5-2 Optimum results achieved by linear programming, simulated annealing and the genetic algorithm

	Objective 1	**Objective 2**	**Objective 3**
Result LP	262.0	0.000	0.0
Result SA	262.0	0.000	0.0
Result GA	262.1	0.009	4.0

Results of the optimisation after 50,000 generations are given in Figure 5-3 - Figure 5-5.

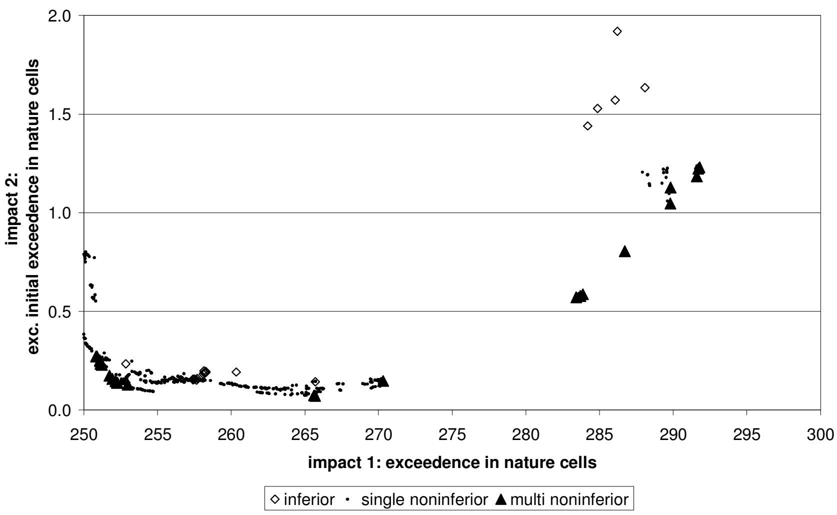

Figure 5-3 Results of optimisation problem 3, Pareto surface projected on objective function 1 - 2

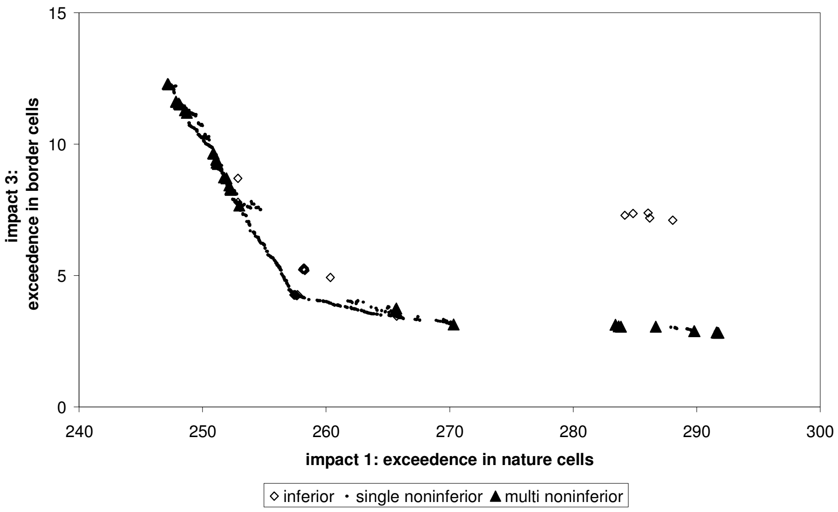

Figure 5-4 Results of optimisation problem 3, Pareto surface projected on objective function 1 - 3

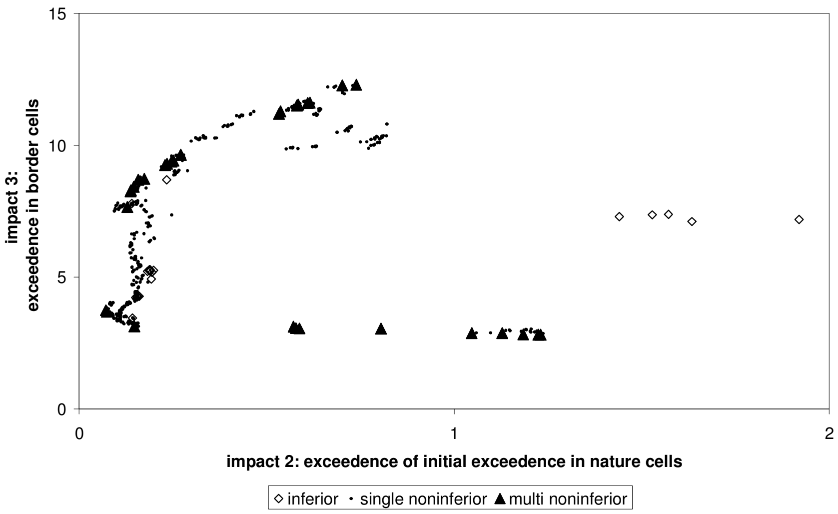

Figure 5-5 Results of optimisation problem 3, Pareto surface projected on objective function 2 - 3

Legend of figures *5-3 - 5-5*

- triangle: Solutions that are Pareto-efficient for more than one pair of objective functions (Pareto score >1)
- small dot: Solutions that are Pareto-efficient for one pair of objective functions (Pareto score =1)
- open dot: Pareto inferior solutions.

5.7.5 Conclusions and discussion

The true optimum could be identified by linear programming and simulated annealing, whereas the near optimum could be identified by the genetic algorithm. However, the potential level of detail of the results that were achieved with the GA is considerably higher than with LP and SA. With SA and LP the optimisation problem was treated with less detail, as only uniform types of land use could be allocated to single raster cells. With the genetic algorithm the problem was treated at a higher level of detail, as the allocated land use types within raster cells were expressed as a fraction of the total area within. Since the optimisation problem was designed for a comparison between the three techniques, the optimum solution consisted of a configuration with uniform land use within single raster cells. However, in real world problems it would be an advantage to have access to solutions at a detailed scale. The computational costs of the genetic algorithm were relatively high, but the results consist not only of the single optimum but also of the Pareto fronts. The number of raster cells corresponded to the maximum number that could be handled by the LP for this particular optimisation problem, but SA and GA could have handled a larger number of raster cells too.

The results of the GA possibly could be improved by formulating the optimisation problem as a single objective optimisation problem. The experiences with optimisation problem 2 showed that a reduction of the number of objective functions can improve the efficiency of the algorithm. Pareto fronts can be constructed by carrying out multiple calculation sessions with different weight ratios to the objective functions.

The results for optimisation problem 3 were validated to some extent by applying various methods to the same problem. Achieving some form of circumstantial validation of the results without applying other techniques is not trivial for the current problem since the SRM model plays a role in all objective functions. The most suitable technique for circumstantial validation of a problem that is too large to be handled by a global optimisation technique would consist of resizing the optimisation problem in such a manner that it just can be solved by a global technique. If the GA performs well for the smaller problem, the reliability of the results for the larger problem has slightly increased.

Apart from circumstantial validation, it is also important to consider the accuracy and reliability of the simulation models involved, and the precision that is feasible in an implementation of the results in practice. It is likely that results of optimisation with heuristic techniques are often rather near optimal than optimal. The results can be appropriate if the differences between near optimal and true optimal are sufficiently small. Therefore it is important to be able to assess the precision of the impact models and the precision that is feasible in the implementation phase. There should be accordance between the level of detail in solutions that are generated by the GA and the level of detail that is feasible in simulations of impact models and in implementation of solutions in practice. This property can be manipulated to some extent by changing the way the problem is coded into the genetic algorithm. It should be avoided that the algorithm generates many different solutions that do not lead to different impacts and do neither differ significantly from the viewpoint of simulation and implementation.

Whether or not an optimisation problem can be solved with GA depends on the properties of the problem. As yet, no generic rules for the assessment of the suitability of GA for a particular problem have been identified. Analysis of the experiences with GA in a number of case studies resulted in the identification of five properties that determine the feasibility of the application of a genetic algorithm to an optimisation problem:

- size of the search space;
- continuity of the search space;
- computational cost of impact models;
- shape of the search space.

These properties will be discussed in the next sessions.

Size of the search space

The size of the search space is determined by the number and the ranges of the variables that are involved in the optimisation problem. The higher the number and the larger the ranges, the more difficult it is to solve an optimisation problem.

Continuity of the search space

The continuity of the search space indicates in which degree global optimiser can be found by hill climbing. The higher the number of local optima, the more difficult an optimisation problem can be solved. The type of spatial optimisation problem that is described in this chapter probably has only a few local optima. If there were many local optima, the size of the current problem would make it almost impossible to identify the global optima. The success of the simulated annealing algorithm that was applied by Loonen en Heuberger [Loonen et al., 2004] confirms this assumption, since the optimisation consisted of a sequence of very small changes to properties of solutions. Only very few stagnation points were found. These could be 'overtaken' by changing temporarily the weights that were allocated to the different objective functions.

Computational costs of impact models

The computational costs are typically concentrated in the impact models in ecological and agricultural optimisation problems. The impact models often consist of highly complex simulation models in which spatial distributedness and sometimes also a temporal dimension are taken in account. The required calculation time may become a serious limitation for heuristic techniques under these circumstances. Many researchers tried to reduce the computational costs by gradually increasing the level of detail of the simulation models. Other options consist in paralellization and upscaling techniques [e.g. Haberlandt et al., 2002, Seppelt and Voinov, 2002].

Shape of the search space

The shape of the search space refers to the part of the collection of all possible solutions that is (near) optimal. Some optimisation problems have quite a large proportion of near optimum solutions that do not differ significantly, whereas other problems may have only a very small number of near optimum solutions. The latter type of optimisation problems is obviously more difficult to solve.

6. An analysis of different strategies for the prioritisation of groundwater quality prediction studies with a sequential numerical game[6]

6.1 Abstract

Groundwater quality prediction studies are carried out to increase the reaction time when drinking water companies have to respond to breakthroughs of contaminants. Drinking water companies exploit numerous wells and need to decide on research priorities for these wells, as budgets are limited. The reliability and accuracy of predictions improve if more funds are invested in data-collection and prediction studies, but there is no clear decision model available to determine the required level of (un)certainty. Hence, it is unclear which prioritisation strategy is optimal. Unnecessary losses can occur if inappropriate strategies are followed. A decision analysis of strategies for prioritising prediction studies is presented, where the problem is posed as an optimisation problem with an explicit loss function. A sequential numerical game was set up in order to assess the effectiveness of different strategies. There were significant differences between the performances of strategies. The most successful strategy used the anticipated uncertainty reduction of additional studies as one of the prioritisation criteria and takes the uncertainty of predictions in to account.

Keywords: uncertainty reduction, technological forecasting, groundwater quality prediction, research prioritisation, game theory, decision analysis.

---

[6] Adapted from: Proceedings of the Modelcare conference of 2005 in The Hague.

6.2 Introduction

Regional drinking water supply systems in densely populated areas typically consist of multiple sources and sinks, interconnected by transport pipes. The sources are production wells that pump groundwater or surface water, the sinks represent water users. Predictive studies on the chemical composition of pumped groundwater are carried out in order to reduce the risks of failure of production wells due to contamination. These studies function as an early warning system, providing time for taking counter measures and thus enabling a reduction of the potential consequences in case some contaminant would leave pumped groundwater quality unsuitable for drinking water production. Contamination of wells can lead to high economic costs because the construction of a new well at a different location involves considerable investments in research and infrastructure. If a contamination reaches a well before a replacing well has become available then the required capacity needs to be obtained temporarily from other wells which also may involve high economic and environmental cost. In the worst situations there may be insufficient finance, time or spare capacity and the supply is hampered. Generally the number of feasible remedial actions decreases if the available reaction time is reduced and the costs of remaining options increase. Early recognition of an upcoming breakthrough can therefore reduce the adverse impacts. The quality of groundwater has deteriorated in many regions over the past decades due to agricultural and industrial pollution, national and international standards for drinking water quality have become more stringent and prediction studies have therefore gained importance. As a result, many drinking water companies need to spend substantial budgets on monitoring and prediction of groundwater quality. Yet, there seems to be no uniform strategy applied to the prioritisation problem of prediction studies. Prioritisation of research is required as budgets are limited, but which prioritisation strategy to choose is not a trivial question: should decision makers aim for minimization of total risk or minimization of maximum risk? What is a suitable operational definition of the risk of well failure? How should decision makers account for the uncertainty of predictions? Rational and consistent methods are needed in order to spend available budgets efficiently. However, prioritisation of prediction studies is often based on ad hoc strategies, as more advanced strategies require complex assessments due to the inherent uncertainty in predictions and to the complexity of many present-day regional drinking water supply systems. Best professional judgment, expert judgment and educated guesses may result in sub optimal prioritisation. *Freeze et al.* [1987] developed a method for the assessment of data worth in groundwater contamination problems. Finkel and Evans [1987] evaluated the benefits of uncertainty reduction in environmental health risk management. Reichard et al. [1990] provided a health risk oriented benefit-cost analysis as a conceptual framework for groundwater management under uncertainty. These studies emphasized the importance to show that the value of data-collection strategies can and should be valued in terms of their expected impact on decision making. The problem of trend detection in water quality data and the optimal design of monitoring networks and sampling strategies have received considerable attention over the past decades [see e.g. Dixon et al., 1996]. The focus of the aforementioned studies considered rather methods to determine the cost and value of information than methods to determine the optimal distribution of an already specified budget, as is the subject of this paper. In a more general sense, decision making under uncertainty has been addressed in mathematics by probability theory and utility theory. In contrast with the rare literature on prioritisation of prediction studies, many papers were dedicated to quantifying the uncertainty of predictive simulations. Monte Carlo simulations, Kalman filtering, Kriging and other techniques have been applied for assessment of the uncertainty of data and model results [e.g. Carrera et al., 1984, Dagan, 1986, Gelhar, 1986, Neuman, 1987, van Geer, 1987, Delhomme, 1978]. The use of some of these techniques is currently on its way to become common practice in applied research. The integration of the achievements of the latter studies in decision making strategies for prioritisation of prediction studies has thus become an interesting option and forms the starting point of the analysis that is presented in this paper.

In the next section the objectives of predictive studies and some operational definitions for relevant system properties are discussed. The analysis results in the identification of a number of possible criteria for prioritisation of groundwater quality prediction studies. This section is followed by a description of the general setup and results of the game experiment, which was used as an environment for testing the performance of a number of strategies. The paper is concluded with conclusions and a discussion of the results.

6.3 Methods

We constructed a conceptual sequential game model to investigate the effectiveness of various prioritisation strategies, measured in terms of losses due to breakthroughs. Apart from the strategies, the game consists of a "stochastic well properties generator", including time series of concentrations of pumped groundwater, an uncertainty reduction function and a loss function, related to the impact of an upcoming exceedence of a concentration limit. By allocating research budgets to wells, the virtual players/decision makers can reduce the uncertainty of predictions of the future concentration of pumped groundwater of these wells. Uncertainty reduction results in an increase of the expected *minimum reliable reaction time* and sometimes in a sufficiently reliable prediction of a exceedence of a quality standard for drinking water, i.e. a prediction of the *maximum reliable reaction time* (Figure 6-2 and Figure 6-3).

6.3.1 Risk of breakthrough

A breakthrough occurs if the concentration of a solute in pumped groundwater exceeds a certain threshold value. In this context, threshold values determine the suitability of water for drinking water production based on toxicological and/or technical considerations. Prediction studies can contribute to the reduction of risks of breakthroughs by expanding the available time for remedial actions.

A general definition of risk according to the British Standard Institution describes risk as "the combined effect of the probability of occurrence of an undesirable event, and the magnitude of the event". [Griffith, 1981]

The operational definition of the risk of well failure due to the breakthrough of a contaminant we define accordingly as:

$$R = P * I \qquad (6\text{-}1)$$

Where:

R	risk
P	probability of breakthrough
I	impact of breakthrough

Some management options in the case of a threatening well failure[7] are directed at reducing the probability that the event will take place, other options are directed to a reduction of the impact of a breakthrough. There are various remedial actions possible if well failure due to the breakthrough of a pollutant is expected. The major counter measures are:

- transferring of the production to a newly constructed well;
- transferring production to another well as a provisional solution;
- timely installing appropriate purification capacity;
- changing the caption zone of a well by changing the pumping regime and thus influencing the future composition of the pumped groundwater;
- purchasing land and modifying the land use in order to reduce the pollutant load that enters the groundwater.

The impact of well failure varies among wells because:

- the construction cost of new wells vary according to local conditions and according to the required capacity;
- the economic and environmental impact of temporary transfer of production to another well as a provisional solution depends on the role of a well within a regional distribution network;
- the reaction time varies.

Not all wells are therefore of equal importance to the functioning of a regional drinking water supply system. Some wells may play a crucial role

---

[7] Only well failures caused by contamination of groundwater are considered in this paper.

within a system while others can be missed with relatively small consequences, for instance if the production can easily be transferred to other wells where spare capacity is available. Allocation of research budgets for groundwater quality prediction proportional to the potential consequences of well failure is then a logical step. However, assessment of the impact of a well failure is a complex task. Temporary transfer of the production capacity from a contaminated well to other wells within the same regional supply system works out differently for every well, due to the relatively large spatial variability of many parameters. Wells, purification plants and transport facilities vary in capacity and in economic and environmental efficiency. The inherent combinatorial explosiveness of network systems makes that there are often many ways to transfer the needed capacity to other wells. Each solution will have different impacts on economic, environmental and other objective categories and nonlinearity and interdependency of well's impact relations as a function of pumping rate make it generally infeasible to identify the optimum (minimum negative impact of well failure) without computer-based modelling and optimisation techniques. Valuation of the relevant impact categories such as for instance economic costs, environmental costs and reliability of supply is required before the optimal configuration can be identified. A method for multi-objective optimisation of drinking water production by means of genetic algorithms is presented in [Vink and Schot, 2001]. According to this method the optimal i.e. minimized negative impact of a particular well failure is determined by optimisation of the production configurations of a regional system with and without that well respectively.

The magnitude of adverse impacts of breakthroughs also depends on the available reaction time (Figure 6-1). The available reaction time at the instant that an upcoming breakthrough is predicted reliably is therefore one of the decisive factors that determine the actual impact of a breakthrough and thus the magnitude of the risk.

$$Ib = f(Ib0, rt) \qquad (6\text{-}2)$$

Where:

Ib	manifest impact of breakthrough
Ib0	potential impact of breakthrough (Ib=Ib0 if rt=0)
rt	reaction time

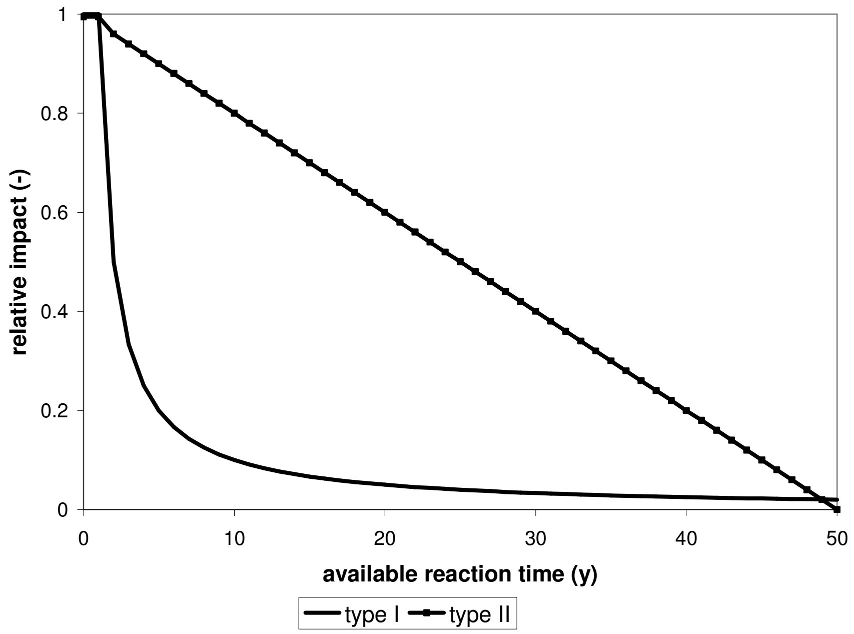

Figure 6-1 Two extreme scenarios for impact - reaction time functions

The relevance of reaction time concerns not only the financial means that are required for remedial actions; it also concerns the feasibility of measures. In cases that ample reaction time is available there may be sufficient time to construct a new well without the need for urgency, but when an upcoming contamination is predicted only shortly before the breakthrough occurs, provisional measures may be necessary in order to maintain continuity of supply. Urgent measures are usually more expensive and provisional measures more often invoke adverse environmental impacts, such as high energy requirements for transportation or damage to groundwater dependent ecosystems due to changed production configurations. Impact-reaction time functions vary widely according to case-specific conditions but all cost - reaction time functions have in common that an increase of reaction time will never result in an increase of total failure impact. Two extreme scenarios of generalized cost - reaction time functions are shown in Figure 6-1.

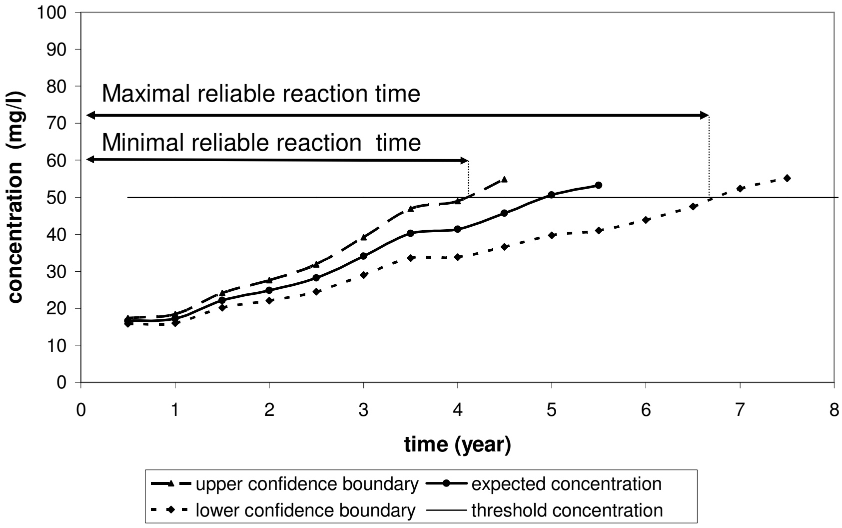

Figure 6-2 Minimum and maximum reliable reaction times

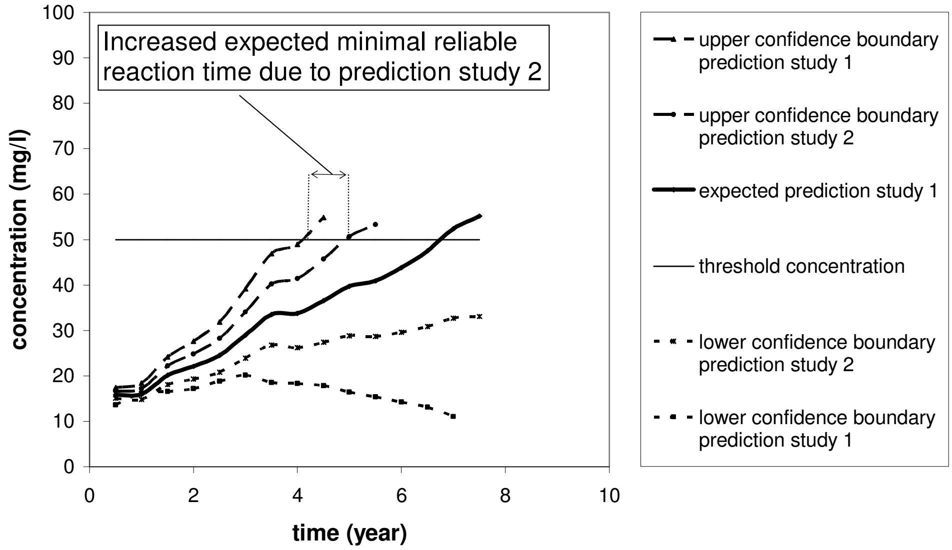

Figure 6-3 Increased expected minimum reliable reaction time due to an additional prediction study

Probabilities of breakthroughs are like all probability statements concerning the future essentially subjective and without a physical basis [Szaniawski, 1998]. They are rather a result of an imperfect interpretation of an incomplete representation of a set of conditions and processes. This implies that probabilities of breakthroughs depend on the accuracy and reliability of prediction studies. The reliability of the statements can be defined at least partially by applying uncertainty assessment techniques and by making the basis of predictions as explicit and accessible to verification as possible. A prediction study that includes a quantitative uncertainty analysis not only produces a predicted concentration time series but also an upper and lower boundary of the associated confidence interval time series (Figure 6-2). The size of the confidence intervals increase with increasing time horizon due to the possible accumulation of prediction errors and due to the fact that predictions for the longer term correspond to parts of the caption zone where monitoring activities generally are scarcer then near the well. Figure 6-3 shows how additional research can narrow the space between the upper and lower boundaries of the confidence intervals. Increased accuracy of predictions at a uniform confidence level results in narrowed boundaries of the confidence intervals. The concepts minimum and maximum reliable reaction time correspond to the time period between the present and the time instant when the threshold concentration intersects with the upper and lower boundaries of the confidence interval time series (Figure 6-2). Improved accuracy due to additional prediction studies will result in a narrowed confidence interval time series and consequently in an increase of the expected value of 'minimum reliable reaction time' and a reduction of the risk of well failure.

The probability that a concentration in pumped groundwater will exceed the threshold concentration at a time instant t can be determined if the probability density function of the predicted concentration at that time instant is available. If the confidence boundaries of all prediction studies apply a uniform confidence level then the 'discounted' impact (equation 2) that corresponds to the minimal reliable reaction time can be used as a measure of risk.

The potential reduction of risk depends on the uncertainty reduction that forms the result of a prediction study. The uncertainty reduction per unit of research budget can be highly variable. Dependent on site-specific conditions such as the size of the caption zone and the spatial and temporal variability of relevant geohydrological variables it may be feasible to improve the accuracy of predictions significantly at relatively low cost for some wells whereas for other wells high investments are required for only a modest improvement of the accuracy of predictions. Besides, it can be assumed that there is a decreasing marginal utility of research investments within a particular time domain and that consequently the uncertainty reduction depends also on the amount of budget that has been spent already.

In practical situations research priorities are allocated according to many different considerations, partly related to the potential impact and partly related to the probability of a breakthrough of a well. Assessment of the probability of a breakthrough may for instance be based on changes of trends in observed concentrations or on the pollutant load in the caption zone. In this study the prioritisation problem is considered without these types of information. The question which total budget should be allocated to prediction studies is also beyond the scope of this paper; we consider a situation where a periodical budget has been defined and is to be distributed over the available wells. The performance of different prioritisation strategies cannot be compared easily in field experiments as the frequency of well failure due to contamination is too low for obtaining significant results at acceptable levels of cost and time. Comparison of strategies that are applied to different regional supply systems is furthermore hampered by the large variability of the relevant properties of individual wells and of the distribution network. Identification of the optimal strategy on a basis of the generalization of field data is therefore considered hardly feasible. An analytical mathematical analysis is considered infeasible too. If possible at all it would probably lack the flexibility of handling a variety of often discontinuous and nonlinear impact-reaction time and cost-uncertainty reduction functions. We chose to investigate the prioritisation problem in the setting of a sequential game experiment. The adopted approach enables a discrete and flexible handling of possibly discontinuous and nonlinear functions. In this paper the problem was analysed under generalized conditions, but for specific drinking water supply systems it can be applied within the framework of a decision-support model. There, the game properties can be adjusted to system-specific conditions by which the effectiveness of strategies in specific situations can be investigated.

6.4 Sequential game

We constructed a sequential game model in order to investigate the effectiveness of various prioritisation strategies for groundwater prediction studies, measured in terms of losses due to breakthroughs. By allocating research budgets to wells the players can reduce the uncertainty of predictions about the future concentration of pumped groundwater from these wells. Uncertainty reduction results in an increase of the expected minimal reliable reaction time and thus in a reduction of potential losses due to breakthroughs.

The properties of the game experiment consider:

- a cost-uncertainty reduction function;
- an impact-reaction time function.

A number of wells each provided with:

- a concentration time series;
- a failure impact;
- a prediction difficulty class;
- a concentration variability class.

A number of players each provided with:

- a periodical budget to be distributed over the wells according to the players strategy;
- a set of confidence boundaries for each well that are adjusted through budget allocation.

The uncertainty reduction function is defined according to:

$$Ci(p,i,t,t') = \left(1 - Urf\,(p,i,t)\right) * Ci(p,i,t-1,t') \qquad (6\text{-}3)$$

with:

$$Urf = \frac{B}{Pd} \qquad (6\text{-}4)$$

subject to:

$$(t' - t) \le Mh \qquad (6\text{-}5)$$

Where:

Ci(p,i,t,t')	confidence interval (mg)
p	player index
i	well index
t	time step index
t'	prediction time index
Urf	uncertainty reduction factor (-)
B	fraction of total available budget per time step (-)
Pd	prediction difficulty class (-)
Mh	maximum time horizon of uncertainty reduction (T)

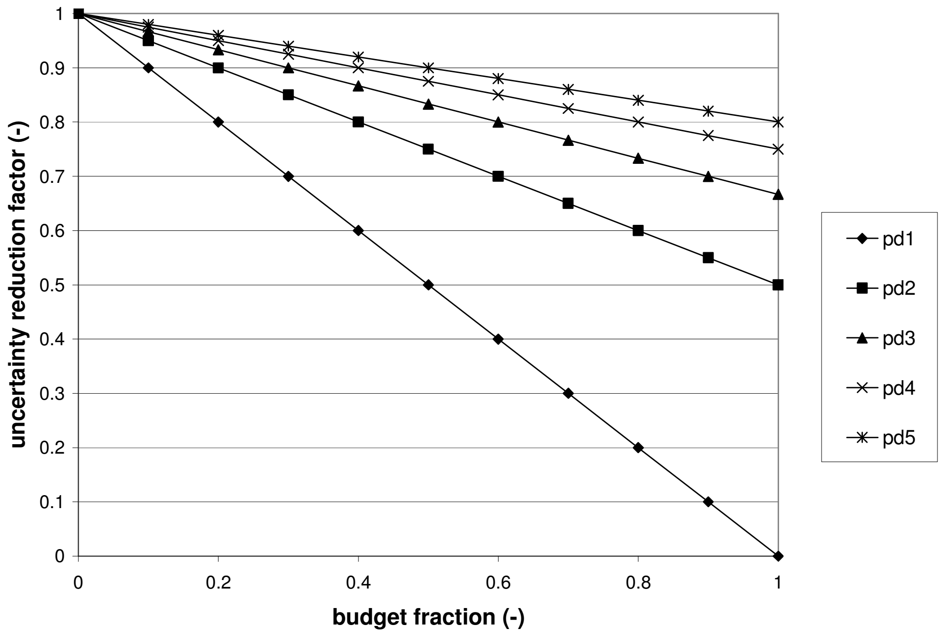

Figure 6-4 Uncertainty reduction functions for different classes of prediction difficulty

6.4.1 Loss function

The performance of the strategies is expressed through a penalty score. The score represents the sum of the manifest impact of wells that show a breakthrough. The score is summed at the first instant that the breakthrough was predicted reliably, which corresponds to the instant that the lower confidence boundary of a prediction intersects with the threshold concentration line. The manifest impact that corresponds to a reaction time of 1 time unit is summed in case a breakthrough is not predicted reliably before the instant of breakthrough.

$$\textit{Penalty score} = \sum_{i=1}^{n} \{ \mathrm{f}(\mathrm{Ib0[i]}, \mathrm{rt[i]}) \} \qquad (6\text{-}6)$$

subject to:

$$[Lbnd \cap Cthr \neq \phi] \; for \; (t' - t) \leq Mh \qquad (6\text{-}7)$$

Where:

n	number of wells
$Ib0$	potential failure impact[8]
$rt[i]$	true reaction time; period between present and moment that the threshold is exceeded
$Lbnd$	lower boundary of confidence intervals
$Cthr$	threshold concentration

The remaining game properties are described in the appendix.

6.4.2 Prioritisation strategies

The general objective of prioritisation of prediction studies is to minimize the negative impact of well failures due to contamination of groundwater. How this objective is achieved best is not clear on beforehand. A wide range of strategies were investigated. According to each strategy research budgets were allocated to the available wells every time step, proportional to the relative weight of a wells priority score to the sum of all priority scores. The total budget to be spent at each time step is 1. The strategies that were investigated are stated in the following section:

1. Priority score = random. To provide a reference performance for the experiment the first strategy denotes the random strategy. It represents a player/decision maker who assigns priorities for allocation of research budgets at random

2. Priority score = 1/(threshold-current concentration); represents an ‘ad hoc’ decision maker who allocates budgets according to the difference between the threshold concentration and the current concentration of pumped groundwater

[8] The units of failure impact depend on the impact category it refers to.

3. Priority score = potential failure impact; represents a decision maker who allocates budgets according to maximum potential failure impact (that varies among wells)

4. Priority score = 1/minimum reliable reaction time; represents a decision maker who applies priorities for maximizing the minimum reliable reaction time

5. Priority score = failure impact*anticipated change of reaction time /minimum reaction time; represents a decision maker who allocates budgets for maximizing the minimum reliable reaction time, weighted for potential failure impact and taking anticipated uncertainty reduction in account. The relations between allocated budget and reduced uncertainty of predictions as they are defined in the game are known by this player

6. Priority score = anticipated impact reduction / anticipated budget requirements; represents a decision maker who allocates budgets for early reliable prediction of breakthroughs. Maximization of early reliably predicted breakthroughs implies investments in research where chances are best to make the lower boundary of the confidence intervals intersect with the threshold concentration. The relations between allocated budget and reduced uncertainty of predictions as they are defined in the game are known by this player

The general features of the game represent the essential properties of the prioritisation problem, but it is poorly known what the ranges of values are that may occur for these features in the real world. The orders of magnitude of the model inputs are based upon estimates of real world data, but different drinking water systems possibly show large differences with respect to system properties. The performance of the various strategies was therefore investigated for large ranges of input values.

A total number of 16 game runs were carried out. Every run consisted of 1000 simulated well failures. An enumeration of the input values for the game runs is presented in Table 6-1 and in the appendix of this chapter.

6.5 Results

The results of the simulations show that strategy 6 performs best in all conditions that were investigated (Figure 6-5, Figure 6-6). The difference between the average relative performance of strategy 6 and other strategies was in all cases larger than 30%. The largest differences occurred when a type 2 impact-reaction time function was applied. The success of the strategy of player 6 considered both the percentage of breakthroughs that were predicted reliably (Figure 6-7) and the average reaction time at the instant of reliable prediction (Figure 6-8). Strategy 5 resulted in a minimal total risk level throughout a game run as compared to the other strategies, but this did not result in minimal losses.

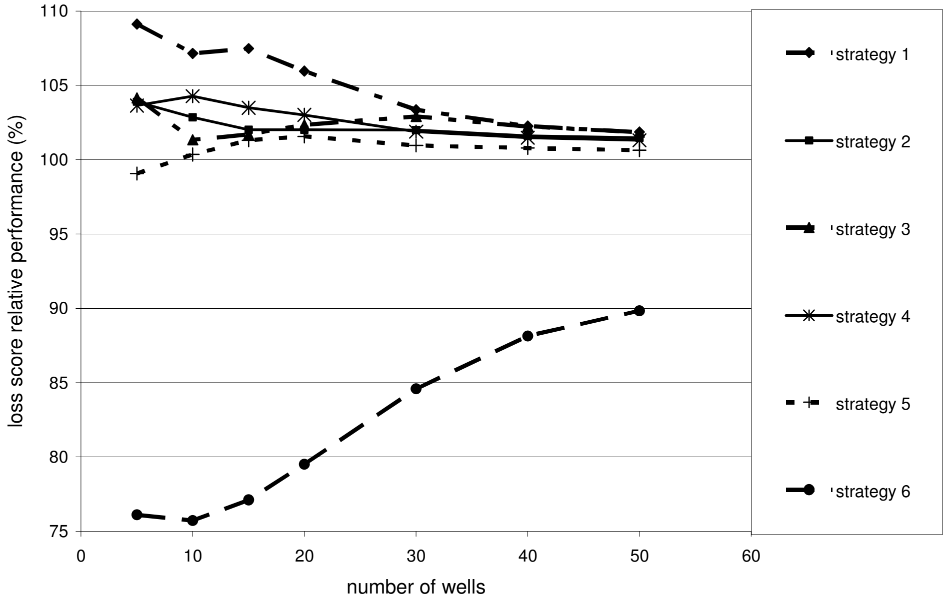

Figure 6-5 Average performance of players relative to average loss score per run (linear impact – reaction time function)

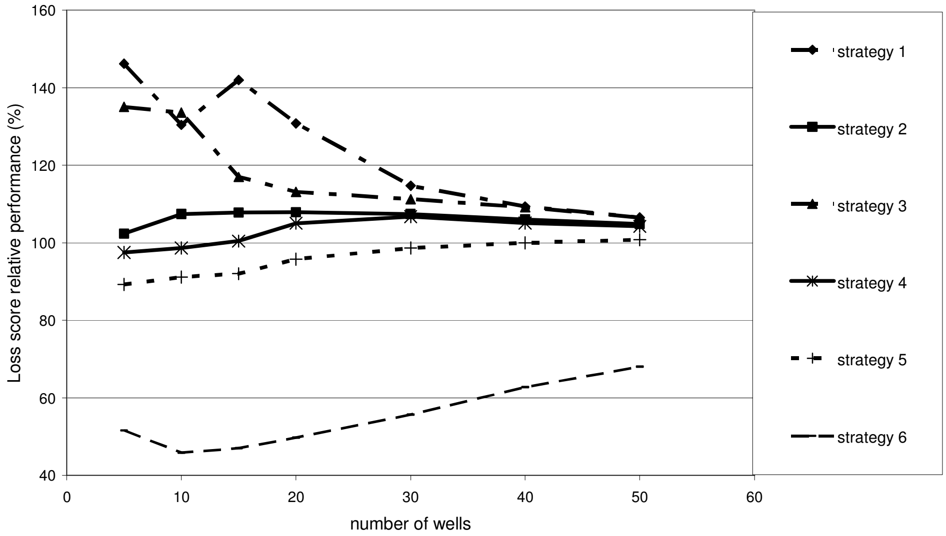

Figure 6-6 Average performance of players relative to average loss score per run (asymptotic impact – reaction time function)

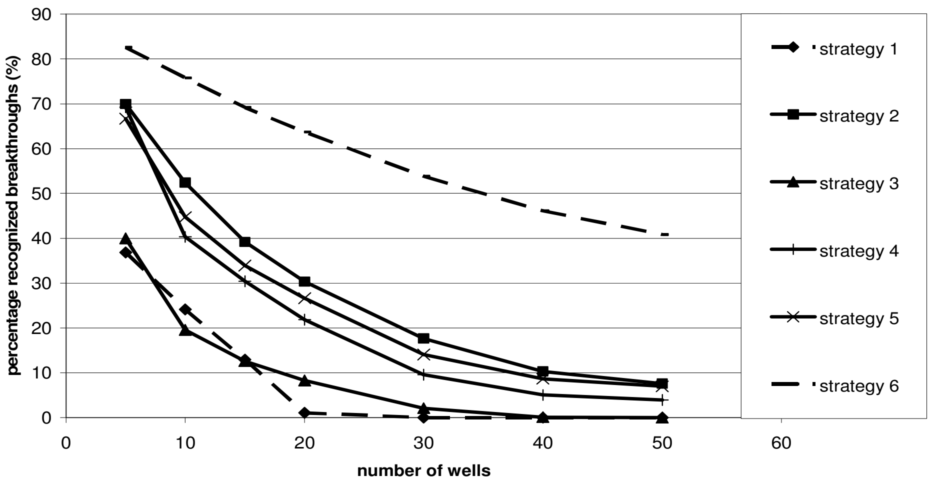

Figure 6-7 Percentage of recognized breakthroughs

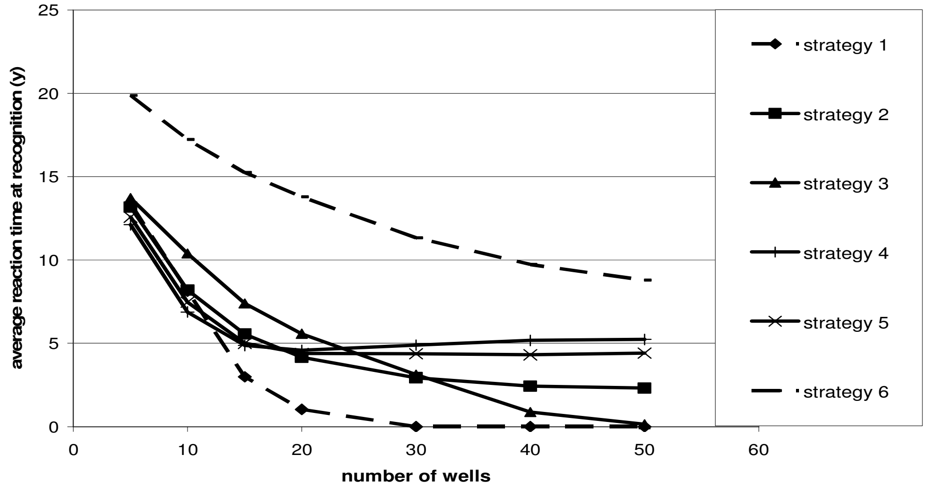

Figure 6-8 Average reaction time at instant of reliable prediction of breakthrough

6.6 Discussion and conclusions

Significant differences in performance were shown to exist between the various strategies that were investigated. It pays to formulate strategies carefully: strategies are suboptimal if the objective function of a strategy differs from the pay off function, or in this case, loss function of the well. Since the uncertainty of a prediction is important to the loss function, it should be taken in account in the strategies for prioritisation, as is shown in the relatively good results of strategies 4, 5 and 6, where uncertainty of predictions is taken in account. Confidence boundaries of time series of predicted groundwater concentration can be used to improve the effectiveness of prioritisation of prediction studies. The concept of "reliable reaction time" can contribute to a better integration of prediction studies and decision making.

The definition of appropriate strategies requires explicit and accurate loss functions. Maximization of minimum reliable reaction time (strategy 5) may seem intuitively a good strategy for minimization of total risk, but was less successful than strategy 6, which is better focused on the loss function. The uncertainty within risks often consists not only in the probability of events, but also in the consequences. Both risks and loss functions are therefore often hard to define accurately. It is therefore desirable that problems with stochastic loss or pay off functions are approached with a formalised, explicit and rational strategy. The use of numerical game experiments helped assessing the effectiveness of different strategies.

Application of strategy 6 in real situations requires both that prediction studies include quantitative uncertainty assessments and that the cost – uncertainty reduction functions of the wells can be assessed. Prioritisation in practice could improve if effective methods would exist for assessing on beforehand how much uncertainty reduction can be expected from additional studies. Uncertainty assessment techniques and related optimisation techniques of monitoring networks probably can be applied for this purpose as far as the use of additional observations is concerned (parameter uncertainty). However, the costs involved in assessing the budget requirements for a particular uncertainty reduction were not taken in account in the simulations.

Confidence intervals of time series of predicted groundwater concentration can be used to improve the effectiveness of research prioritisation of prediction studies. The sequential game approach was helpful in identifying the relevant properties for this prioritisation problem.

The validity of this experiment is based on the assumed analogy between the simulation game and the real world. In the real situation there is often more information available about relevant system variables such as land use and observed changes of concentrations in monitoring wells. Within the game model described in this paper this type of prior information was not considered. It is therefore not pretended that the model here presented is adequate for a specific case.

The use of numerical game experiments helped assessing the effectiveness of different strategies. Application of strategy 6 requires both that prediction studies include quantitative uncertainty assessments and that the cost – uncertainty reduction functions of the wells can be assessed. Prioritisation in practice could be improved if methods become available for assessing the expected uncertainty reduction due to additional studies.

APPENDIX to Chapter 6

Game procedure

In pseudo programming language the procedure of the game can be represented as:

Construct hypothetical time series of pseudo observed concentrations of pumped water for every well in the game.

Repeat

Determine the distribution of research budgets for all players this time step;

Calculate for each player for each well the new boundaries of the confidence intervals as a function of the allocated budget and the cost-uncertainty reduction function;

Increment time;

Update pay off scores;

If there was a breakthrough, or all players have a reliable prediction that a breakthrough will occur then define new properties of a replacing well;

Until the total numbers of breakthroughs corresponds to the number of plays that was chosen.

Definition of time series of concentration

At the start of the game a number of virtual wells are created and the failure impact of these wells is defined randomly. Fictitious time series of concentration are generated by using random functions. In the next section the procedure is described in pseudo programming language:

```
While no concentration exceeds a predefined threshold concentration
  do
  begin
          for all wells in the game;
          begin
                  if the trend duration has expired
                  begin
                          define a random trend
                          define a random trend duration
                  end
          Conc[t]=conc[t-1]+random noise+random trend;
          end
end
```

Concentration time series are defined according to:

$$C(t) = C(t-1) + Tr + N \qquad (6\text{-}8)$$

With:

$$-Tr\max{}^*Wd[i] \le Tr \le Tr\max{}^*Wd[i] \qquad (6\text{-}9)$$

$$-nf * Ctr * Wd[i] \le N \le nf * Ctr * Wd[i] \qquad (6\text{-}10)$$

$$C(t=1) = Random * Ctr * 0.5 \qquad (6\text{-}11)$$

$$Trd\min \le Trd \le Trd\max \qquad (6\text{-}12)$$

Where:

C(t)	concentration at time t (mg/l)
Tr	random trend (mg/l) (-Trmax<Tr<Trmax)
N	random noise (mg/l)
Tr_{max}	maximum trend (mg/l)

Wd[I]	concentration variability class of well [i] (-)
Nf	noise factor (-)
Ctr	threshold concentration (mg/l)
Trd	trend duration (time steps)
Trd_{min}	minimum duration of trend (time steps)
Trd_{max}	maximum duration of trend (time steps)

Constants

Game properties that were kept constant throughout the numerical experiments are listed in Table 6-1.

Table 6-1 Constants of the game experiment

Name	Value
Wd	1-5
Nf	0.05
Ctr	50
Trd	10 - 30
Trd_{min}	10
Trd_{max}	30
Maxhorizon	10 – 50
Pd	1-5

Table 6-2 Calculation scheme of input combinations

Run	Impact-reaction time function (-)	Maximum concentration variability class (-)	Maximum prediction horizon (time steps)	Maximum prediction difficulty class (-)
1	asymptotic	5	10	5
2	asymptotic	5	10	1
3	asymptotic	5	50	5
4	asymptotic	5	50	1
5	asymptotic	1	10	5
6	asymptotic	1	10	1
7	asymptotic	1	50	5
8	asymptotic	1	50	1
9	linear	5	10	5
10	Linear	5	10	1
11	Linear	5	50	5
12	Linear	5	50	1
13	Linear	1	10	5
14	Linear	1	10	1
15	Linear	1	50	5
16	Linear	1	50	1

PART C: SYNTHESIS

7. Analysis of heuristic optimisation techniques as a tool for decision support

7.1 Circumstantial validation

Heuristic techniques are powerful and flexible optimisation techniques. In Part B of this thesis it was demonstrated that highly complex problems can be solved by the genetic algorithms that were developed for this study. As a result of the heuristic, inductive character of these techniques, they can be used in a very flexible manner, without the need to reformulate existing models. It is expected that in the nearby future this property will lead to a strong increase in the application of heuristic techniques. The rapidly increasing computational power of desktop computers enforces this tendency.

However, it is not evident that the global optimum, or global Pareto fronts and surfaces in case of multiple objective optimisations, are indeed found when there is stagnation or convergence during the optimisation process. Therefore, validation of the results is needed. Direct validation is generally unfeasible because there are no global optimisation techniques at hand that are suitable to solve highly complex, nonlinear and interdependent optimisation problems. However, partial confirmation of the results is feasible. This *circumstantial validation,* as it is called in this thesis, acknowledges the fact that complete validation is in principle impossible, but that an acceptable level of validation for practical purposes can be provided in many cases.

Three different methods for circumstantial validation have been applied in this thesis and are enumerated in the next section.

- Analytical inspection of particularly the extreme ends of Pareto fronts.

- Application of the genetic algorithm to similar, slightly simplified problems that allow application of other techniques, whereas the degree of difficulty of the modified problem is similar, if not identical, from the viewpoint of optimisation by a genetic algorithm or other heuristic technique.

- Reformulation of the optimisation problem in such a way that one or several unique solutions along the Pareto front are known and therefore can be partially verified for these solutions, whereas the degree of difficulty of the optimisation problem remains the same, from the viewpoint of solving it by the genetic algorithm.

7.2 Valuation

Valuation of environmental and other non-economic objective categories is always required in decisions and becomes explicit by this approach. Effectiveness of decision making is hampered by the absence of an integrated and coherent valuation system of environmental values such as energy, CO2 emission, waste production, use of materials, use of environmental capital and biodiversity. A formalised approach to optimisation problems is not only desirable if all impacts can be quantified easily. Many real world optimisation problems contain 'soft' elements that cannot be quantified without discussion. Yet, even in these cases the formulation of a management problem as an optimisation problem has clear advantages. Objectives are formulated explicitly, which results in an improved discussion between experts and/or stakeholders, as differences between opinions become more clearly.

7.3 Benefits

Economic and environmental benefits of the application of DSS and optimisation techniques can be significant. Optimum solutions of complex problems are rarely trivial and different solutions result in considerable differences of impacts. Gradual progress towards generally accepted formal quantitative relations becomes possible when calculated optima can be compared with expert judgments. Generally excepted impact models and valuations can be applied consistently throughout the optimisation process, which is unfeasible in the case of pure expert judgments.

7.4 Polder or poker?

The use of decision support systems, with or without optimisation modules, is not in all situations effective. It is relatively simple to apply decision support systems if there is only one decision maker, or a group of decision makers that support the decision support system. Decision support systems can facilitate the decision process substantially if the negotiation attitude is cooperative. If stakeholders/decision makers with different interests are to use a decision support system jointly, transparency might result in polarisation. DSS are more suitable for a "polder model" than for a "poker model" where players tend to "keep their cards against their breasts". Transparency is required when decision support systems are meant to be used by stakeholders with conflicting interests, as it is unfeasible to hide information within a decision support system and maintain acceptance by all users of the decision support system. An open participatory decision process is then required. A decision process where a single institute or person has sufficient authority to determine the rules of the negotiation process may enforce acceptance of a particular decision support system. The use of decision support systems is unsuitable if the stakeholders have conflicting interests, do not agree on transparency and do not support the results of the decision support system. A decision support system as a common tool is not useful in those conditions as long as the actors disagree on the definitions ad assumptions of a DSS.

7.5 Continuing lines

Decision support systems and heuristic optimisation techniques with results presented in Pareto fronts are not only instruments to find "the optimal solution" but also offer a way to better understand the problem by analysing results of different scenarios. They form to my opinion a continuation of a line that started quite some time ago, when calculations were introduced to support decisions. At first, only one, or at most a few scenarios were calculated. Later, impacts of many scenarios, eventually presented as Pareto fronts were quantified by application of a simulation-optimisation approach. Presently, many complex optimisation problems with a spatial dimension can be solved effectively by heuristic techniques, but optimisation problems with both spatial and temporal dimensions seem still a bridge too far. Perhaps the line can be continued here by joining heuristic techniques with game modelling.

There is also another line of which this thesis aims to be a continuation: quite a long line, that started in the early days of man's existence, when people probably used instinct as their principal cerebral tool, Relatively recently, people developed optimisation techniques for single objective problems. They applied weighting factors for multiple objective problems, thus reducing them to single objective problems. However, using of weighting factors is not always a satisfactory approach, as has been discussed in this thesis. Using Pareto fronts is an interesting alternative; it enables determination of weighting factors aposteriori. Many investigations have showed that intuitive and implicit decisions are often irrational and inconsistent [e.g. Kahneman et al. 1982]. I am the opinion that the complexity of many present-day optimisation problems requires a formalised, rational and explicit approach. If optimisation problems are suitable for solving by genetic algorithms, the continuation of this long line consist of 6 basic steps:

1. Identify the objective categories and criteria.
2. Determine the impact models and the indicators.
3. Code a problem into a genetic algorithm.
4. Optimise and determine the Pareto fronts.
5. Validate results by circumstantial validation.
6. Choose the appropriate solution.

As it is possible to determine the weighting factors aposteriori, it is possible to verify whether the selected solution is consistent with decisions taken previously. At present, such a comparison is rarely made. It is very well possible that, for instance, the ratio of actual weighting factors for environmental objectives as compared to economic objectives vary widely in recent governmental projects. If a government or an enterprise wants to spend a certain available budget consistently with respect to the values it attributes to particular objectives, it is necessary that the weighting factors are known. The weighting factors related to investments in goods can be determined by means of a life cycle assessment (LCA) or by other techniques. These assessments should be carried out with respect through all relevant objectives. Thus, it can be achieved that investments for improvement of ecological quality, or any other relevant objective, are done there where a maximum yield can be expected, based on explicit considerations. Over the last decades quite a number of authors have proposed new valuation methods for non-economic goods and services [e.g. Costanza et al., 1997].

There are certainly some setbacks to such an approach to multiple objective optimisation problems. It requires a degree of transparency that might result in a polarisation between different groups of stakeholders involved. In some cases a less transparent, more implicit approach can avoid this polarisation. Furthermore, it is in many cases quite difficult to handle uncertainty and avoid bias in the quantification of impacts of decisions. However, a generally accepted pricing system for non-economic costs and benefits within a sector, a region, a state or even larger organisation levels could facilitate negotiation processes and improve the consistency and effectiveness of actions in public and private domains. Finally, from the viewpoint of the historical development of mankind, continuing the lines from intuitive, implicit societies to open and rational societies, requires a gradual shift to a more explicit approach for multiple objective optimisation problems. Genetic algorithms can play a useful role in that evolution.

8. References

- Aerts, J. (2002), Spatial decision support for resource allocation. PhD thesis, University of Amsterdam, 2002.
- Aguado E. (1974), I. Remson, M. F. Pikul and W. A. Thomas, Optimum pumping to prevent dewatering, J. Hydraulic Div. Am. Soc. Civ. Eng., 100, 860-877.
- Anonymous (1996), Emissieplafonds voor nitrogenk en maximale stikstofgiften per gemeente in Nederland. Een eerste orde benadering. RIVM-rapport nr. 722108019.
- Anonymus (2005), Decision support systems. http://www.informationbuilders.com/decision-support-systems-dss.html.
- Azapagic A. and R. Clift (1999), Life cycle assessment and multi-objective optimisation, Journal of Cleaner Production 7 135-143.
- Azencott R. (1992), ed., Simulated Annealing: Parallelization Techniques, John Wiley & Sons.
- Beasley D., D. R. Bull and R. R. Martin (1993), An overview of genetic algorithms, University Computing.
- Bleeker A., H.W. ter Maat and P.S.C. Heuberger (1999), Bouwstenen emissiegericht nitrogenkbeleid in Drenthe; een regionale beschrijving van een optimale emissieverdeling op basis van de draagkracht van de natuur in de EHS. RIVM-rapport nr. . 722108024.
- Braat L.C., A. Van Amstel, E. Nieuwhof, J. Runhaar and J. B.Vos (1987), Verdroging in Nederland, Publikatiereeks Milieubeheer, Ministerie van VROM, Den Haag, The Netherlands.
- Brookes C.J. (2001), A genetic algoritm for designing optimal patch configurations in GIS. int. j. geographical information science, 2001, vol. 15, no. 6, 539- 559
- Buckles B.P., and F.E. Petry (1992), Genetic algorithms, Institute of Electrical and Electronics Engineers, Inc.
- Busacker R.G., and P. J. Gowen (1961), A procedure for determining a family of minimum-cost flow patterns. Operations Research Office Technical Report 15, J. Hopkins University, Baltimore.

- Carrera J., E. Usunoff and F. Szidarovszky (1984), A method for optimal observation network design for groundwater management. J. Hydrol., 73, 147-163.
- Chamberlain, E. H. (1948), An Experimental Imperfect Market. Journal of Political Economy 56: 95-108.
- Cieniawski S. E., J. W. Eheart and S. Ranjithan (1995), Using genetic algorithms to solve a multiobjective groundwater monitoring problem, Water Resources Research 31(2) 399-409.
- Costanza R., d'Arge, R., de Groot, R., Farber, S., Grasso, M.,Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P. and M. van den Belt (1997), The value of the world's ecosystem services and natural capital. Nature Vol. 387.
- Dagan G. (1986), Statistical theory of groundwater flow and transport: Pore to laboratory, laboratory to formation and formation to regional scale. Water Resources Research 22 No 9.
- Dantzig G.B. (1963), Linear Programming and Extensions, Princeton University Press, Princeton, NJ.
- Deb and Kalyanmony (1999), Multi-objective genetic algorithms: Problem difficulties and construction of test problems, Evolutionary Computation Fall99 Vol. 7.
- Delhomme J.P. (1978), Kriging in the Hydrosciences. Adv. Water Resour. 1(5), 251-266.
- Dettinger M.D. and J. L. Wilson (1981), First-order analysis of uncertainty in numerical models of groundwater flow, 1, Mathematical development: Water Resources Research, 17, 149-161.
- Dixon W. and B. Chiswell (1996), Review of aquatic monitoring program design. Wat. Res. 30 (9) 1935-1948.
- Doherty, J. (2002), Manual for PEST, Fifth Edition,Watermark Numerical Computing, Australia.
- El Magnouni, S. and W. Treichel (1994), A multicriterion approach to groundwater management. Water Resour. Res. 30(6) 1881-1895.
- Erisman, J.W., M.H.J. Klein, H.M. Beije, J. Latour, A. Bleeker, N.J.P. Hoogervorst, P.M. van Egmond, R. Reiling, A.H. Bakema, M.J.A. van Adrichem, P.S.C. Heuberger (1997), Perspectieven voor de

oplossing van de nitrogenkproblematiek in de Nederlandse natuur. RIVM-rapport nr. 722108021.

- Erisman, J.W., Bleeker, A., van de Velde, R., Bakema, A.H., van Grinsven, J.J.M., Wiertz, J, en N.J.P. Hoogervorst (1996), Indicatieve studie naar de nitrogenkproblematiek in het ROM-gebied Z.O.-Friesland. RIVM-rapport nr. 722108011.
- Eshelman L.J., R. Caruna and J. D. Schaffer (1989), Biases in the crossover landscape. In J.D. Schaffer, editor, Proceedings of the Third International Conference on Genetic Algorithms, 10 - 19, Morgan-Kaufmann.
- Finkel A.M. and J.S. Evans (1987), Evaluating the benefits of uncertainty reduction in environmental health risk management. Journal of the Air Pollution Control Association, 37(10):1164-1171.
- Freeze R.A., R. Zapf-Gilje, J. Massman, L. Cahn, L. Smith and B. James (1987), Advances in the assessment of data worth for engineering decision analysis in groundwater contamination problems. In: Groundwater flow and quality modelling. E. Custodio, A. Gurgui and J.P. Lobo Ferreira (1987), Eds. Boston: D. Reidel.
- Geer, van F.C (1987), Applications of Kalman filtering in the analysis and design of groundwater monitoring networks. Thesis. Delft University Press.
- Gelhar, L. W. (1986), Stochastic Subsurface hydrology; from theory and applications. Water Resources Research Vol. 22 No 9.
- Gelhar, L. W. (1993), Stochastic Subsurface hydrology;.Prentice Hall.
- Goldberg, D.E. (1989), Sizing populations for serial and parallel genetic algorithms, Proceedings of the Third International Conference on Genetic Algorithms.
- Gorelick S. M., C. I. Voss, P. E. Gill, W. Murray, M. A. Saunders and M.H. Wright (1984), Aquifer reclamation design: The use of contaminant transport simulation combined with nonlinear programming, Water Resour. Res., 20(4), 415-427.
- Gorelick S. M., I. Remson and R. W. Cottle (1979), Management model of a groundwater system with a transient pollutant source, Water Resour. Res. 15(5) 1243-1249.

- Grefenstette J.J. (1990), Genetic algorithms and their applications. In A. Kent and J.G. Williams, eds., Encyclopaedia of Computer Science and Technology, 139 - 152, Marcel Dekker.
- Griffith, R.F. (1981), A Background in Risks., in: Farmer F.R. and Warner, F. eds. Living with uncertainty: Risks in the Energy scene, London, Oyez – IBC.
- Grootjans, A.P. (1985), Changes of groundwater regime in wet meadows. Thesis Groningen University, The Netherlands.
- Haberlandt U.,V. Krysanova, A. Bardossy (2002), Assessment of nitrogen leaching from arable land in largeriver basins. Part II: regionalisation using fuzzy rule based modelling. Ecological Modelling 150 (2002), 277-294
- Harbaugh A.W., Edward R. Banta, Hill M.C. and Michael G. McDonald (2000), Modflow 2000 The U.S. Geological Survey modular groundwater model; User guide to modularization concepts and the groundwater flow process, OR 00-92, U.S. Geological Survey.
- Heuberger, P.S.C., A.H. Bakema, M.J.A. van Adrichem (1997), Optimalisatie van ruimtelijke emissieverdelingen. RIVM-rapport nr. 408143003.
- Hill M. C. (1992), A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional groundwater flow model using nonlinear regression. OR 91-484, U.S. Geological Survey.
- Hill M.C., Edward R. Banta, Arlen W. Harbaugh and Evan R. Anderman (2000), Modflow 2000 The U.S. Geological Survey modular groundwater model; User guide to the observation, sensitivity and parameter estimation processes and three post processing programs; OR 00-184, U.S. Geological Survey.
- Hill, M.C. (1998), Methods and Guidelines for Effective Model Calibration, U.S. Geological Survey Water Resources Investigations Report 98-4005, U.S. Geological Survey, Denver, Colorado.
- Hillier F.S. and G.J. Lieberman (1980), Introduction to Operations Research. 3rd ed. Holden-Day Cal.
- Holland, J.H. (1975), Adaptation in natural and artificial systems MIT Press.

- Horton, B., (1996), A method of using a genetic algorithm to examine the optimum structure of the Australian sheep breeding industry: open-breeding systems, MOET and AI. Australian Journal of Experimental Agriculture 36, 249–258.
- Ingber, L. and B. Rosen (1992), Genetic algorithms and very fast simulated reannealing: a comparison . Math Comp. Mod. 16 , 87-100, 1992.
- Ingber, L. (1989), Very Fast Simulated Re-Annealing. Math Comp. Mod. 12 , 967-973, 1989.
- Jones, R.M. (1998), An analysis of computer-supported co-operative work systems to support decision-making in regional planning, Computers, Environment and Urban Systems, Vol. 22 No 4, 335-350.
- Kahneman D. and A.Tversky (1982), Judgement under uncertainty: heuristics and biases. eds: D. Kahneman, P.Slovic and A.Tversky. Cambridge Univ. Press. pp48.
- Karmarkar N. (1984), A new polynomial time algorithm for linear programming. Combinatorica, Vol.4, pp.373-395
- Keeney and Raiffa (1993), Decisions with multiple objectives, Preferences and Value trade-offs. Cambridge Univ. Press.
- Kirkpatrick, S., C.D. Gelatt, Jr. and M.P. Vechi (1983), Optimisation by simulaterd annaealing. Science 220 (4598), 671-680, 1983.
- Kuo C.H., A.N. Michel, and W.G. Gray (1992), Design of optimal pump-and-treat strategies for contaminated groundwater remediation using the simulated annealing algorithm. Advances in Water Resources AWREDI 15, No. 2: 95.105.
- Last F. T. and R. Watling (1991), Acid deposition, its nature and impacys. Proc. Roy. Soc. Edinburgh.
- Lazo, J. and W. D. Schulze (1992), Can contingent valuation measure nonuse values?, American Journal of Agricultural Economics Vol. 74 Issue 5, p1126-1133.
- Leeuw, de F.A.A.M. and J.A. van Jaarsveld (1992), Determination of source-receptor relations for acidifying components (in Dutch) RIVM-report 723001009.

- Lenstra J.K., A.H.G. Rinnooy Kan (1991), A. Schrijver (eds.), History of Mathematical Programming: a collection of personal reminiscences, North-Holland, Amsterdam.
- Loonen W., P.S.C.A. Heuberger, H. Bakema and P.P. Schot, Application of a genetic algorithm to minimize agricultural nitrogen deposition in nature reserves. Agricultural systems (accepted)
- Makinde-Odusola, B. A. and M. A. Marino (1989), Optimal control of groundwater by the feed-back method of control, Water Resour. Res. 25(6),1341-1352.
- Massman J. and Freeze R. A. (1987), Groundwater contamination from waste management sites: the interaction between risk-based engineering design and regulatory policy, 1, methodology. Water Resour. Res. 23(2): 351-367.
- Massman J. and Freeze R. A. (1987), Groundwater contamination from waste management sites: the interaction between risk-based engineering design and regulatory policy, 2, results. Water Resour. Res. 23(2): 368-380.
- Mayer, D.G., Schoorl, D., Butler, D.G., Kelly, A.M. (1991), Efficiency and fractal behaviour of optimisation methods on multiple-optima surfaces. Agricultural Systems 36, 315–328.
- Mayer, D.G., Belward, J.A., Burrage, K. (1996), Use of advanced techniques to optimize a multidimensional dairy model. Agricultural Systems 50, 239–253.
- Mayer, D.G., Belward, J.A., Burrage, K. (1998). Optimizing simulation models of agricultural systems. Annals of Operations Research 82, 219–231.
- Mayer D.G., J.A.Belward, K.Burrage (2001), Robust parameter settings of evolutionary algorithms for the optimisation of agricultural systems models. Agricultural Systems 69 (2001) 199–213
- McDonald M.G. and Harbaugh A.W. (1984), A modular three-dimensional finite difference groundwater flow model, U.S. Geological Survey
- McKinney, Daene C. and D.P. Loucks (1992), Network design for predicting groundwater contamination. Water Resour. Res. 28(1) 133-147.

- McKinney Deane C. and Min-Der Lin (1994), Genetic algorithm solution of groundwater management models, Water Resour. Res. 30(6).1897-1906.
- McKinney Deane C. and Min-Der Lin (1992), Design methodology for efficient aquifer remediation using pump and treat systems. In: Mathematical Modelling in Water Resources, Ed.T. Russel et al., pp 695-702, Elsevier Science, New York.
- Metropolis, N., A. Rosenbluth, W. Rosenbluth and A.H. Teller, Equation of State Calculations by Fast Computing Machines, J. Chem. Phys. 21 (6), 1087-1092, 1953.
- Michalewicz Z. (1996), Genetic algorithms+data structures = evolution programs, 3rd edition Springer.
- Mintzberg, H. (1976), Planning on the left side and managing on the right side, Harvard business review, Vol 54, No. 4, pp. 49-58.
- Molz F.J. and L. C. Bell (1977), Head gradient control in aquifers used for fluid storage, Water Resour Res. 13(4), 795-798.
- Neumann J. von and O. Morgenstern (1944), Theory of Games and Economic Behaviour. Princeton press.
- Nijkamp P. (1979), Multidimensional spatial data and decision analysis. Chichester Wiley.
- Papadimitriou C. H. and K. Steiglitz (1985), Combinatorial optimisation - Algorithms and complexity.
- Partidário M. R. (1996), Strategic environmental assessment: Key issues emerging from recent practice Environmental Impact Assessment Review 16:31-55.
- Pareto V. (1906), Manual of political economy. Kelley Pubs; Reprint edition (1969)
- Parsons, D.J. (1998), Optimising silage harvesting plans in a grass and grazing simulation using the revised simplex method and a genetic algorithm. Agricultural Systems 56, 29–44.
- Pincus, M. (1970), A Monte Carlo method for the approximate solution of certain types of constrained optimisation problems, Oper. Res. 18, 1225-1228, 1970.

- Power, D.J. (2003), A Brief History of Decision Support Systems. DSSResources.COM, World Wide Web, http://DSSResources.COM/history/dsshistory.html, version 2.8.
- Purao S., K. J. Jain and D. L. Nazareth (1999), Supporting decision making in combinatorially explosive multicriteria situations, Decision Support Systems 26 225-247.
- Rajabi S., K. W. Hipel and D. M. Kilgour (1996), Water supply planning under interdependence of actions, Theory and application, Water Resour. Res. 35(7) .2225-2235.
- Reichard R, C. Cranor, R. Raucher and G. Zapponi (1990), Groundwater contamination risk assessment. A guide to understanding and managing uncertainties. IAHS publication N0. 196.
- Reichard E.G. and J.S. Evans (1989), Assessing the value of hydrogeologic information for risk-based remedial action decisions. Water Resour. Res. 25(7): 1451-1460.
- Runhaar, J., C. R. van Gool and C. L. G. Groen (1996), Impact of hydrological changes on nature conservation areas in The Netherlands, Biological Conservation 76 269-276.
- Schaffers H. (1984), Computerondersteuning en planning, Mogelijkheden en beperkingen bij regionale watervoorziening, thesis Technical University Twente (in Dutch), Enschede, Netherlands.
- Schot, P.P., Willem Loonen and Kees Vink (2005), U vraagt en wij draaien; Ruimtelijk optimaal verdelen van kwaliteitswensen. Landschap 22-1 pp. 55-62
- Seppelt R. (2000), Regionalised optimum control problems for agroecosystem management. In: Ecological Modelling. 131(2-3), 121-132
- Seppelt R. and A.Voinov (2002), Optimisation methodology for land use patterns using spatially explicit landscape models. Ecological Modelling 151 pp. 125–142
- Sprague, R. H. Jr. and H. J. Watson (1979), Bit by Bit: Toward Decision Support Systems, California Management Review, vol. XXII, no. 1, pp. 60-68

- Stanners, D. and P. Bourdeau (eds.)(1995). Europe's environment; the Dobris assessment. European Environmental Agency, Copenhagen, Denmark.
- Stear, E.B. (1987). Control paradigms and self-organization in living systems. In F. E. Yates (Ed.). Self-organizing systems: The emergence of order. New York: Plenum Press.
- Stoner, R. F., D. M. Milne, and P. J. Lund (1979), Economic design of wells, Q. Journal of Engineering Geology.
- Stoner, R. F. (1981), Economic aspects and policy issues in groundwater development, World Bank Staff Working Paper 469.
- Szaniawski K. (1998), On science, inference, information and decision making, selected essays in the philosophy of science. Eds. Adam Chmielewski and Jan Wolenski, Kluwer.
- Todd, D. K. (1980), Groundwater hydrology, John Wiley and Sons, New York.
- Van Dam, J.D. (ed), P.S.C. Heuberger. J.M.M. Aben, W.A.J. van Pul (2001). Effecten van verplaatsing van agrarische nitrogenkemissies. Verkennning op provinciaal niveau. RIVM-rapport nr. 725501003, 2001.
- Vanderbei R.J.(1996), Linear Programming: Foundations and Extensions, Kluwer Academic Publishers, Boston, 1996.
- Vanderbei R.J. (1996), Linear Programming: Foundations and Extensions, Kluwer Academic Publishers, Boston.
- Vink C., Heuberger P.,Bakema A.(2002), Ruimtelijke Optimalisatiemethoden, Rivm CIM, Intern CIM rapport nr. M004/2002
- Vink, K. and P.P. Schot (2002), Multiple-objective optimisation of drinking water production strategies using a genetic algorithm, Water Resources Research, Vol. 38, No. 9, doi:10.1029/2000WR000034, 2002
- Wald, A. (1950), Statistical Decision functions, Wiley & sons New York.
- White, J. P. (1985), A multiple objective interactive Lagrangean relaxation approach, Eur. J. Oper. Res.,19(1) 82 90.

- Whitehead, J.C. and G. C.Blomquist (1991), Measuring contingent values for wetlands: effects of information about related environmental goods, Water Resour. Res. 27(10) 2523-2531.
- Willis R. L. and W. G. Yeh (1987), Groundwater systems planning and management Prentice-Hall, Englewood Cliffs, N. J.
- Witte, J.P.M., C. L. G. Groen, R. Van der Meijden and J. G. Nienhuis (1993), DEMNAT: a national model for the effects of water management on the vegetation. in: Hooghart, J. C., Posthumus, C.W.S. (Eds.), The Use of Hydro-Ecological Models in The Netherlands, TNO Committee on Hydrological Research, Proceedings and Information no. 47, Delft, . 31-51.
- Wright S.J. (1997), Primal-Dual Interior-Point Methods, Society for Industrial and Applied Mathematics, Philadelphia
- Zander P., H. Kächele (1999). Modelling multiple objectives of land use for sustainable development. Agricultural Systems 59 311-325
- Zheng, C. (1997), ModGA: A Genetic Algorithm Based Groundwater Flow and Transport Optimisation Model MODFLOW and MT3D, Report to DuPont Chemical, 95 pp.
- Zheng, C. (1997), ModGA_P: Parameter Estimation Using Genetic Algorithms, Report to DuPont Chemical, 35 pp.

Abbreviations

GA	Genetic Algorithm
SA	Simulated Annealing
GP	Goal Programming
NN	Neural Network
LP	Linear Programming
NLP	Nonlinear Programming
DSS	Decision Support System
SDSS	Spatial Decision Support System
LCA	Life Cycle Assessment
MOO	Multiple Objective Optimisation
QP	Quadratic Programming
MC	Monte Carlo technique
DP	Dynamic Programming

Summary

Management of hydrological and other natural resources is becoming increasingly complex because of their increasing scarcity, the increasing number of actors and objectives involved, and because of the increasing rate of change of technological, environmental and economic conditions.

The risks of choosing suboptimal solutions have become more pronounced because of this increasing complexity.

Yet, at the same time computers have become more powerful and computer based heuristic optimisation techniques have become more suitable for complex, nonlinear and interdependent optimisation problems.

Consequently, the hypothesis of this thesis is that application of heuristic optimisation techniques to complex spatial environmental problems with multiple objectives can improve the identification of (near) Pareto-efficient solutions and thus contribute to more effective decision making.

Evaluation of this hypothesis within the framework of this thesis consisted of the application heuristic optimisation techniques, to a suite of case studies.

The study consisted of four case studies, of which three comprised the application of genetic algorithms. The fourth case consisted of a sequential game experiment. It was investigated whether genetic algorithms can be applied successfully to a suite of complex optimisation problems in the environmental and hydrologic fields. The pro's and con's of the use of these techniques and the conditions for effective application were studied, particularly focussed on handling multiple objectives and validation of results.

In case study 1 the calibration of a groundwater model was approached as a multiple objective optimisation problem. Regional drinking water production was formulated as a multiple objective optimisation problem in case study 2 and case study 3 consisted of various optimisation problems concerning allocation of agriculture and nature landuse types. In the fourth

study, the optimal strategy for the prioritisation of groundwater quality prediction studies was searched with a set of game experiments that enabled simulation and evaluation of various strategies.

In case studies 1 -3, the *spatial* dimension is the major dimension that contributes to the complexity, but in the fourth case it is rather the temporal dimension.

The results of the case studies confirm the hypothesis of this dissertation that the application of heuristic techniques to complex optimisation problems in spatial planning and resource management enables better, more efficient decision making. The genetic algorithms that were built specifically for the case studies provided a powerful, stable and flexible optimisation technique. Pareto-optimality and uniqueness of solutions proved to be effective, unbiased fitness criteria for identifying trade-off curves. In these three case studies a certain degree of tuning of the genetic algorithm was necessary for the more complex versions of the problems. Consequently, it is considered essential to validate the results of heuristic optimisation techniques. Although complete validation is principally not possible, several ways of *circumstantial validation* could be achieved. Three different methods of circumstantial validation were applied:

1. Formulation of 'dummy' problems that are similar to the real optimisation problem, but constructed in such a way that one or more properties of solutions along the Pareto front of the 'dummy' optimisation problem are known; the results of the optimisation of the real problem can thus be partially validated.

2. Analytical inspection of particularly the extreme ends of Pareto fronts.

3. Application of the genetic algorithm to similar, but simplified problems that allow application of other techniques, such as linear or nonlinear programming, while maintaining a similar degree of difficulty of the modified problem from the viewpoint of optimisation by a heuristic technique.

Like in case study 1, case study 4 shows how uncertainty of model characteristics can be incorporated in the optimisation approach. An explicit formulation of the prioritisation problem enabled comparison of concepts that were conceived by deduction with the numerical results of the game model. Thus, “expert judgment” and intuition, that presently are in practice the major bases for decision making on this prioritisation problem, could be strengthened by a more scientific approach.

Decision making with respect to real-world problems with conflicting objectives requires valuation of incommensurable objectives. Valuation of non-economic impact categories requires inevitably communication, negotiation and sometimes even confrontation between decision makers and stakeholders. Explicit valuation on a basis of ‘willingness to pay’ of individuals offers good possibilities to differentiate between various kinds of environmental capital and services, particularly if the valuers are well informed. The definition of formal and explicit objectives, valuations and quantification methods with respect to environmental and other non-economic issues is therefore desirable. Optimisation techniques and decision support systems have the potential to improve the transparency, efficiency and consistency of decisions. However, experiences in practice showed that the use of decision support systems is not in all situations effective. Having confidence in the results of such a system is the principal condition if only one stakeholder, or a group stakeholders with the same interests uses a decision support system. If stakeholders with different interests are to use a decision support system jointly, then the negotiation attitude needs to be cooperative too. A decision support system is suitable for a “polder model” but less suitable for a “poker model” where players tend to “hide their cards against their breasts”.

Samenvatting

Management van hydrologische en andere natuurlijke bronnen wordt steeds complexer door toenemende schaarste van bronnen, het toenemende aantal betrokken actoren, het toenemende aantal doelstellingen en de toenemende snelheid waarmee technologische, milieuhygiënische en economische omstandigheden veranderen.

Het risico dat suboptimale keuzen worden gemaakt is groter geworden als gevolg van deze toegenomen complexiteit.

Tegelijkertijd zijn computers echter krachtiger geworden en is de toepassing van digitale heuristische optimalisatietechnieken daardoor effectiever geworden bij het oplossen van complexe, nonlineaire en interdependente optimalisatieproblemen.

De hypothese van dit proefschrift is dat gebruik van heuristische optimalisatietechnieken bij complexe ruimtelijke, milieuhygiënische problemen met meerdere doelstellingen de identificatie van Pareto-efficiënte oplossingen kan faciliteren en daardoor kan bijdragen aan effectievere besluitvorming.

Toetsing van deze hypothese in dit proefschrift bestond uit de toepassing van heuristische optimalisatietechnieken, op vier case-studies.

In dit proefschrift zijn vier case-studies beschreven, waarvan er drie betrekking hebben op het gebruik van genetische algoritmen. De vierde case study betreft de analyse van een optimalisatieprobleem met een sequentieel spel-experiment. Er is onderzocht of deze heuristische optimalisatietechnieken met succes kunnen worden toegepast in deze case-studies op hydrologisch en milieukundig gebied. De voor- en nadelen van het gebruik van deze technieken en de voorwaarden voor effectieve toepassing zijn onderzocht. Daarbij is met nadruk het omgaan met meerdere doelstellingen en de validatie van de resultaten onderzocht.

In case study 1 is de calibratie van een numeriek grondwatermodel benaderd als een optimalisatieprobleem met meerdere doelstellingen. Drinkwatervoorziening op regionale schaal is geformuleerd als een optimalisatieprobleem met meerdere doelstellingen in case study 2. Case study 3 bestaat uit een analyse van enkele optimalisatieproblemen die betrekking hebben op de toedeling van de landgebruikstypen natuur en landbouw. In de vierde case study is onderzocht welke strategie voor het prioriteren van ruwwaterkwaliteitvoorspellingstudies optimaal is. Dit is onderzocht met behulp van een serie numerieke spel-experimenten.

In de case studies 1 t/m 3 is de ruimtelijke dimensie de belangrijkste veroorzaker van de complexiteit van de optimalisatieproblemen, terwijl in de vierde case study eerder de temporele dimensie de oorzaak van de complexiteit is.

De resultaten van de case studies 1 t/m 3 bevestigen de hypothese van deze dissertatie; de toepassing van heuristische optimalisatietechnieken op complexe optimalisatieproblemen in ruimtelijke planning en het beheer van natuurlijke hulpbronnen maken een betere, efficiëntere besluitvporming mogelijk. De genetische algoritmen, die speciaal voor deze case studies zijn gemaakt, zijn een krachtige, stabiele en flexibele optimalisatietechniek gebleken. Er is aangetoond dat Pareto-optimaliteit en uniciteit van oplossingen effectieve, objectieve criteria voor fitheid zijn bij de identificatie van optimale oplossingen. In deze case studies is enige mate van aanpassing van de genetische algoritmen aan de specifieke aard van de optimalisatieproblemen nodig gebleken, m.n. voor de meest complexe varianten. Op grond van deze bevinding kan geconcludeerd worden dat het essentieel is dat met heuristische technieken verkregen resultaten zo goed en volledig mogelijk worden gevalideerd. Hoewel volledige validatie principeel onmogelijk is, zijn er verschillende manieren van 'onvolledige validatie' ontwikkeld en toegepast. Drie verschillende vormen van deze onvolledige validatie zijn toegepast:

1. Formulering van 'variantproblemen' die lijken op het eigenlijke optimalisatieprobleem, maar die zodanig zijn geformuleerd dat één of meerdere eigenschappen van Pareto-efficiënte oplossingen van het variantprobleem bekend zijn. Op deze wijze kan aannemelijk gemaakt worden dat ook de oplossingen van het eigenlijke optimalisatieprobleem optimaal zijn.

2. Analytische inspectie van m.n. de uiteinden van Pareto fronten.

3. Toepassing van het genetische algoritme op gelijkende, maar zodanig vereenvoudigde problemen dat ook met nadere technieken, zoals lineaire en nonlineaire programmering, oplossingen kunnen worden berekend. Hierbij is het zaak ervoor te zorgen dat de moeilijkheidsgraad van het probleem bezien vanuit het functioneren van het genetisch algoritme niet vermindert.

Evenals in case study 1 is ook in case study 4 getoond hoe onzekerheid van modeleigenschappen in de benadering van het optimalisatieprobleem kan worden verwerkt. Een expliciete formulering van het prioriteringprobleem maakte het mogelijk om verschillende, met deductie gegenereerde concepten te vergelijken met de numeriek gegenereerde resultaten van het spel-experiment. Op deze wijze konden "expert judgment" en intuïtie, die in de huidige praktijk de belangrijkste basis voor de besluitvorming m.b.t. dit prioriteringprobleem zijn, versterkt worden met een wetenschappelijke benadering.

Besluitvorming t.a.v. problemen met meerdere, conflicterende doelen waarvan de waarden niet objectief in een gemeenschappelijke schaal zijn uit te drukken vereist waardering (in de zin van waardetoekenning) van de verschillende doelen. Waardering van niet-economische effectcategoriën vereist communicatie, onderhandeling en soms zelfs confrontatie tussen de verschillende actoren. Expliciete waardering, gebaseerd op de mate van 'bereidheid te betalen' ('willingness to pay') van personen biedt goede mogelijkheden om verschillende vormen van milieukapitaal en -diensten onderling te kunnen vergelijken, m.n. wanneer de betrokken personen goed geïnformeerd zijn. Het vaststellen van formele en expliciete doelen, waarderingen en waarderingsmethoden voor milieuhygiënische en andere vraagstukken die niet uitsluitend economisch relevant zijn is daarom wenselijk. Optimalisatietechnieken en beslissingondersteunende systemen kunnen de transparantie, efficiëntie en consistentie van besluiten verbeteren. In de praktijk is echter gebleken dat beslissing-ondersteunende systemen niet in alle situaties effectief is. Wanneer slechts één actor, of een homogene groep actoren gebruik maken van een adequaat beslissingondersteunend systeem, dan is het hebben van vertrouwen in de met het systeem verkregen resultaten de belangrijkste voorwaarde voor succes. Wanneer verschillende actoren gezamenlijk gebruikmaken van een beslissingondersteunend systeem, dan dient niet alleen vertrouwen te bestaan in de resultaten, maar ook de attitude van de betrokkenen coöperatief te zijn. Een beslissingondersteunend systeem voor meerdere actoren is geschikt voor een 'poldermodel', maar veel minder voor een 'pokermodel', waar de spelers geneigd zijn hun 'kaarten tegen de borst te houden'.

Acknowledgements

I would like to thank Dr. Paul Schot for his invaluable friendship and support throughout the complete period that I worked on this thesis. I also thank Prof. Dr. Peter Burrough and Prof. Marc Bierkens for their support and advice.

I thank all the collegues at the Department of Environmental Studies for their friendship and all these inspiring exchanges of ideas and opinions.

This study could not have been completed without the support of Kiwa Water Research, RIVM-MNP, the Dutch national research institute for public health, environment and nature, WML and Hydron drinking water companies and the provincial authorities of Limburg province.

I thank drs. Aldrik Bakema and drs. Peter Heuberger from RIVM-MNP and many friends and collegues from Kiwa's Water System Management department, headed by Dr. Arthur Meuleman, for their cooperation over the past years; it definitely contributed to the completion of this thesis.

I want to express my gratitude towards Jeroen Vos for his supervision of the lay-out of my thesis nad the design of the cover, and to Ron Ottelé and Menno van der Woude for their friendship and their creative ideas that are valuable to me since so many years. Finally I would like to thank my dear wife Darja for her support, her patience with all my ongoing projects, plans and ideas and her confidence in me.

Curriculum vitae

Kees Vink werd geboren op 25 april 1960 in Putten. De middelbare schooltijd werd doorlopen aan het College Groevenbeek in Ermelo. Van 1980 tot 1984 studeerde hij Wijsbegeerte aan de Vrije Universiteit te Amsterdam. Zonder deze studie af te ronden stapte hij in 1984 over naar de studie Tropische Cultuurtechniek aan de Hogere Bosbouw- en Cultuurtechnische School in Velp, waar hij in 1988 zijn diploma behaalde. In het kader van deze studie voerde hij gedurende 7 maanden onderzoek uit in het Peruaanse Andesgebergte in de omgeving van Cuzco. Zijn afstudeeronderwerp heeft hij in Egypte uitgevoerd; het betrof een numeriek modelonderzoek naar de verdamping in de Qattarra depressie, in Noord-West Egypte. Aansluitend studeerde hij een jaar aan Silsoe College, behorende bij Cranfield University in Engeland, waar hij in 1990 zijn Master of Science graad in 'Irrigation and Water Management' ontving. Zijn master-thesis handelde over het optimaal economisch ontwerp van grondwaterputten. Hij is in de twee daaropvolgende jaren eerst enige tijd werkzaam geweest bij Tauw Infra Consult als milieuadviseur en vervolgens als hydroloog bij Kiwa in Nieuwegein. In de periode 1992 - 1994 is hij in dienst van het Ministerie van Buitenlandse Zaken als hydroloog gedetacheerd geweest in Ouagadougou, Burkina Faso bij het project Bilan d'Eau. Dit project werd uitgevoerd door ingenieursbureau Iwaco. Terug in Nederland is hij enige tijd als hydroloog werkzaam geweest in Den Bosch bij Iwaco, later opgegaan in Royal Haskoning. In 1997 is hij aangesteld als promovendus bij de vakgroep Milieukunde van de Universiteit Utrecht, met als begeleider dr. Paul Schot. Sinds 2001 is hij werkzaam als zelfstandig ondernemend hydroloog, verbonden aan het samenwerkingsverband van water- milieu en bodemadviseurs "Elementair". Hij is getrouwd met Darja Vos en heeft twee kinderen, Koos en Katinka.