Incorporating Input Data Uncertainty in Water Quality Management Models: Examples for Non-Seasonal and Seasonal Wasteload Allocation Programs

Andrews Kwabena Takyi

A Thesis

Presented to the University of Manitoba

in Partial Fulfilment of

the Requirements for the Degree of

Doctor of Philosophy

in the Department of Civil and Geological Engineering

Winnipeg, Manitoba, Canada 1995

ISBN 0-612-16322-9

Name_____

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

CIVIL

SUBJECT TERM

0543 U·M·I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

INCORPORATING INPUT DATA UNCERTAINTY IN WATER QUALITY MANAGEMENT MODELS: EXAMPLES FOR NON-SEASONAL AND SEASONAL WASTELOAD ALLOCATION PROGRAMS

BY

ANDREWS KWABENA TAKYI

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

To my dear wife Peace and daughter Emmanuella

I hereby declare that I am the sole author of this thesis.

I authorize the University of Manitoba to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Manitoba to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Abstract

Input information uncertainty (Type II uncertainty) is incorporated in water quality management models in order to identify efficient management options under stochastic stream conditions. Different approaches may be used to incorporate Type II uncertainty in non-seasonal and seasonal management models, because different levels of importance and concerns about the sources of Type II uncertainty may be identified for these models.

For non-seasonal management models, the Multiple Realization (MR) approach, which incorporates several sets of possible design stream conditions simultaneously in a single optimization model, is used to generate 'true' reliability-waste treatment cost trade-off relationships for river systems with more than one critical section. Large computational resources are often required to solve the MR model. To reduce the computational effort, new algorithms are developed to reduce the size of the optimization model. The MR model and the new solution algorithms are demonstrated for BOD wasteload allocation based on the Willamette River in Oregon. Results show that all the algorithms produce efficient solutions that are similar to the solutions based on an existing heuristic method. However, these new algorithms drastically reduce the CPU time required by the existing method to solve the MR model. One of the algorithms, that utilizes the generalization capability of artificial neural networks, appears to be the most computationally efficient, if several thousand realizations of possible design stream conditions are required to represent the stochastic water quality system.

For seasonal management models, a non-stationary Markov Chain approach for designing seasonal uniform treatment wasteload allocations that are robust with respect to the length of recorded design flow data is developed. Unlike many existing models for determining seasonal wasteload allocation, the Markov Chain approach considers a wider range of possible low flow conditions than those in the recorded data. It also incorporates explicitly the probabilistic

relationships between low flow data at adjacent gaging stations. The seasonal Markov Chain model is also demonstrated for the management of BOD waste on the Willamette River. The results of this application indicate that the management decisions derived for the Markov Chain approach are less sensitive to short and sparse recorded flow data than those for an existing approach that calculates the probability of water quality violation as the ratio of the number of years with water quality failure to the total number of years of recorded stream data.

Acknowledgements

I wish to express my sincere thanks to my advisor, Dr. Barbara J. Lence, for her financial support for my studies, and for her interest, advice, and encouragement during the period of research towards my doctoral degree. I would also like to express my gratitude to my advisory committee members, Drs. J. W. Eheart, A. S. Alfa, D. H. Burn, and S. P. Simonovic, for their valuable suggestions and thoughtful review of this research work. My sincere thanks also goes to all my colleagues, and the professors of the Water Resources Group at the University of Manitoba for assisting me in many ways during my studies in Winnipeg.

The financial support of the University of Manitoba Graduate Fellowship Award, the Petro-Canada Resources Graduate Fellowship in Environmental Issues in Engineering, and the Natural Sciences and Engineering Research Council of Canada (NSERC) are gratefully acknowledged.

Throughout my studies at the University of Manitoba, the Ansah, Dueck, Nortey, and Mensah families have been of great help to me in many ways. I wish to thank these families for being kind to me. I would also like to express my gratitude to the Kusi-Sekyere and Ntiamoah families for their assistance during the preparation of this document.

Finally, I am indebted to my wonderful wife, Peace, and my dear daughter, Emmanuella, for their moral support, patience, and the sacrifices they made, while I worked on my Ph.D. program.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Water quality management modelling is typically a multi-objective problem. The common objectives or goals are to minimize pollution control cost, maximize water quality, maximize reliability, and maximize equity. Other objectives may include minimizing regret, minimizing the number of water quality violations, minimize the duration of the longest water quality violation episode, minimizing the maximum water quality violation, and maximizing robustness. The management problem is often formulated as a mathematical programming model that optimizes one of the above objectives subject to limits on one, or more, of the other goals.

The mathematical programming model is generally used to generate information about management options for a river system under a set of non-seasonal or seasonal steady state conditions. The alternative management strategies identified with this model may be analyzed further using more sophisticated dynamic water quality models

to refine the choice of management decisions. However, many water quality systems are poorly defined with respect to the pollutant transport model (Type I uncertainty) and the input information (Type II uncertainty). Therefore, the information generated by either dynamic or steady state mathematical programming models may be unreliable. These models may fail to identify the efficient management alternatives that have to be analyzed further. Also, the pollution abatement decisions based on poor information may result in failure to meet the management goals.

Generally, the frequency, duration, and magnitude of violations of a given water quality standard are indices of pollution control performance that represent the reliability, resiliency, and vulnerability, respectively, of water quality management options under Type I and II uncertainty. For a given pollution control option, the variability or sensitivity of these indices to input or design information reflects the robustness of this option to design inputs (i.e., how satisfactory the option performs under a wide range of possible future pollution discharges and background water quality conditions). The reliability criterion describes how likely the water quality goals may be achieved, while the resiliency and vulnerability criteria give indications of the ease with which the water quality system recovers from a failure and the severity of the consequences of water quality violations, respectively.

This research presents two different methods that may be used to address problems with Type II uncertainty in non-seasonal and seasonal water quality management models. For non-seasonal management models, the Multiple Realization (MR) approach, which incorporates several sets of possible design stream conditions simultaneously in a single optimization model, is used to generate 'true' reliability-waste treatment cost trade-off relationships. In the case of seasonal models, a Markov Chain approach is

developed to design management decisions that are robust to limited recorded low flow data, since the availability of these data, especially during non-critical water quality seasons, may be one of the important concerns of uncertainty in seasonal water quality modelling.

Research directed toward incorporating uncertainty in water quality management models typically concentrates on Type II uncertainty. With the exception of the stochastic dynamic programming (SDP) models utilized by Lohani and Hee (1983) and Cardwell and Ellis (1993), there are two widely used methods for incorporating Type II uncertainty in non-seasonal water quality management models. These are:

1. Combined simulation and optimization (SIOP) modelling; and

2. Chance constrained optimization (CCP) modelling.

Generally, an SIOP model is utilized to allow easy mathematical and statistical representation of complex water quality management systems and to extract important information for planning or managing such systems. Cumulative distributions of objective function values (e.g., waste treatment cost) or decision variables (e.g., waste treatment levels of polluters) are some of the important results produced by a traditional SIOP model. Often, information derived from these cumulative distributions include trade-offs between the objectives or decisions and the reliability of these solutions. However, for a multi-dimensional decision space, the trade-off information obtained may represent an inefficient solution set, because most of the solutions produced by the traditional SIOP model are often inferior to the nondominated solutions for the objective function, or decision variables, and reliability.

For the CCP approach, constraints that ensure acceptable water quality are formulated in a probabilistic form. To transform the probabilistic model into a deterministic one, which is amenable to solution with existing optimization algorithms, several simplifying assumptions about the physical system, the uncertain input information and their distributions, and the management model are often required. Therefore, the results from such a model may not be dependable for many practical applications.

Traditional water quality management programs, also known as direct regulation, determine the required pollutant removal levels at point sources so that specified water quality standards are maintained along the receiving stream. The optimization models, which are used to design these programs, select the same waste treatment strategies for all periods of the year (i.e., the wasteload allocations are non-seasonal) based on stream background conditions during the critical water quality periods of the year. These non-seasonal wasteload allocations typically may be classified as least cost (LC) allocations, uniform treatment (UT) allocations, or some combination of LC and UT allocations. For the LC programs, the resulting management solution minimizes the total waste treatment cost for all polluters, while maintaining acceptable water quality standards at pre-selected locations, or checkpoints, in the stream. Under this program, dischargers who have lower marginal waste treatment costs per unit water quality improvement at the critical checkpoint(s) of the stream typically remove a higher percentage of the waste they produce than those that face higher marginal waste treatment costs. UT programs determine the same level of waste removal for all polluters in the interest of equity, and the management solution for these programs normally minimizes the uniform waste treatment level subject to constraints that ensure acceptable water qauility at the stream checkpoints.

Due to the increasing public concern about the deterioration of the quality of many fresh water resources, the large amounts of money spent on waste treatment facilities, and the increasing complexity of water quality systems, recent research on water pollution control has focussed on innovative programs for managing the quality of fresh water resources. These innovative programs include seasonal programs and incentive-based programs, such as effluent charge (EC) programs, transferable discharge permit (TDP) programs, and subsidy programs.

Seasonal models appear to be the most widely applied innovative water quality management program in practice. These seasonal models may be LC, UT, or incentive-based. A survey in 1983 indicated that 45 states in the United States used some form of seasonal program (Lamb and Hull, 1985). Generally, seasonal programs are more cost efficient than equivalent non-seasonal programs, because seasonal programs utilize the varying assimilative capacity of a river during different periods of the year. Several studies have demonstrated the cost efficiency of seasonal water quality programs (see, e.g., O'Neil, 1980 and Eheart et al., 1987). However, most of these studies, do not address the concerns of the increased input information requirement and the limited historical data normally available for these seasonal programs.

Although seasonal programs take advantage of the temporal variability of the assimilative capacity of water bodies, the management models used to analyze these seasonal programs are often complex and the historical input information available is often short or sparse. Furthermore, the traditional methods for incorporating uncertainty into simple non-seasonal water quality management models may not be adequately and easily applied to such complex management models. One way to address the uncertainty in complex water quality systems is to develop management models that produce solutions

that are robust to limited recorded stream data.

1.2 Research objectives and scope of work

This research introduces approaches that may be used to address Type II uncertainty in non-seasonal and seasonal water quality management programs. The objectives of the research are:

1. to develop general methods that improve the SIOP approach for incorporating uncertain input information in non-seasonal water quality management programs;
2. to develop a seasonal water quality management model that designs allocations that are robust to limited historical input information; and
3. to demonstrate (1) and (2) for the control of biochemical oxygen demanding (BOD) waste discharges in an example river basin.

In this thesis, different approaches are presented for addressing Type II uncertainty in non-seasonal and seasonal programs, because different levels of importance and concern about the sources of input information uncertainty may be identified for these programs. These differences result from different mathematical formulations of the most commonly implemented non-seasonal and seasonal direct regulation programs. For example, while UT is often implemented for seasonal programs, LC as well as UT models are often analysed for non-seasonal programs.

The first research objective is to improve the SIOP approach so that accurate trade-off relationships can be developed for the water quality management objectives, or

decisions, and reliability, or probability of water quality violation. This improved SIOP approach, the MR approach, is demonstrated for some traditional water quality management programs and for an example river basin. The approach easily allows incorporation of uncertainty from all stream background information and the inter-relationships among these information. The large computational resources required by the MR model are drastically reduced by using techniques that identify and incorporate only potentially important sets of design stream conditions in the MR model.

The poor water quality input information for many river basins exacerbate problems of Type II uncertainty for seasonal management programs because historical information for average flow periods of the year is often unavailable. Although such average flow conditions are non-critical for non-seasonal water quality programs, they may constitute design information of some periods of the year for a seasonal program. The lack of historical information during non-critical water quality periods may be attributed to the inability to measure these data due to adverse hydrological, meteorological, or physical conditions. Sparse or limited historical data for non-critical periods may also be due to the failure to measure this information because of their irrelevance in the design of many non-seasonal management programs which often require data for only the critical water quality period of the year. Therefore, there is a potential for a higher risk of water quality violation for solutions of seasonal programs than that specified by the threshold risk value included in a model. This higher risk of not meeting the water quality goal may be reduced by risk-equivalent modelling approaches that design management strategies that are robust to limited recorded data. Since low flow and high temperature values for different periods of the year are the most important stochastic design stream conditions for seasonal programs, and since temperature data can often

be obtained or estimated for many locations, the seasonal model developed in this research (Objective 2) considers output performance robustness with respect to short and sparse low flow data only.

In Chapter 2, a review of research that incorporates Type II uncertainty in (1) non-seasonal direct regulation surface water quality management models, (2) air and groundwater quality management models, and (3) seasonal water quality management models, is presented. Deterministic surface water models are also reviewed in Chapter 2. In Chapter 3, a new combined simulation and optimization modelling approach is presented and demonstrated for BOD waste management based on information for the Willamette River basin in Oregon. This new approach, called the Multiple Realization (MR) method, includes several stream background conditions in a single optimization model to obtain an efficient solution at a given reliability level. However, the MR approach may create large optimization problems that require large computational resources for even small water quality systems. Therefore, new algorithms are developed in Chapter 4 to reduce the size and computational resource requirements of the MR model. In Chapter 5, a neural network tool is developed to enhance the computational efficiency of one of the algorithms presented in Chapter 4. In Chapter 6, a non-stationary Markov chain model for a seasonal water quality management program is developed and demonstrated for BOD waste discharges into the Willamete River. The Markov chain model is used to design robust management decisions by incorporating uncertainty in low flow data due to limited historical records. Summary, conclusions, and recommendations for future research are included in Chapter 7.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews the literature in three broad areas of modelling, namely: (1) deterministic and stochastic surface water quality management modelling of non-seasonal direct regulation programs; (2) incorporation of uncertainty in groundwater quality and air quality management models; and (3) seasonal water quality management modelling. Groundwater quality and air quality management modelling under uncertain environmental conditions have been included in this review because the mathematical formulations of these problems are similar to those of stochastic surface water quality management models. Also, the MR approach, which is introduced and modified in this thesis, has been applied often to groundwater quality management problems, but not to surface water quality management problems.

The main goal of water quality management in rivers is to prevent or minimize degradation of the quality of surface water bodies, such as streams, estuaries, reser-

voirs and lakes from pollution, so that beneficial uses (e.g., domestic, industrial, and agricultural water supply, recreation, and fishing) and the aquatic ecosystem are not significantly impaired. However, this goal must be achieved without jeopardizing regional socio-economic development. Therefore, surface water quality management is a challenge that involves many stakeholders, including the public, environmental groups, industries, regional authorities, and governments. The management decisions or policies used to attain the water quality goal depend on what the major stakeholders would like to obtain from the water body, the socio-economic cost of the decisions, and the uncertainty and risk of not attaining the desired goal with selected management decisions.

Generally, the quality of surface water bodies are managed based on guidelines and policies developed by national, state, or regional environmental protection authorities or boards formed by governments or based on laws made by government. Different forms of these policies or laws that are developed by water quality policy makers in consultation with major stakeholders may apply to a stream segment, a river basin, or rivers in a sub-region or country.

In many countries, the environmental protection authority enforces the water quality management policies by granting pollution license to waste dischargers, such as municipal and industrial waste treatment plants. This license may either specify an ambient water quality standard to be maintained in a river by waste discharges or give a permit to the polluter to discharge a certain amount of waste into the river. Generally, compliance of these policies is overseen by the environmental protection agency that may issue penalties to waste dischargers who violate their license. However, in recent times, the public, environmental groups, and other stakeholders have embarked on co-operative

action to share the responsibility of effective monitoring and reporting arrangements to ensure compliance of water quality policy. In the United States and other developed countries, environmental groups and other stakeholders of stream quality resources have sometimes taken legal action and sought damages against waste dischargers who violate water quality standards.

Simulation and optimization models are used to generate important data about the water quality system to be managed, including efficient and reliable options for protecting present and future beneficial uses of the river and the aquatic ecosystem, while minimizing socio-economic cost of the management decisions. The data obtained from these models complement other information, such as existing and projected regional economic development levels and the experience of water quality policy makers in deriving pollution management plans or decisions.

Systems analysis methods for water quality management modelling were introduced in the 1960s. For a given set of stream background conditions and water quality goals (e.g., minimize cost, maximize equity, maximize water quality, or maximize reliability), these techniques may be used to determine the optimal pollution abatement decisions for diffused, or non-point, sources of pollution (e.g., land management practices) and for point sources of industrial and municipal waste discharges (e.g., waste removal levels of polluters) into a stream.

The main components of a water quality management system are a hydraulic simulation model, a pollutant transport simulation model, and a water quality management optimization model. The physical representation of this management system is usually comprised of reaches along a river, defined in such a way that the hydraulic, hydro-

logic, background water quality, and pollutant discharge characteristics within a reach are homogeneous. Thus, a new reach is defined by the confluence of a tributary, a marked change in the physical environment of the river, or the presence of new sources of pollution (i.e., point and non-point sources). The modelling effort of the management system often requires further discretization of each reach into segments whose boundaries represent water quality checkpoints. In most studies, the checkpoints are pre-selected such that if water quality standards are maintained at these checkpoints then acceptable water quality is guaranteed along the whole stretch of the river.

The hydraulic simulation model is linked to the pollutant transport simulation model, which computes coefficients or functions that relate the water quality level at each checkpoint to pollution at given locations and therefore to pollution abatement decisions at these locations. For pollutant transport simulation models that are linear with respect to pollution control decisions, the coefficients that represent the water quality improvements at the checkpoints per unit waste reduction by each polluter are often called transfer coefficients. These transfer coefficients or functions and the management objectives are used to formulate an optimization model that ultimately determines the water quality management decisions for the river system.

2.2 Non-seasonal water quality management models

2.2.1 Deterministic models

Deterministic systems analysis techniques used to model water quality management systems include simulation models (Loucks and Lynn, 1966; Warren and Bewtra, 1974; and Orlob, 1982) and optimization models (Sobel, 1965; Kerri, 1966; Loucks et al., 1967; ReVelle et al., 1968; Anderson and Day, 1968; Arbabi and Elzinga, 1975; and Biswas 1981). The simulation approach invokes sets of feasible pollution control decisions for polluters and then evaluates the resulting water quality management objectives for each set of control decisions. Selection of the optimal solution is based on a visual inspection of values of the resulting water quality levels and other management objectives, or based on some type of multi-criteria ranking procedure. While this approach is easy to apply for complex water quality systems, even for moderately sized systems, the number of possible feasible sets of options for waste treatment levels for point source polluters and land management practices for non-point polluters could be very large. Ranking the feasible decision sets and selecting the optimal solution for a typical multi-objective water quality system may be a difficult task. Furthermore, in order to incorporate input information uncertainty in such a modelling effort, the number of evaluations of water quality goals could increase to an impractical level and the ranking of the sets of management decisions can be very complicated and untractable.

Loucks and Lynn (1966) develop four mathematical simulation models to predict the probability distribution of the minimum dissolve oxygen (DO) concentration down-

stream of any waste treatment facility. Their models use the Markov chain modelling approach to predict the probability of having less than a given DO concentration, for any consecutive time period, downstream from a point source waste discharge with known effluent characteristics. These models consider a wide variation of stream and sewage flow conditions and the resulting variation in treatment plant efficiencies and stream assimilative capacities. The authors determine the distributions of minimum DO concentrations in a hypothetical stream for 1, 2, and 3 consecutive day-long periods. Although, an optimal set of waste treatment levels are not selected, the authors present a discussion of how these mathematical simulation models may be used to determine required treatment plant efficiencies for various sewer flows.

Warren and Bewtra (1974) use a deterministic simulation model to study the effects of combined sewer overflows on water quality in a typical Ontario, Canada community. The pollutants considered are BOD, total suspended solids, total phosphates and fecal coliform. The models predict stream quality variations due to different scenarios of interceptor sewer capacity, storage vessel or plant capacity, and treatment plant capacity. By including a cost function for all system components in the model, the authors are able to analyze and evaluate various methods of pollution abatement decisions.

The most common optimization techniques used for water quality management in the literature include linear programming, LP (Sobel, 1965; Kerri, 1966; Loucks et al., 1967; ReVelle et al., 1968; Anderson and Day, 1968; Arbabi and Elzinga, 1975; and Biswas, 1981), dynamic programming, DP (Liebmann and Lynn, 1966; Dysart and Hines, 1970; Nocholson et al., 1970; Shih, 1970; Dysart, 1970; Newsome, 1972; Warn, 1978; Hahn and Cembrowicz, 1981; and Somlyódy et al. 1993), nonlinear programming, NLP (Graves et al., 1972; Hwang et al., 1973; Bayer, 1974; Pratishthananda and Bishop,

1977; and Burn and McBean, 1987). LP is one of the most widely used systems analysis methods for water quality management modelling. This popularity may be attributed to the important role that DO plays in water quality determination (Streeter and Phelps, 1925; O'Connor and Dobbins, 1958, Camp, 1963; Dobbins, 1964; Thomann, 1963, 1974; and O'Connor and DiToro, 1970) and the ease with which stream DO response to traditional BOD waste discharge, and waste treatment cost functions for BOD removal, may be incorporated into a linear optimization problem. Such LP formulations typically utilize the superposition property of the steady state Streeter-Phelps model (Streeter and Phelps, 1925) for DO and BOD or the Camp-Dobbins modification (Camp, 1963; and Dobbins, 1964) of this BOD-DO model. Mixed integer linear programming (MIP), which is classified as an LP technique in this thesis, is also used by some researchers for water quality management modelling (see, e.g., Loucks et al., 1981; and Burn, 1989)

In addition to LP being the easiest of the optimization techniques to solve, general-purpose solution packages are available for solving LPs. Also, the dual variables which are obtained through the LP solution may provide useful information about the physical river system, the water quality goals, and the alternative ways by which desired objectives can be achieved.

LP techniques for the optimum management of water quality in an estuarine environment are described by Thomann and Sobel (1964). Although the authors present results from estuarine water quality simulation models, which are linked to the LP optimization formulation, they do not solve the optimization problem.

Sobel (1965) outlines the nature of regional water quality systems and presents LP models for several DO improvement problems. The models include an LC program

and a model that maximizes the benefit-cost ratio subject to constraints on acceptable DO levels in a river. These LP models are contrasted with an MIP formulation of the traditional UT program. The author does not demonstrate the models for any river system.

Kerri (1966) uses an LP formulation to analyze five BOD waste management alternatives for the Willamette River in Oregon. The basic LP model is based on an LC program, and acceptable water quality levels are defined by a DO standard within a critical reach of the river. He obtains cost curves for municipal waste dischargers based on the work of Logan et al. (1962) and develops industrial waste treatment cost curves based on adjustments to the cost curves for the municipal treatment facilities. His analysis is for a 5.00 mg/l DO standard during a stream flow of 10,200 m^3/day at Salem, Oregon.

Several LP models for water quality management, similar to those of Sobel (1965) and Kerri (1966), and based on actual river basins or on hypothetical systems are reported in the literature (see, e.g., Loucks et al., 1967; ReVelle et al., 1968; Anderson and Day, 1968; Arbabi and Elzinga, 1975; and Biswas, 1981).

Sobel (1969) uses a Chebyshev criterion to develop a water quality management model for optimal waste discharge schedules from a holding pond. His model maximizes the minimum water quality in the objective function subject to limits on a waste treatment cost budget and on the amount of waste that can be discharged into a river. Although the model formulation is nonlinear, the author transforms it to an equivalent LP problem.

Brill et al. (1976) use an LP to examine economic efficiency and equity of wasteload

allocation by developing a trade-off curve between these two water quality management objectives for the Delaware Estuary in the Eastern United States. They define three measures of equity namely, (1) range equity, (2) summed deviation equity, and (3) maximum efficiency equity, based on waste removal levels. Their models have objectives that maximize equity subject to a waste treatment cost budget and water quality constraints. They vary the total program budget from the cost of the LC program to the cost of the UT program, and obtain a trade-off between total waste treatment cost and each of the three equity measures.

Michels (1987) formulates five LP models that incorporate reserve capacity in the wasteload allocation process. In her models, reserve capacity is defined as the percentage or amount of the polluters' allowable effluent load that may be set aside to provide higher water quality in order to accommodate future growth in waste production or uncertainty in the water quality system. Apart from one of her models, the General Least Cost Model Which Includes Reserve, all of the others require an initial wasteload allocation program as a starting point from which further reductions are considered to provide for a reserve in waste assimilative capacity.

Application of DP for water quality management is becoming increasingly popular because the technique does not require (1) convexity of waste treatment cost curves and (2) linearity of the pollutant transport mechanism (see, e.g., Liebmann and Lynn, 1966; Dysart and Hines, 1970; Nocholson et al., 1970; Shih, 1970; Dysart, 1970; Newsome, 1972; Warn, 1978; Hahn and Cembrowicz, 1981; Lohani and Hee, 1983; Cardwell and Ellis, 1993; and Somlyódy et al. 1993). Therefore, current nonlinear and highly advanced water quality simulation models, such as QUAL2E (Brown and Barnwell, 1987) and WASP5 (Ambrose et al., 1991), can be linked to a DP based management program.

Furthermore, discrete water quality management options can easily be incorporated in a DP model.

For a DP application to water quality management, river reaches correspond to stages, water quality parameters or measures are the state variables, and pollution control actions are the decision variables. The objective function, which may be waste treatment cost, or some other utility function, serves as the return function. Water quality simulation models connect state variable and stage combinations.

A DP formulation for an LC BOD wasteload allocation is presented by Liebman and Lynn (1966). The authors use the Streeter-Phelps BOD-DO simulation model and DO standards that change along the river for a system based on a simplified version of the Willamette River in Oregon.

Dysart and Hines (1970) and Dysart (1970) present a two-dimensional DP approach for water quality control for the case where there is significant interaction of wastes in a stream. Unlike most previous studies that (1) consider a single pollutant and a single water quality parameter or (2) assume independence between the stream quality effects of multiple pollutants, this work explicitly includes the interactions between different pollutants in the same stream. The model is demonstrated for the LC management of thermal and organic waste discharges into the Chattahoochee River in Georgia.

Nicholson et al. (1970) use a DP model to control water quality through both point source treatment levels and reservoir releases for low flow augmentation. Their model outlines a methodology for evaluating alternatives for alleviating detrimental downstream DO effects caused by thermal stratification in reservoirs. The authors apply their methodology to the Clark Hill Reservoir on the Savanna River in Georgia.

Other deterministic DP applications for stream quality management include analyses of the Trent River in the United Kingdom (Newsome, 1972 and Warn, 1978), the Neckar River in Germany (Hahn and Cembrowicz, 1981), and the Nitra River in Slovakia (Somlyódy et al., 1993).

An NLP model may often be the most realistic representation of a water quality management system because the objective functions (e.g., the cost functions of waste treatment plants) and the pollutant transport functions are generally nonlinear. However, NLP models have been less popular in water quality management applications than the other management models due to (1) the complexity of such approaches, (2) the relatively large computational requirements, and (3) the increased difficulty in incorporating input information uncertainty.

Graves et al. (1972) detail an application of NLP to the problem of LC water quality control in an estuary. The model allows for the possibilities of at-source treatment, regional treatment plants, and by-pass piping. Data from the Delaware Estuary are used to demonstrate a large-scale problem.

A river basin-wide water quality management model based on an NLP is presented by Hwang et al. (1973). Their model controls thermal and organic waste discharges into the Chattahoochee River. The related water quality standards used are the minimum DO concentration, the maximum allowable BOD concentration, the maximum allowable stream temperature, and the allowable rise in stream temperature. The NLP model is solved using the Generalized Reduced Gradient Method.

The work of Bayer (1974), Pratishthananda and Bishop (1977), and Burn and McBean (1987) present other examples of the application of NLP modelling for stream

quality management systems. Bayer (1974) uses an NLP to design an LC BOD wasteload allocation for polluters on the Willamette River in Oregon, and concludes that the results obtained are better than the results based an LP and a DP. Pratishthananda and Bishop (1977) solve a regional water resource planning problem, which includes a wastewater treatment plant capacity expansion, for Salt Lake County in Utah using a nonlinear multilevel transportation problem. Burn and McBean (1987) use a nonlinear optimization model to determine the optimal removal of coliform bacteria for a case study based on the Grand River in Ontario.

2.2.2 Stochastic Models

The results produced by the deterministic efforts described in the last section and the management information extracted from these results depend on the input information, the characteristics of polluters' effluent, and the type of pollutant transport simulation model used. For a given stream quality management system and pollutant transport model, many of the inputs are uncertain or stochastic in nature. Therefore, a realistic modelling approach will incorporate the uncertainty in using the correct transport model (Type I uncertainty) and the appropriate input information (Type II uncertainty) into the transport and management models. This section presents the sources of Type I and Type II uncertainty, the modelling techniques used to incorporate Type II uncertainty in water quality management models and their drawbacks, and the applications that address Type II uncertainty in the literature.

Type I uncertainty may arise from the lack of complete understanding of the mechanics or relationships between parameters, variables, and model output, the poor

choice of assumptions on which the model formulation is based, and the presence of objectives or elements of the water quality system that cannot be represented mathematically. Type II uncertainty is caused by the stochastic nature of water quality input information, the limited amount of input information available for many river basins, and the errors made in measuring, processing, and recording data. Uncertain water quality input information may include design flows and temperatures, reaction rates, stream velocity, background water quality, effluent concentrations and discharge rates of point source polluters, waste treatment cost coefficients for polluters, and the amount and sources of non-point pollution.

There are few attempts to incorporate Type I uncertainty in stream quality management models (Cardwell and Ellis 1993), probably because there is the tendency to address uncertainty in pollutant transport models by developing more sophisticated models that more accurately represent the physical, chemical, and biological processes rather than including such considerations into a management model framework. On the other hand, there have been several studies directed towards incorporating Type II uncertainty in water quality management models (Loucks and Lynn, 1966; Lohani and Thanh, 1978, 1979; Burn and McBean, 1985, 1986; Ellis, 1987; and Burn, 1989). This emphasis may be attributed to concerns related to limited water quality data available for many river systems and to the high cost of water quality data collection programs.

The most popular techniques used to address Type II uncertainty in water quality management models are CCP and SIOP. However, there are other less frequently reported methods such as stochastic linear programming, SLP (Sobel, 1965) and SDP (Lohani and Hee, 1983 and Cardwell and Ellis, 1993).

The SIOP procedure for water quality management modelling consists of generating a number of realizations or scenarios of design stream conditions, solving a water quality simulation and an optimization model for each of the realizations to obtain an objective function value and a corresponding set of pollution abatement decisions, ranking the objective function values, and obtaining a cumulative probability distribution based on the objective function values, their ranks, and the total number of realization utilized. Some of the important information from the cumulative distribution are the trade-offs between the objectives, or decisions, and the probability of water quality violation. However, if the decision variables of the management model are multi-dimensional, accurate evaluation and ordering of the vectors of management decisions and the corresponding objective functions can be a complicated task. This complication arises because there are no guarantees of unique correspondence between the optimal value of the objective function and the corresponding vector of decision variables. Thus, for different input information, the same optimal value of the objective function may be obtained, but the optimal vectors of decision variables and the corresponding probabilities of system failure may be different. Under such situations, some of the solutions produced by a traditional SIOP model may be inferior to the 'true' trade-off curve for the objective function and the reliability of water quality maintenance (see, e.g., Morgan et al., 1993).

To incorporate Type II uncertainty, the CCP approach formulates the management problem such that the maintenance of acceptable water quality levels at pre-selected checkpoints are represented with probabilistic constraints. Each of these probabilistic water quality constraints is assigned a lower bound for the reliability level that must be achieved. The CCP model, for a given set of constraint reliability levels, is transformed to a deterministic equivalent model which may be solved using existing mathematical

programming algorithms. The CCP technique often requires several assumptions to simplify the water quality management model and the solution procedure. In fact, most of the CCP models for water quality management found in the literature make assumptions about (1) the distribution of the background data, (2) the management objectives, or (3) the physical water quality system (see, e.g., Lohani and Thanh, 1978, 1979; Burn and McBean, 1985, 1986; Fujiwara et al., 1986, 1987; Ellis, 1987).

Sobel (1965) incorporates Type II uncertainty in water quality management modelling using an SLP model to optimize water quality objectives. The SLP model is reformulated as a stochastic quadratic programming problem with a new objective that minimizes the variance of DO improvement in all reaches and constraints that impose a budget limitation on total waste treatment cost.

Lohani and Thanh (1978) develop a CCP model for stream quality management that assumes that streamflow is the only stochastic input to the model. A chance constraint is developed for the streamflow at each pre-selected water quality checkpoint such that a given risk of violating the performance requirement is not exceeded. The authors apply their model for LC and UT management of seven BOD waste dischargers on the Hsintien River in Taiwan. The correlation between streamflow and some of the other stochastic stream variables such as travel time and reaeration are ignored in their model. Furthermore, the spatial distribution of streamflow is not incorporated in the CCP model.

Lohani and Thanh (1979) analyze the CCP model for stream BOD control for four different operating policies for BOD waste discharge management on the Hsintien River. They also apply the CCP model for a monthly BOD removal scheme and compare the

results with an annual scheme for a given degree of reliability of meeting a desired DO limit for each reach. As expected, they showed that the monthly BOD removal scheme is more cost efficient than the annual scheme. Lohani and Saleemi (1982) extended this CCP model to include waste water flow, deoxygenation rate, reaeration rate, and waste strength as random variables.

Lohani and Hee (1983) consider streamflow as a stochastic input in a DP formulation of a water quality management system. Their study solves a wasteload allocation problem using a DP model that treats inflow as a random variable and reformulates the single stage reliability constraints into deterministic equivalents.

Fujiwara et al. (1986) develop and demonstrate a CCP model that improves on the model developed by Lohani and Thanh (1978) by removing the requirement for *a-priori* knowledge of the initial DO deficit at the top of intermediate reaches for the water quality system. This improved model also considers streamflow as the only stochastic input information to the model. Therefore, solutions obtained from the improved model may also have limited practical use as in the case of the model developed by Lohani and Thanh (1978).

Fujiwara et al. (1987) propose and demonstrate a LC linear CCP model for the control of BOD waste discharges into a river. They consider storm water discharges and streamflows in the main river and the tributaries as random variables. By obtaining the joint probability distribution of the stochastic storm water discharges and streamflows, they transform the CCP model formulation into an equivalent linearly constrained program having a staircase structure. This CCP model is rational in terms of incorporating the spatial dependence of storm water discharges and the streamflows in the main river

and tributaries in a stochastic model. A major drawback of this model is the assumption that travel time, reaction rates, sedimentation-scour coefficients, and BOD runoff are independent of streamflow. The assumption of independence between streamflow and these input information is made to reduce the complexity of the stochastic model.

Burn (1989) uses a SIOP approach to incorporate the uncertainty in stream background conditions in a management model. Monte Carlo simulation is used to produce several sets of steady state water quality transfer coefficients to reflect the uncertainty in streamflow, stream velocity, and reaction rates that occur in the water body. Constraints for this model maintain acceptable water quality at selected locations along the river. For each set of water quality constraints, he utilizes a MIP model with a minimum cost objective to determine discrete waste removal levels for each of the point source polluters along a river. Although not a required assumption for his approach, in his illustrative example he assumes complete colinearity between the flow at all portions of the water body, such that the flow at the most downstream portion of the river is used to determine the flow in all other reaches based on some functional relationship. He generates other components of the transfer coefficients used in the water quality constraints (e.g., travel time and reaction rates) based on synthetic flows and on functional relationships reflecting the historically observed correlation between individual model components and flow. The expected variability of the components of the transfer coefficients is preserved by adding a zero-mean normally distributed noise term to the deterministic functional relationships for these components.

Burn (1989) obtains cumulative distributions for the total cost of waste treatment and the corresponding vector of BOD waste treatment levels for the point source polluters based on the solutions of the integer programming model for the different sets of

water quality constraints based on the different transfer coefficients. He indicates that the non-exceedence probability for a given budget is related to a measure of the risk of not attaining the desired water quality level. He demonstrates, for a case study, that this cumulative probability level and the probability of no water quality violation are closely related at the low levels of risk of water quality violations that are of interest to risk-averse decision makers.

Burn and McBean (1985) represent the stochastic water quality management system with a linear CCP model. They consider Type II uncertainty and characterize this uncertainty using First Order Error Analysis. Since it is difficult to solve an SLP model if there is uncertainty in the transfer coefficients of the water quality constraints, the authors reformulate the traditional water quality management model so that the uncertain transfer coefficients are in the objective function only. Their model determines a wasteload allocation that maximizes the sum of the water quality improvements at all checkpoints. The constraints limit the other objectives of the management model (e.g., the total cost of waste treatment and the equity among waste dischargers) to pre-assigned bounds. A dual problem is then formulated from this model to transfer the stochastic coefficients in the objective function to the right hand side of the constraints. This makes the stochastic model amenable to solution using available linear CCP techniques. The authors demonstrate their model for the management of BOD waste discharges in an example river based on the Speed River near Guelph,
Ontario. They show that a lognormal distribution for the transfer coefficients provides a more satisfactory fit to the distribution of DO deficit than that provided by a normal distribution.

Burn and McBean (1986) apply the stochastic model of Burn and McBean(1985) to

BOD-DO management modelling for the Schuylkill River basin near Reading, Pennsylvania. They include equity concerns in this application and obtain trade-off curves for waste treatment cost and water quality response at different equity levels. The authors observe a fairly good overall agreement between DO deficit obtained from their stochastic planning model and a Monte Carlo simulation model at the twelve water quality checkpoints used.

Ellis (1987) develops an LC nonlinear CCP model for the allocation of BOD waste removal efficiencies for point source discharges into a river. The water quality goal used is a probabilistic restriction on the maximum allowable DO deficit at selected locations along the river. The author assumes randomness in all variables present in the Camp-Dobbins modification of the Streeter-Phelps equation for steady state BOD-DO interaction in a river (Camp, 1963; and Dobbins, 1964). He casts the uncertainty in the random variables into the matrix of DO improvement coefficients by using stochastic simulation to estimate the mean and the variance-covariance for these coefficients. The coefficients are assumed to be normally distributed. He presents a variation of CCP, called Imbedded CCP, which allows nonexpected values of the first and second moments of the random coefficients to be used for analysis. The model ignores the dependence among the vectors of water quality improvement coefficients at pre-selected checkpoints. The author also presents a joint chance constraint programming formulation of the water quality optimization problem and notes that it is generally impractical to assume independence among the row vectors of the matrix of water quality improvement coefficients.

Recent research by Burn and Lence (1992) develops four water quality management models to obtain trade-offs between total waste treatment budget and (1) maximum

water quality violation, (2) maximum regret, (3) total water quality violation, and (4) total regret. Two of the four models use Chebyshev's criterion; they are the minimize maximum violation and minimize maximum regret models. In order to reflect alternative hydrologic, meteorologic, and pollutant loading conditions, five sets of input data are assumed to represent all possible input data scenarios. Thus, these scenarios are used to account for uncertainty in input data. They define a violation as a shortfall in the water quality goal at a checkpoint and a regret as the difference between the water quality violations that occur under one input scenario and those that occur under a different scenario, which represents the actual input conditions. The authors apply the models to the management of BOD discharges in the Willamette River in Oregon.

Cardwell and Ellis (1993) present stochastic dynamic programming models for waste load allocation from multiple point sources. Their models include both Type I and Type II uncertainty and are capable of employing sophisticated water quality simulation models. Frequency- and maximum violation-based regret management models are used to incorporate Type I uncertainty that includes (1) the Streeter-Phelps equation, (2) QUAL2E, and (3) WASP4 for stream BOD discharges into the Schuykill River near Reading, Pennsylvania. To incorporate Type II uncertainty into the stochastic dynamic programming models, Markov transition matrices specific to each of the three water quality simulation models are developed using Monte Carlo simulation. However, due to the large computational burden, only one hundred Monte Carlo simulations are used to determine the transition matrices for the Streeter-Phelps and QUAL2E models, while as few as ten are used to estimate transition matrices for the WASP4 model.

To determine a non-uniform wasteload allocation for a stochastic water quality management system, the analytical and statistical formulations are often complicated and

nonlinear, and may be difficult to solve. Besides, as indicated in Chapter 1, a typical real life water quality management problem is multi-objective in nature. For such multi-objective management problems, the objectives often conflict with one another. Also, some of the objective functions (e.g., maximizing reliability) and constraints may be implicit functions of the decision variables. Monte Carlo simulation is one of the easiest and most effective methods of accurately evaluating values of such objective functions. The common approach for obtaining a reliability level is to derive the reliability value, corresponding to a specified solution, from the cumulative probability function of the other objectives (e.g., minimum cost) or from the other sets of decision variables, based on single realization solutions from the SIOP model. For example, if the cumulative probability of exceedance of the cost of waste treatment corresponding to a given LC solution is 0.80, a common approximation is to assume that, if implemented, this solution will ensure acceptable water quality 80% of the time. To improve the estimation of the reliability value of a given solution, the decision variables of the solution are invoked to determine the proportion of the total number of Monte Carlo simulations for which acceptable water quality is obtained (i.e., reliability value). However, for some water quality management systems, these approaches may produce inefficient decisions, if the solutions, whose reliabilities are evaluated, are based on the traditional SIOP model. When reviewing stochastic models which account for probabilistic phenomena, Fuessle et al (1987) wrote:

> "Despite the complementarity of optimization and simulation, difficulties arising from stochastic aspects of the planning problem remain. For example, it has not been shown how to search efficiently for a good solution using optimization and simulation interactively."

Most of the studies that incorporate uncertainty in water quality management models include reliability, but ignore other important risk indices such as resiliency, vulnerability and robustness. However, all these risk indices are measures of system performance that may offer important insights and information to decision makers for formulating successful water quality policies under environmental and system uncertainty. The information generated by incorporating these indices in a management model are particularly important for water quality systems in which frequency, duration, magnitude, and variability of these indices indicate the extent of environmental damage under critical hydrological and background water quality conditions. The importance of these risk criteria are illustrated and discussed in detail for water resources management systems by Glantz (1982), Hashimoto et al. (1982a, 1982b), and Fiering (1982a, 1982b, 1982c, 1982d).

Glantz (1982) raises questions about societal response to water supply estimates and the benefits to society of good forecasts and subsequent water supply policy as well as the cost of erroneous ones. He also raises the issue of who should be responsible for the socio-economic implications of failed water management policy based on erroneous forecasts. His study is based on actual forecast of total water supply available and subsequent water allocation policy made by the Bureau of Reclamation for irrigators in the Yakima Valley in Washington.

Hashimoto et al. (1982a) discuss reliability, resiliency, and vulnerability criteria as applied to water resources systems. They derive mathematical expressions for these criteria and utilize the expressions to evaluate the possible performance of water resources systems. The authors use a water supply reservoir with a variety of operating policies to illustrate trade-off relationships among expected project benefits, reliability, resiliency,

and vulnerability. They indicate that by using such improved descriptions of possible nature of poor system performance the possible risks of various project alternatives could be better understood.

In a companion paper, Hashimoto et al. (1982b) define and use a robustness criterion in a water-supply system planning model. The authors define robustness as a measure of the likelihood that the actual cost of a proposed project will not exceed some fraction of the minimum cost of a system designed for the actual conditions that occur in the future. The model is applied to the planning of the expansion of a water supply system under demand uncertainty in southwestern Sweden.

In a series of four papers, Fiering (1982a, 1982b, 1982c, 1982d) explore the use of robust estimators and resilient ecosystem concepts from statistics and biology, respectively, in water resources systems modelling, planning, design, and operation. Detailed discussion of resiliency as applied to water resources and other systems is presented. The author gives several indices of resiliency. He develops a probability distribution of time to failure of a system, which is related to resiliency, by assuming a Markov behaviour of the passage of the system from one state to another, given a state dependent control policy. Simulation and canonical analyses are used to estimate resiliency. The capability of conventional performance indices of reservoir operation policy (e.g., storage index, detention time, capital cost, incremental reliability, and incremental storage index) to replicate information contained in resiliency indices is investigated. The author also examines the extent to which the conventional and resiliency indices can be estimated from basin geometry. Based on a hypothetical multi-reservoir system, he shows that the simulation approach for estimating resiliency gives imperfect, but encouraging, results and suggests areas of further research in mapping basin character-

istics and configuration into performance indices. The canonical analysis produces high canonical correlations, indicating that there is significant replication of information. The author develops a classification which suggests those basins and structural characteristics that would indicate less replication and consequently improved description of system performance.

2.3 Air quality management models

Models for the planning and management of air quality in pollution-prone industrialized areas have the same mathematical formulation as those for water quality management in streams. The traditional objective in air quality management is cost minimization of pollution abatement decisions and the typical constraints ensure acceptable air quality at pre-selected locations. The mathematical formulation of a deterministic model for the LC air quality management problem can often be represented as an LP since, in many cases, the cost curve for air pollutant removal is convex and can be linearized. Like stream quality management, the coefficients of the environmental quality constraints for air quality management (also referred to as transfer coefficients) are computed based on the environmental background conditions (e. g., pollution emission levels, wind speed, wind direction, and other meteorological variables) and a pollutant transport model.

Concerns about the randomness in the transfer coefficients, due to the variability of meteorological factors, has motivated research that incorporates the probabilistic nature of these coefficients in air quality management models (Ellis et al., 1985; 1986; and Fuessle et al., 1987). CCP and simulation are the most popular approaches used to include the uncertainty in the random transfer coefficients in these models. A brief

review of the research that considers meteorologic uncertainty in mathematical models for air quality planning and management is given here.

Ellis et al. (1985) extend a large scale deterministic optimization model for the selection of acid rain abatement strategies for Eastern North America to a probabilistic form by incorporating the uncertainty associated with the meteorological inputs in the pollutant transport model. The uncertainty in the meteorological inputs are reflected in lognormally distributed transfer coefficients for the constraints that limit wet sulfur deposition at pre-selected locations. The authors describe three methods of incorporating transfer coefficient uncertainty in an LP framework. These methods are Two-Stage Linear Programming Under Uncertainty (LPUU), CCP, and SLP. They develop a composite CCP-SLP model for the acid rain abatement problem. Two extreme categories of transfer coefficient correlation-covariance structure are analyzed: (1) complete colinearity which assumes complete dependence between all transfer coefficients and (2) complete noncolinearity which assumes complete independence between all transfer coefficients. The authors observe that the solutions from the complete colinearity-based model are so conservative that feasible abatement strategies can only be attained for wet sulfur deposition limits in excess of those considered to be environmentally desirable. For the complete noncolinearity case, they observe that there is no clearly defined system risk level associated with a given level of total system expenditure or vector of pollution abatement decisions. Therefore, they conclude that the solutions from the complete noncolinearity model may have limited utility to decision makers. Beta functions are used to develop upper and lower bounds for the relationships between system cost and the associated cumulative probability of exceedence for the complete noncolinearity solutions. This work suggests that the correlation structure of transfer

coefficients is very important when incorporating meteorologic variability in acid rain abatement modelling.

In a companion paper, Ellis et al. (1986) use a more realistic correlation-covariance structure, referred to as limited colinearity, for the transfer coefficients in the composite CCP-SLP model. The limited colinearity case lies between the extreme bounds of complete colinearity and noncolinearity. The authors utilize a solution procedure based on identification and delineation of meteorologic sub-systems so that individual sub-systems can be modeled as completely colinear or noncolinear as found appropriate. This approach retains the basic LP model formulation and is used to generate useful insights about the characteristics of the overall system performance. Although this approach is an improvement over those based on the two extreme correlation-covariance structures, the authors correctly state that the results may only serve in a didactic role.

Fuessle et al. (1987) develop a general CCP model for the selection of power plant sites, to minimize the cost of flue gas treatment and electrical transmission, and to satisfy pre-specified air quality standards. The model and solution approach use LP and simulation interactively to overcome the difficulties encountered when incorporating non-normally distributed and statistically dependent stochastic transfer coefficients in existing CCP efforts. This approach also reflects the confidence limits of distribution parameters. The mathematical formulation of the traditional nonlinear CCP model is simplified based on the assumption that the maximum and minimum electrical capacity at a site are linear functions of waste treatment efficiency. Thus, a linear interpolation is used to obtain a relation between the optimal treatment efficiency and two variables that compose the optimal electrical capacity at each potential power site. This interpolation is only valid if a constant sulfur dioxide emission per megawatt can be assumed

for different sizes of power plants. For other environmental planning or management problems, this assumption may be stated as: the waste produced per unit plant size for each polluter is constant. The assumption of a constant ratio for the waste produced to plant size may be spurious for many environmental management problems, e.g., the control of point source BOD discharges into a stream.

Ponnambalam and McBean (1990) use a CCP model of a hypothetical acid rain abatement program to show that the optimal total treatment cost (i.e., the optimization model output) is a nonlinear function of the significance level for the chance constraint and the variances and covariances among the uncertain input transfer coefficients. With this example, they show that correlation among input parameters should not be ignored in uncertainty analysis.

2.4 Groundwater quality management models

The mathematical representation of a stochastic hydraulic aquifer remediation system is similar to those of stochastic surface water quality and air quality management systems. For hydraulic aquifer remediation systems, however, hydraulic conductivity is often considered the most important uncertain input. Therefore, most studies in this area consider the spatial variability of the hydraulic conductivity field of an aquifer to be the only stochastic input.

Wagner and Gorelick (1987) explicitly incorporate parameter estimation into a model for the optimal design of an aquifer remediation scheme. They use first-order first- and second-moment analysis to characterize the parameter uncertainty in a non-linear CCP model. The chance constraints ensure that pollutant concentration does

not exceed a pre-assigned concentration level at each checkpoint within the aquifer with a specified minimum reliability. In the CCP model, the authors assume that the pollutant concentration at a checkpoint is a function of the uncertain parameters, and is normally distributed. The model is demonstrated for hypothetical steady state and transient stochastic aquifer reclamation design.

Wagner and Gorelick (1989) formulate a stochastic management model, referred to as a multiple realization model, for an aquifer remediation system with several reclamation wells. To improve the reliability of the solutions, the multiple realization model incorporates a number of possible scenarios of the stochastic hydraulic conductivity field in a single optimization model. The authors observe that the computational needs and time requirements limit the application of their model to 30 hydraulic conductivity realizations for a hypothetical example. In general, this could be a weakness of the practical application of the multiple realization model since a large number of vectors of input information may be required to adequately characterize the uncertainty in the inputs. Furthermore, their model does not indicate how trade-off relationships between management decisions and reliability could be obtained. However, environmental quality planners and managers often need such trade-off relationships for decision making.

Morgan et al. (1993) develop a Mixed-Integer-Chance-Constrained programming (MICCP) model for incorporating the uncertainty in all the LP constraint coefficients in a groundwater remediation problem. The approach is similar to the multiple realization method of Wagner and Gorelick (1989), except that it allows some realizations to be violated and solutions for estimated reliability levels less 100% to be obtained, hence increasing the computational complexity. It uses Monte Carlo simulation to include the uncertainty in the magnitude and the spatial distribution of the hydraulic conductivity

field. The authors use a solution technique which allows a certain proportion of the total number of Monte Carlo simulations to fail. The proportion of simulations allowed to fail is considered to be an estimate of the risk of not providing adequate remediation decisions. The MICCP model is demonstrated for a hypothetical aquifer and shown to be capable of incorporating uncertainties in both the right- and left-hand-side coefficients of the constraints of an optimization-based management model. However, the approach results in a large optimization model and requires large computational capacity and CPU time. In their application, the authors use only 100 Monte Carlo simulations to represent the stochastic hydraulic conductivity, due to the limitation in the size of the optimization model. The typical number of Monte Carlo simulations required for many real life problems may be in the thousands, thus precluding the use of the technique in its present form for several real life environmental management systems.

Ranjithan et al. (1993) present a feed-forward neural network approach for selecting a few important scenarios or realizations of the spatial hydraulic conductivity field to include in the MICCP model developed by Morgan et al (1993). The goal of the authors is to reduce the computational effort required to solve the MICCP model. The hydraulic conductivity values at each grid of a discretized hypothetical aquifer and the spatial mean and standard deviation of the log-hydraulic conductivity of a given realization are used as inputs of one training example for the neural network. The output, or target, is the total pumping rate required by the optimal pumping strategy for this realization. Thus, the total pumping rate for a given realization is used as a surrogate for its level of importance in the MICCP model. Starting with 100 unscreened realizations, the authors select 25 important realizations and develop a trade-off curve between pumping cost and reliability level for reliability values greater

than 80%. The results show that the neural network approach is computationally more efficient than the solution method used by Morgan et al. (1993). Compared with the original MICCP model results, the trade-off curve for Ranjithan et al. (1993) is accurate only for approximate reliability levels greater than 92%.

Recently, Ritzel et al. (1994) applied a genetic algorithm (GA) to solve a hypothetical stochastic groundwater containment problem adapted from Morgan et al (1993). The GA is essentially a search technique, structured after the biological processes of natural selection and evolution, for solving optimization problems. The authors use several GA techniques to obtain trade-off relationships between reliability and cost of the hydraulic containment problem. They compare the results obtained with those of the MICCP solution and observe that one of the GA techniques, the Pareto GA, nearly reproduces the trade-off curve from the MICCP method. The flexibility of the GA technique in allowing for fixed costs, nonlinearities, and discontinuities in the optimization model is demonstrated and discussed. However, the CPU time requirement for the GA technique does not indicate a higher computational efficiency than the MICCP approach for the case study demonstrated by the authors, and the capability of the GA to be implemented on a parallel computer should be explored.

2.5 Seasonal water quality management models

Traditionally, waste discharge management programs have been designed using constant design conditions for the entire year. These programs, referred to as non-seasonal programs, use some annual critical or 'near-critical' low flow, temperature, and other background water quality conditions as input data for the management model. The

design flow commonly used is the annual 7Q10 or a similar statistic. The annual 7Q10 is the 7-day averaged low flow which has a return period of 10 years. This implies that, if all other input information is assumed to be certain, then the designed management strategy based on the 7Q10 low flow has a 10% probability of at least one water quality violation event in any year. Typical design temperatures include the highest temperature on record, the highest monthly average temperature, the 90th percentile value of the recorded temperature, or some other projected worse-case situation. The other background water quality conditions are often represented by some typical (e.g., average) annual values of non-point source pollution contributions.

There has been a recent move to design seasonal waste discharge programs that utilize the seasonal changes in stream assimilative capacity and achieve water quality at decreased social and economic cost of pollution control. For example, in spring when flows are high and temperatures and decay rates are low, most non-ice covered streams have the capacity to assimilate a larger amount of BOD waste than they do in summer months when flows are low and temperatures are high. Designing different waste removal levels for the different seasons generally results in higher total waste discharge and lower waste treatment costs than those that result under non-seasonal waste discharge programs.

O'Neil (1980) uses an LP model to determine the optimal discharge policy for a hypothetical seasonal TDP management strategy for BOD discharges on the Lower Fox River in Wisconsin. A TDP program assigns a specified amount of waste discharge as a right to each polluter and allows sales or leasing of these rights among polluters. The results indicate that, for a stream DO standard of 6.2 mg/l, cost savings of approximately $23 million could be achieved by using the seasonal TDP program instead of a

non-seasonal UT policy. He uses four seasons and three scenarios of design flow and temperature combinations. He observes that as the number of seasons increases, the design of the management strategy becomes more complex. Under such situations, the administrative costs of pollution control may increase, while the reliability of meeting a water quality goal becomes less predictable.

Boner and Furland (1982) use a seasonal discharge program to study the management of BOD, nitrogenous wastes, and thermal discharges in a stretch of the Chattahoochee River near the City of Atlanta. This study shows that substantial cost savings over a non-seasonal program could be achieved by utilizing the dynamic assimilative capacity of the river. About 13% of the total year-round nitrification and secondary treatment cost, 10% of the total capital cost, and 23% of the user-charge cost (i.e., expansion debt service and coverage for local capital costs, and all additional operation and maintenance costs) could be saved if a seasonal program is implemented instead of a non-seasonal one.

Lence (1985) develops methods of analysis and assessment for Dynamic TDP programs. She identifies three TDP programs capable of exploiting varying water quality conditions for the control of BOD waste. These programs are Seasonal Static, Dynamic Feedforward, and Dynamic Feedback Permits. Seasonal Static Permits allow a constant rate of discharge during a time period in the year designated as a 'season'. Dynamic Feedforward Permits index the allowable discharge to observed stream conditions (e.g., flow and temperature) while the Dynamic Feedback Permits relate allowable discharge to the official water quality indicators which influence and are influenced by the pollutants (e.g., the available DO in streams). She raises questions regarding the combination of possible policy choices that would yield the highest cost efficiency, e.g.,

the type of control mechanism, and the location of monitoring devices (stream gauges or DO meters) that would produce the most efficient and reliable DO control system.

Kilgore (1985) and Eheart et al. (1987) show that a two-season TDP program could achieve a cost savings of up to 30% over a non-season TDP program for the Willamette River in Oregon. They indicate that the potential cost savings may depend on the selection of seasons and the corresponding durations. They show that an optimal summer season duration for which program cost approaches a minimum can be found.

Although there is no dispute about the superior cost efficiency of seasonal waste management programs compared to non-seasonal ones, there are concerns related to the accuracy of the computed probability of water quality violation, the increased amount of information requirements, the administrative complexity, and the operational difficulty of these programs. Nonetheless, it has been argued that the significant cost savings from these seasonal programs would justify continued investigation of this management approach (Kilgore, 1985; Lence, 1985; and Eheart et al., 1987).

The basis for comparison between the seasonal and non-seasonal programs is only valid if the risk of water quality violation is the same for both types of programs. The acclaimed cost efficient seasonal models allow a larger waste discharge into the stream and therefore are more likely to cause a violation of the water quality standard or goal. Risk of water quality violation, or certainty of model outcome, is one of the concerns about seasonal waste management programs that cannot be quantified in monetary terms. Risk Equivalent Seasonal Programs have therefore been developed for time varying programs (Rossman, 1989). These programs refer to risk as the probability of incurring a water quality violation, and are designed to maintain the same risk and

level of water quality standard as non-seasonal programs.

Rossman (1989) presents a waste discharge program that varies with the natural assimilative capacity of the receiving river body and maintains the same risk of water quality violation as non-seasonal water quality management programs for a single discharger case. He defines risk as the probability of incurring one or more water quality violations in a given year, and uses an NLP model to find the seasonal discharge limits that minimize total cost while keeping an acceptable risk of water quality violation. The model is applied to controlling ammonia toxicity, BOD, chlorine, and lead effluents in the Quinnipiac River in Connecticut and the Uncompahgre River in Colorado. In both river basins it is demonstrated that the risk-based waste load allocation is more cost efficient than the existing practice (i.e., a non-seasonal management program) for that system.

Two Risk Equivalent Seasonal Waste Discharge Programs for river basins with several dischargers are developed by Lence et al. (1990). They adopt the definition of risk from Rossman (1989). The two management programs have objectives of minimizing average uniform treatment and maximizing total waste discharge. BOD waste management for the Willamette River in Oregon is used for a case study, and it is demonstrated that both seasonal programs have significant cost savings compared with a non-seasonal model with the same risk of water quality violation. The authors investigate effects on the total waste treatment cost due to (1) the summer season length for a two-season program and (2) the DO standard. They show that the optimal length of the summer season ranges from two to five months depending on the water quality goal. Although a high cost efficiency is demonstrated for the Risk Equivalent seasonal programs, the authors express concern about a possible increase in environmental damage relative to

the corresponding non-seasonal program, and suggest that future research be directed towards developing new measures that account for environmental damage.

There is still some pessimism about the performance of the Risk Equivalent Seasonal Programs because of sparse or short historical water quality input information for many river basins. Seasonal programs are generally more complex and require more information than non-seasonal programs. For example, while year-round BOD waste discharge programs use design flows from low flow seasons and design temperatures from summer, seasonal models need design information for all seasons of the year. Some of this design information is either unavailable or unreliable for certain periods of the year (e.g., winter low flows). Therefore, there is a higher level of uncertainty in the input information for seasonal programs than for non-seasonal ones. Model output uncertainty is also expected to be higher for the former. To date, no work has explicitly accounted for uncertainty in the input information due to limited historical record for the design of risk equivalent seasonal waste management programs.

Lence and Takyi (1992) apply a modified regionalized sensitivity analysis (RSA) for assessing the effect of unreliable stream records on the design of a seasonal UT program that controls BOD discharges into the Willamette River. They consider flow and temperature in winter and summer as the uncertain input information. The RSA is conducted for waste management scenarios that have different season length combinations and DO goals. Their results indicate that the design of the seasonal program for the case study is generally more sensitive to uncertainty in summer flow and temperature data than to uncertainty in winter data. They also show that downstream flow data are more important than upstream data in low-flow periods. The authors further show that the degree to which uncertain stream conditions affect the seasonal management

model outcome depends on the water quality goals of the governing agency and the lengths of the seasons examined.

Chapter 3

Multiple Realization (MR) Model for Water Quality Management

3.1 Introduction

This chapter presents an improved SIOP approach, the Multiple Realization (MR) modelling method, that may be used to derive efficient solutions and accurate cost-reliability trade-off relationships for a water quality management model. Unlike the traditional SIOP approach which obtains solutions of a management model based on single realizations of stream conditions, the MR method incorporates several realizations simultaneously in an overall management model in order to produce efficient solutions.

The solution approach for the MR model begins by solving a single optimization model which incorporates all the realizations of background stream conditions that are used to represent the stochastic water quality system. Thus, the first solution corresponds to a reliability level of one. To solve the MR model at the next lower

reliability level, one of the realizations included in the optimization model is dropped and the optimization model is solved again. This process continues successively until the entire trade-off relationship is obtained. The set of realizations incorporated in the optimization model to obtain efficient solutions at given reliability levels cannot be selected at random. In general, at each stage of the solution process, the selected realizations must be chosen in order to obtain the best objective function value and the highest reliability level of the overall model solution.

In this chapter, the heuristic procedure by Morgan et al. (1993) for obtaining the realizations to include in the MR model at each reliability level is presented and demonstrated for the example surface water quality management problem based on the Willamette River. Furthermore, a number of approaches for selecting the appropriate realizations at a given stage of the solution process for a minimization problem have been developed in this research work. These approaches include: (1) an algorithm that selects a proportion of the total number of realizations with the highest objective function values based on the traditional SIOP solution; (2) an algorithm that selects a small number of the most critical realizations at each checkpoint; and (3) an algorithm that selects the important realizations based on the generalization capability of a feed-forward neural network.

The MR approach for water quality management utilizes the individual strengths of Monte Carlo simulation and optimization modelling. In an MR model, Monte Carlo simulation may easily be used to represent the statistical nature of complex equations or mathematical formulations in describing stochastic water quality systems. Thus, the Monte Carlo simulation normally permits a more detailed and realistic representation of the physical characteristics, background information, and the objectives of a

water quality management system than that allowed by optimization models. On the other hand, through the dual variables and sensitivity analysis, optimization models may provide important information about the physical river system, the water quality management objectives, and the alternative efficient ways by which desired objectives can be achieved.

The next section presents a general formulation of a basic deterministic water quality management model, which is used later in this chapter to develop the MR model. Then, examples of non-inferior realizations, that are responsible for inefficient solutions from the traditional SIOP approach, are demonstrated. The rest of the chapter gives a general MR model formulation, which overcomes the drawbacks of the traditional SIOP approach, when non-inferior realizations of background stream conditions are present for a given management system. The heuristic MR model solution approach due to Morgan et al. (1993) is also given and demonstrated for BOD wasteload allocation based on an example river basin.

3.2 General formulation for a water quality management model

Traditionally, water quality management problems have normally been represented mathematically as optimization models that minimize total waste treatment cost of point source polluters or a surrogate of this cost subject to constraints that ensure acceptable quality level of the receiving river. In some cases, other objectives, such as equity, are included in the model in the form of constraints that limit these objectives

to acceptable values. The optimization model formulation has typically taken the form:

$$\min f(\mathbf{X}) \tag{3.1}$$

subject to

$$\mathbf{A}(\mathbf{X}) \geq \mathbf{b} \tag{3.2}$$

$$\mathbf{r}(\mathbf{X}) \leq \mathbf{q} \tag{3.3}$$

$$\mathbf{X} \leq \mathbf{h_X} \tag{3.4}$$

$$\mathbf{X} \geq \mathbf{w_X} \tag{3.5}$$

where

$\mathbf{X}$ = vector of decision variables for water quality management, (e.g., the vector of percentage waste removal levels by point source polluters);

$f(.)$ = objective function (e.g., the cost function for waste treatment for all polluters);

$\mathbf{A}(.)$ = matrix of functions that characterize water quality improvement along the river;

$\mathbf{b}$ = vector of minimum allowable water quality improvement along the river;

$\mathbf{r}(.)$ = vector of secondary objective functions (e.g., equity);

$\mathbf{q}$ = vector of upper limits on the secondary objective functions;

$\mathbf{h_X}$ = vector of upper limits on the decision variables; and

$\mathbf{w_X}$ = vector of lower limits on the decision variables.

Generally, $f(\mathbf{X})$ is a nonlinear cost function of the pollution abatement decisions $\mathbf{X}$. However, this function is often approximated by a piecewise linear convex relationship for waste removal levels above primary treatment efficiency. $\mathbf{A}(\mathbf{X})$ represents the functions that relate pollution abatement decisions to water quality improvements along the stream. The matrix $\mathbf{A}$ is obtained, or developed, based on a pollutant transport simulation model, pollutant characteristics and discharge rates, and a given set of design stream conditions such as low flow, stream velocity, stream temperature, reaction rates, and the background water quality index. The elements of the vector $\mathbf{A}(\mathbf{X})$ may be linear or nonlinear with respect to the management decisions, depending on the pollutant simulation model used. For simulation models that are linear with respect to pollution management decisions (e.g., the Streeter-Phelps equation or the Camp-Dobbins modification of the Streeter-Phelps BOD-DO model), the elements of the matrix $\mathbf{A}$ are constants that typically represent the impact or transfer coefficients obtained by simulating the water quality improvement along the river per unit change in each management decision for a given set of design stream background conditions. The elements of $\mathbf{b}$ are also calculated based on the background conditions, the pollutant transport model, and the water quality standard. *Equation* 3.2, often known as the water quality constraint set, ensures that acceptable water quality levels are maintained along the river. The set of constraints represented by *Equation* 3.3 limit the other objectives of the model such as equity to pre-assigned values. *Equations* 3.4 and 3.5 constrain the management decisions to lie within maximum and minimum limits, respectively.

Type II uncertainty in the general water quality management model *Equations* 3.1 - 3.5 is in (1) the coefficients of the objective (cost) function, (2) the coefficients of the elements of the matrix $\mathbf{A}$, and (3) the vector $\mathbf{b}$. Most studies attempt to incorporate

only the uncertainties in **A** and **b**. For the traditional SIOP model, each realization of stream conditions comprises of a unique set of **A** and **b**.

It is often difficult to know or express the preferences of decision makers for water quality objectives, therefore trade-off relationships among water quality goals are normally generated to simplify the task of screening potential solutions and selecting the final pollution abatement strategy. The screening and final selection process use information from the trade-off relationships, further detailed or sophisticated water quality simulation, and subjective evaluation of the relative importance of the water quality management objectives to be achieved.

3.3 Non-inferior realizations of water quality response

The input data and parameters for multi-discharger water quality management systems are typically spatially distributed and optimal waste management solutions (e.g., the vector of waste removal levels) based on two different Monte Carlo realizations of the water quality conditions may be **non-inferior** to each other. Thus, it may be impossible to achieve acceptable water quality levels at all locations along the modelled stream under both realizations of design conditions, if the optimal management solution based on either realization of background conditions is implemented. On the other hand, if the solution to the management model based on Realization A can ensure acceptable water quality for stream background conditions represented by Realization B, but the reverse is not possible, then Realization B is said to be **inferior** to Realization A.

Non-inferior realizations may be common, for example, if the critical water quality responses (e.g., the lowest value(s) of the stream DO level) for different realizations of

water quality design conditions occur at locations that are far apart. Different realizations may have different critical stream locations that are far apart because critical combinations of the components of stream background conditions (e.g., low flow, stream velocity, temperature, reaction rates, and the background water quality index) may occur upstream for one realization but downstream for another realization. Non-inferior realizations of water quality conditions may also occur if there are multiple critical water quality locations for possible stream background conditions. In general, the presence of non-inferior realizations of stream background conditions is responsible for the inadequacy of the traditional SIOP approach for stochastic multi-objective water quality management modelling. The poor performance of the traditional SIOP approach under non-inferior stream conditions is due to the different critical locations that control the pollution abatement decisions for these stream conditions.

Two non-inferior realizations of water quality responses are illustrated using Problems P1 and P2 below. Here, the non-inferior water quality responses are used to represent the constraints in an LC water quality management model. The hypothetical river system simulated is derived from a simplified version of an example river basin. This hypothetical river system has two point source waste dischargers and three water quality checkpoints. The two realizations of water quality response are generated using Monte Carlo simulation. The mathematical formulations for Problems P1 and P2 are as follows:

Problem P1:

$$\min Z_c = 9.4Y_{11} + 13.5Y_{12} + 9.1Y_{21} + 12.6Y_{22} \tag{3.6}$$

subject to

$$0.4241X_1 \geq 0.1876 \tag{3.7}$$

$$0.4147X_1 + 0.4056X_2 \geq 0.4517 \tag{3.8}$$

$$0.6350X_1 + 1.3802X_2 \geq 1.3639 \tag{3.9}$$

$$X_1 - Y_{11} - Y_{12} = 0.35 \tag{3.10}$$

$$X_2 - Y_{21} - Y_{22} = 0.35 \tag{3.11}$$

$$Y_{11} \leq 0.33 \tag{3.12}$$

$$Y_{12} \leq 0.26 \tag{3.13}$$

$$Y_{21} \leq 0.34 \tag{3.14}$$

$$Y_{22} \leq 0.27 \tag{3.15}$$

Problem P2:

Objective function is the same as for **Problem P1** (*Equation* 3.6)

subject to

$$0.4373X_1 \geq 0.3374 \tag{3.16}$$

$$0.3367X_1 + 0.3225X_2 \geq 0.4229 \tag{3.17}$$

$$0.4465X_1 + 0.9469X_2 \geq 0.6570 \tag{3.18}$$

and *Equations* (3.10) through (3.15)

and all variables are non-negative.

where

$Z_c =$ total waste treatment cost (e.g., in \$ million);

X_i = waste removal level for discharger i (e.g., in %); and

Y_{ij} = waste removal for discharger i associated with treatment segment j of the piecewise linearized cost function (e.g., in %).

The range of possible waste removal levels for each discharger is divided into treatment segments in order to approximate the nonlinear cost functions with piecewise linear relationships. The objective function, *Equation* (3.6), minimizes the total cost of waste treatment (e.g., in $ million) and its coefficients are the unit cost coefficients associated with the linearized waste treatment segments of the dischargers' cost curves (e.g., in $ million/%). The only difference between Problems P1 and P2 are *Equations* 3.7 - 3.9 and *Equations* 3.16 - 3.18. *Equations* 3.7, 3.8, and 3.9 ensure that water quality standards are satisfied at Checkpoints 1, 2, and 3, respectively, based on one Monte Carlo realization of the water quality response for the first realization, while *Equations* 3.16, 3.17, and 3.18 ensure acceptable water quality at Checkpoints 1, 2, and 3, respectively, for a second realization. The coefficients for *Equations* 3.7 - 3.9 and for *Equations* 3.16 - 3.18 are based on water quality responses generated for Realizations 1 and 2, respectively, using the pollutant transport equation for this river system. The left hand side coefficients for each of these equations (*Equations* 3.7 - 3.9 and 3.16 - 3.18) represent the water quality improvement per percentage waste removal (e.g., in $mg/l/\%$) and the right hand side coefficients represent the minimum water quality improvement required (e.g., in mg/l) to maintain acceptable water quality standards for the corresponding checkpoint and realization. *Equations* 3.10 and 3.11 define the waste treatment levels for Dischargers 1 and 2, respectively, as the sum of the treatment segments for each of the waste dischargers. These equations also specify a 35% minimum

level of waste treatment for each discharger. *Equations* 3.12 and 3.13 define the upper limits for treatment segments for Discharger 1, while *Equations* 3.14 and 3.15 define those for Discharger 2.

Solutions for Problems P1 and P2 are given in Rows 1 and 2 of Table 3.1. The first and second columns of this table give the name and the minimum total waste treatment cost, respectively, of the problems analyzed. Columns 3 and 4 contain the optimal waste treatment levels for Dischargers 1 and 2, respectively. The next three columns contain the water quality indeces for Checkpoints 1, 2, and 3, respectively, for Realization 1, while Columns 8, 9, and 10 represent the water quality indeces for Checkpoints 1, 2, and 3, respectively, for Realization 2. Positive water quality index values indicate that an acceptable water quality is maintained at the corresponding checkpoint for the given realization. The value of a positive index equals the excess water quality over the specified standard (e.g., in mg/l) at the corresponding checkpoint. On the other hand, a negative index implies a water quality violation and the magnitude of this index is the shortfall in the water quality standard at that checkpoint for the corresponding realization. A zero index means the water quality standard is exactly achieved at that checkpoint for the corresponding realization.

It can be seen from Table 3.1 that the solutions of Problems P1 and P2 achieve acceptable water quality (i.e., have non-negative indeces) for Realizations 1 and 2, respectively, at all checkpoints. However, the solution of Problem P1 fails to maintain the water quality standard at Checkpoints 1 and 2 for Realization 2, while that of Problem P2 also fails to maintain the standard at Checkpoint 3 for Realization 1. In this example, the critical water quality locations for Realization 1 occur at Checkpoints 1 and 3, while those for Realization 2 occur at Checkpoints 1 and 2. The optimal solution

of either realization cannot be used to manage this hypothetical system successfully, if a water quality violation is not acceptable for both realizations. Realizations 1 and 2 are therefore non-inferior to each other under the LC management model formulated for this hypothetical water quality system. The higher cost of waste treatment for Problem P2 is no indication that its solution can maintain acceptable water quality for Realization 1. A solution that will achieve the water quality goal at all checkpoints for both realizations can be obtained by using a mathematical model formulation consisting of *Equations* 3.6 through 3.18. This mathematical model is referred to as the Combined model in Table 3.1. The third row of Table 3.1 shows that the solution for the Combined model has a higher total cost than that for either Problem P1 or P2, and it maintains acceptable water quality at all checkpoints for both Realizations 1 and 2.

Non-inferior realizations of water quality response may not exist for (1) a river system where the critical water quality section does not change drastically under different sets of potential background conditions and (2) a management model which has only one decision variable, e.g., a model based on a non-seasonal UT program. Under such conditions, traditional SIOP approaches for obtaining a trade-off between the probability of water quality violation and the management objectives, or the vector of decision variables may be adequate.

If there are non-inferior sets of background stream conditions, it would be necessary to use the MR approach which includes several possible realizations of water quality design conditions simultaneously in an overall management model to achieve a solution that maintains the required objectives for all conditions. The fundamental difference between the traditional SIOP approach and the MR approach is that, while the former solves the management model for a single realization at a given reliability level, the

latter includes a set of realizations that depend on the reliability level of interest in a single optimization model.

For a given management system, the MR approach generates several realizations of sets of water quality response functions at pre-selected checkpoints along the river via a pollutant transport model and Monte Carlo simulation of stream background conditions. Water quality constraints for these checkpoints, based on all of the realizations of water quality response functions, are included in a single overall management model. The optimal solution of this management model is used to determine the most critical realization of the sets of water quality responses included in the model. The constraint set representing the most critical realization of stream background conditions is then removed from the overall management model and the solution of this reduced model is determined. To generate the entire trade-off relationship between reliability and the objective function or the vector of management decisions, this process continues successively until the solution corresponding to the least critical water quality realization is obtained. Each successive solution of the overall management model has a corresponding probability of water quality violation which is equal to the sum of the probability of occurrence of all sets of water quality conditions represented by the constraints that have been removed.

3.4 General MR model formulation and solution approach

This section presents the formulation for the MR model for incorporating uncertainty in a water quality management program. The inadequacy of the traditional SIOP approach to generate good trade-offs between water quality objectives (e.g., total waste treatment cost, reliability, resiliency, vulnerability, and equity) is addressed using a single optimization model that incorporates all possible water quality scenarios and ensures acceptable standards at all checkpoints. Developing such trade-offs with the MR model requires the identification of realizations of water quality background conditions that must not be violated, and those that are allowed to be violated, at a given reliability level. The general management model for typical water quality objectives is given below:

$$\min f(X) \tag{3.19}$$

subject to

$$P_{ij}(\mathbf{X}) \geq b_{ij} \quad \forall \quad i = 1, 2, \ldots, I; \quad j \in \Phi_R \tag{3.20}$$

$$r_l(\mathbf{X}) \leq q_l \quad \forall \quad l = 1, 2, \ldots, L \tag{3.21}$$

$$\mathbf{X} \leq \mathbf{h_X} \tag{3.22}$$

$$\mathbf{X} \geq \mathbf{w_X} \tag{3.23}$$

where

$P_{ij}(\mathbf{X})$ = a function that characterizes water quality improvement at checkpoint i for realization j (e.g., mg/l);

b_{ij} = minimum allowable water quality improvement at Checkpoint i for Realization j (e.g., mg/l);

$r_l(\mathbf{X})$ = the l th secondary objective function (e.g., equity);

q_l = the upper limit on the l th secondary objective function;

$\mathbf{h_X}$ = a vector of upper limits on the decision variables;

$\mathbf{w_X}$ = a vector of lower limits on the decision variables;

I = the total number of water quality checkpoints;

L = the total number of secondary objectives;

Φ_R = the set of realizations of water quality responses that would result in the best objective function value and a reliability equal to R; and

all other variables are already defined.

Equation 3.19 minimizes an objective of the management model, e.g., the total cost of waste treatment. The sense of the objective function may also be a maximization, e.g., if $f(\mathbf{X})$ defines the total waste discharged into the river system or the reserve capacity at the most critical water quality location along the river. *Equation* 3.20 is the water quality constraint set. It ensures that acceptable water quality is maintained at all checkpoints for all realizations in the set Φ_R. The reliability, R, or the probability of non-violation of water quality may be computed as:

$$R = \frac{n(\Phi_R)}{S} \tag{3.24}$$

where

$n(\Phi_R)$ = number of Monte Carlo realizations in the set Φ_R (i.e., the number of realizations for which water quality must be maintained at all

checkpoints to obtain the best objective function value and a reliability equal to R); and

S = total number of Monte Carlo realizations included in the MR model.

$P_{ij}(\mathbf{X})$ represents the function that relates pollution abatement decisions to water quality improvements at Checkpoint i for Realization j. The right hand side of *Equation* 3.20, b_{ij}, is the required water quality improvement at Checkpoint i for Realization j. The set of constraints represented by *Equation* 3.21 limit the secondary objectives of the model such as equity to pre-assigned values. *Equations* 3.22 and 3.23 constrain the management decisions to lie within maximum and minimum limits, respectively.

The first task involved in solving the MR model, *Equations* 3.19 - 3.23, is to identify all members of the set Φ_R. After Φ_R has been identified, the appropriate optimization algorithm may be used to solve the model. However, obtaining this set is not trivial. A successive heuristic algorithm for obtaining Φ_R and solving the MR model is presented by Morgan et al. (1993) for an aquifer remediation design which is formulated as a MICCP problem. This heuristic algorithm is described in Section 3.4.2 in the context of stream quality management modelling.

3.4.1 Determination of the elements of Φ_α

The determination of members in the set Φ_R is a successive procedure. The goal is to determine the most critical $100(1-R)\%$ realizations of water quality response (i.e., elements in the set Φ_R') and to eliminate them from the management model. At the first stage of this successive procedure, the water quality constraints for all realizations of the Monte Carlo simulation are included in a single optimization model and solved

using an appropriate algorithm. The solution obtained represents the case in which no water quality violation is allowed for any realization of the total number of Monte Carlo simulations. At the second stage, the most critical water quality realization at the initial stage, $\Re_c^1$, is determined and dropped from the model. The resulting (reduced) model is then solved. The most critical realization for the reduced model, $\Re_c^2$, is then dropped from this model and the resulting problem is solved for the third stage. This process continues until the proportion of the total number of realizations removed from the original optimization model is $(1-R)$. At this stage, the remaining realizations for the reduced model constitute the set Φ_R.

The order in which these S realizations are removed from the MR model to obtain the trade-off relationships gives a level of criticalness of the stream background scenarios included in the model. Therefore, the realization dropped at Stage m of the successive solution process, $\Re_c^m$, has a level of criticalness equal to m. In this thesis, the level of criticalness of Realization j is denoted as U_j. This implies that the lower the value of U_j, the more important or critical Realization j is.

Mathematically, Φ_R is defined as:

$$\Phi_R \;=\; \{G_j : E_j = 0 \quad \forall \;\; j = 1, \cdots, S\} \tag{3.25}$$

where

G_j = index for Realization j;

E_j = 1 if Realization j is dropped from the MR model before stage $(1-R)S$,
0 otherwise.

The solution of the reduced optimization model that incorporates the realizations in the set Φ_R will result in the best objective function value when the reliability is R or

the probability of water quality violation is equal to $(1-R)$. Thus, if S realizations are utilized and the entire trade-off relationship between reliability and objective function, or the vector of decision variables, is desired, the MR model must be solved at S stages.

The most critical realization at stage m, $\Re_c^m$, may be identified based on some measure of criticalness for each realization or on a heuristic procedure (see, e.g., Morgan, 1990 and Morgan et al., 1993). One approach developed in this research obtains the measure of criticalness for a given realization based on the solution (e.g., objective function, decision and dual variables, and reliability) of the management model using the design inputs for this realization only. This approach, which is based on neural networks, is presented and demonstrated in Chapter 5.

3.4.2 The heuristic algorithm of Morgan et al. (1993)

At a given stage of the successive process, the heuristic approach of Morgan et al. (1993) considers each realization that has a binding constraint (i.e., binding realizations) after solving the overall model, as a candidate for the most critical realization. These candidates are alternatively dropped from the overall model and the realization whose elimination produces the largest decrease in cost for an LC problem, when the overall model is solved, is taken as the most critical realization. Thus, for the heuristic approach, the optimization model may be solved several times at each stage of the successive process.

The heuristic procedure described above uses the optimal solution at each stage to create the optimization model for the following stage. This is done by removing all constraints associated with each binding realization of the optimal solution at a given

stage in turn with replacement, to create candidate optimization problems that may be solved to obtain the best objective function value at the following stage. Global optimality at all stages of the solution process is guaranteed if all candidate solutions, instead of the optimal solution, at a given stage are used to create the optimization problems at the following stage. For details of this procedure, interested readers are referred to Morgan (1990) and Morgan et al. (1993). Morgan (1990) shows that for each reliability level the objective function value obtained by using only the optimal solution at one stage to create the problem at the next stage is within 1% of that generated by using the five best candidate solutions at a given level to generate the model for the following level. Furthermore, he observes that using the optimal solution instead of the best five candidate solutions at one stage to create the problem for the next stage reduces the computational time by about 75% for a hypothetical groundwater remediation design that utilizes 100 realizations.

3.4.3 Solving the reduced model

The problems created at each stage of the successive procedure may be solved using any appropriate optimization technique. During the solution process, the reliability R, is decreased from one to $\frac{1}{S}$ at intervals of $\frac{1}{S}$. Thus, the successive heuristic approach described in Section 3.4.2 is used to produce solutions (i.e., values of the objective function and the corresponding management decisions variables) from a reliability level of one to a value equal to $\frac{1}{S}$, or nearly zero.

3.5 Application of the MR model to the Willamette River basin

3.5.1 Description of the water quality system

The MR model and the heuristic solution approach are demonstrated for a non-seasonal control of point source BOD waste discharges in the Willamette River basin. The segment of the Middle Fork of the Willamette River for which the model is applied is 298 km long and has eight major tributaries and ten BOD waste dischargers. Cost data (in 1978 dollars) and wasteload characteristics of the dischargers are obtained from Kilgore (1985). The location and wasteload characteristics of the dischargers are shown in Table 3.2. Column 1 of this table shows the names of the ten major BOD waste dischargers on the Willamete River. The second column summarizes the flow gaging station names and the numbers used to represent these stations. The locations of the waste dischargers and the flow gaging stations relative to Springfield are given in Column 3. Columns 4 and 5 indicate the discharge rate and influent BOD waste concentration, respectively, of the point sources of pollution. The Camp-Dobbins modification to the Streeter-Phelps equation (Camp, 1963; and Dobbins, 1964) is used to simulate DO levels along the river. However, more sophisticated water quality simulation models, such as QUAL2E and WASP5, may be used to simulate water quality level in the MR model. River flow and temperature data are acquired from the United States Geological Survey (USGS). Velocity and reaeration rates are functions of flow and are taken from Worley (1963) and Liebman (1965). Benthic oxygen demand is assumed to be zero and a background DO deficit of 1.0 mg/l is adopted for the Willamette and its tributaries. These assumptions

are consistent with those of several water quality management modelling case studies based on the Willamette River (see, e.g., Liebman and Lynn, 1966; Eheart, 1980; Brill et al.,1984; Kilgore, 1985; Eheart et al., 1987; Lence et al., 1990; and Burn and Lence, 1992).

The 7-day average low flow for Springfield, Harrisburg, Albany, Salem, and Portland, and the highest mean monthly temperature for Harrisburg for the months of June through September are used for this analysis. These months are assumed to represent the summer season and are therefore used to analyze the non-seasonal program. Since the Coast Fork tributary joins the Middle Fork of the river upstream of the segment of interest, and of Springfield, Oregon, the flow records for the Middle Fork and the Coast Fork gaging stations are added and assigned to what is referred to as Station 1, Springfield, flows.

The stretch of the Willamette River modelled is divided into 18 reaches. Thirty-five water quality checkpoints are defined. The stochastic input information considered in this study are the 7-day average low flow at each of the five flow stations, the highest monthly mean temperature at Harrisburg, and the stream velocity and reaction rates for each reach of the river, for the given streamflow conditions.

3.5.2 The management model

An LC program for controlling BOD waste discharges is used to demonstrate the application of the MR model. Monte Carlo simulation, described in Section 3.5.3, is used to generate the multiple realizations. The LC program is an LP model with an objective function, *Equation* 3.19, that minimizes the cost of waste treatment, and decision

variables that are the vector of waste removal levels of the point source dischargers. For the example river basin used, a minimum level of waste removal corresponding to primary treatment (35% removal) is specified for all dischargers. The indicator of water quality is the stream DO, hence, the water quality constraint set, *Equation* 3.20, ensures acceptable DO levels at all checkpoints. Different DO standards are specified for two different sections of the river based on water quality profiles obtained from a preliminary analysis. These standards are 7.10 mg/l DO for the final 82-km downstream segment which has a 48-km estuarine section, and 8.20 mg/l for the remaining upstream segment of the river. Liebman and Lynn (1966) assigned DO standards for different sections of the simplified representation of the Willamette River used in their study. However, the standards used by them were approximately equal to those used by the Oregon State Sanitary Authority to manage the Willamette River during the time of their study. Using different water quality standards for different segments of a river may be required (1) to achieve water quality objectives for different localities situated along the river, or (2) to preserve different aquatic species at different sections of the river.

The function $P_{ij}(\mathbf{X})$ on the left hand side of *Equation* 3.20 relates the DO improvement at Checkpoint i to the waste removal levels at all point source waste discharge locations for Realization j. For the Camp-Dobbins modification to the Streeter-Phelps equation used in this work, $P_{ij}(\mathbf{X})$ is a linear function of $\mathbf{X}$ and its coefficients are the impact coefficients obtained from a simulation of the water quality response. The right hand side coefficient of *Equation* 3.20, b_{ij}, is the minimum allowable DO level improvement required at Checkpoint i and for Realization j. These coefficients are also obtained using the Camp-Dobbins BOD-DO simulation model. The LC model has

no secondary objectives, therefore *Equation* 3.21 is eliminated from this model. The minimum and maximum levels of waste treatment for each discharger are fixed at 35 and 98%, respectively.

3.5.3 Monte Carlo simulation of the water quality response

Monte Carlo simulation is used to generate the possible combinations of the uncertain input data that reflect the uncertainties in the input information for the overall management model described in *Equations* 3.19 -3.23. The adequate number of realizations to generate depends on the type and complexity of the water quality system and on the number of uncertain input data and their distributions. This adequate number of realizations is often obtained by gradually increasing the number of simulations until the statistical properties of the input and output information for the simulations converge for an experimental trial. For demonstrating the MR model using the heuristic solution algorithm presented by Morgan et al (1993), only 100 Monte Carlo realizations are utilized due to a limitation dictated by the computational time required.

The 7-day average low flow data for each of the five flow stations and the highest mean monthly temperature for Harrisburg fit the two-parameter lognormal distribution. To obtain the same correlation structure as that for historical flows at the five stations and to reduce computation time, the Monte Carlo simulations use the 7-day averaged low flow at Albany as a basis for generating flows at the other four stations. The Albany station is used as a base station because its flows are the best predictor of the other station flows (see, Takyi, 1991). For each Monte Carlo simulation, a random stream temperature and 7-day averaged low flow at Albany are generated based

on the two-parameter lognormal distributions for the mean monthly temperature at Harrisburg and for the 7-day averaged low flow at Albany, respectively. Then, the flows at the other four stations are generated on the basis of their regression relationship to the flow at Albany and on normally distributed zero mean noise terms associated with a new observation for the dependent variable of a regression model (see, Neter et al., 1989). The stream velocity and reaeration rates at (20°C) for each reach of the river are subsequently computed based on the generated flows at the five stations, on the functional relationship between flow and stream velocity, and between flow and reaeration rates at 20°C (developed by Worley, 1963 and Liebman, 1965), respectively, and on normally distributed zero mean noise terms whose variances are given as:

$$V_v = aQ^b \tag{3.26}$$

$$V_r = cQ^d \tag{3.27}$$

where

V_v	=	variance of the noise terms for stream velocity;
V_r	=	variance of the noise terms for reaeration rates at 20°C;
Q	=	flow at the reach; and
a, b, c, d	=	coefficients.

For each simulation, the generated temperature, flows, velocities, and reaeration rates of all reaches are used to generate a set of steady state DO response, based on Camp-Dobbins modification to the Streeter-Phelps equation (Camp, 1963; and Dobbins, 1964). This realization of DO response consists of DO improvements, per unit waste removed (i.e., impact coefficients) by each discharger, for each checkpoint, and the minimum required DO improvement at all checkpoints. Thus, each Monte Carlo

simulation produces 350 impact coefficients (i.e., 10 dischargers $\times$ 35 checkpoints) and 35 minimum required DO improvements. Therefore, each realization is represented by 35 water quality constraints.

3.6 Discussion of results

Figure 3.1 is the graph of the total cost of waste treatment, above primary treatment, versus reliability R, for the MR modelling approach using 100 realizations. Also plotted are the results for the traditional SIOP approach that is based on single realizations of these 100 stream background conditions. For the traditional SIOP results, the probability of DO standard violation for a given solution is equal to the proportion of the total number of realizations which fail to achieve the standard at one or more checkpoints when the vector of waste treatment levels required by this solution is invoked. Figure 3.1 shows that most of the solutions based on single realizations are inefficient compared with those based on the MR model. For example, the cost of waste treatment and the corresponding reliability of the LC program solutions based on Realization A alone are \$5.69 million/year and 0.91, respectively. However, for the same cost of waste treatment (\$5.69 million/year), the solution for the MR model maintains a reliability equal to 0.99. Also, the solution for Realization C maintains a 0.77 reliability and requires \$4.74 million/year for waste treatment, while, at this reliability level, the MR model solution requires only \$1.44 million/year for waste removal. For these 100 simulations, the highest reliability attained is 0.91 for the traditional SIOP model and 1.0 for the MR model. Figure 3.1 also shows that the solutions for some of the realizations are as efficient as those for the MR model. However, these realizations that produce efficient

SIOP solutions may be identified only after comparing their solutions with the MR solution.

Any point corresponding to a traditional SIOP solution and lying above or to the right of the curve for the MR model represents an inefficient solution. The distance between a point representing any of the traditional SIOP solutions and another point on the curve for the MR model (having the same cost or reliability as the traditional SIOP solution) gives an indication of the degree of inefficiency of the traditional SIOP solution. The larger this distance, the more inefficient the traditional SIOP solution is. Thus, for the example presented, a larger number of inefficient solutions occur at high levels of reliability (greater than 0.70) than at low levels. Typical threshold values of the reliability level used by decision makers operating in the public domain in managing water quality in rivers are often greater than 0.80. This indicates that information obtained about a water quality system based on the traditional SIOP approach may be unreliable for planning or managing the system.

Figure 3.1 also shows that traditional SIOP solutions with reliability levels and waste treatment costs less than 0.70 and \$1.30 million/year, respectively, are generally close to the MR solution. These realizations represent average stream background conditions. Unlike critical conditions, these average conditions are normally characterized by little spatial and temporal variation.

Although the optimal LC solution based on Realization C is over \$2.00 million/year more than the solution based on Realization D, the former solution is 77% reliable while the latter is only 83% reliable. These two realizations are clearly non-inferior to each other, and show that the total waste treatment cost (i.e., the value of the

objective function) based on a single realization may not represent the best attainable reliability that corresponds with the objective function value. In fact, the scatter of the points representing traditional SIOP solutions suggests the absence of a simple function relating reliability and cost for this approach.

Like all other mathematical modelling methods based on Monte Carlo simulations, the efficacy of the results of the MR model can be guaranteed only if a sufficient number of realizations are utilized. If the number of realizations used is inadequate, the statistical properties of the output from the model will not converge, therefore results and conclusions made after analyzing the management model may not be generalized for the system. Thus, an important criterion of the Monte Carlo simulation approach is that the distribution functions for the input and output data samples are similar to the respective 'unknown parent distributions'. This similarity between sample and parent distributions is a function of the number of realizations utilized which depends on (1) the type and complexity of the model and output distribution; (2) the number of uncertain input data; and (3) the distributions or ranges of the uncertain input data. To determine the appropriate number of realizations to use, normally, an experiment is conducted that requires the number of realizations to be increased until statistically stable output results are obtained.

Morgan (1990) discusses a bionomial distribution method for obtaining the confidence limits of reliability values as a function of the total number of realizations utilized in the MR model and the number of realizations for which the constraints of the model are met. This method may be used to estimate the appropriate number of realization to utilize by ensuring, for example, that the 95% confidence interval for reliability levels of interest is small. However, generally this method may be inadeaquate, since it ignores

the type of system modelled, the complexity of the model used, and the number and characteristics (e.g., distribution and range) of the uncertain input data. Furthermore, the method does not give any indication of the sensitivity of the reliability level to the optimal objective function value. For some systems, even a small change in reliability may result in a large change in objective function value. Therefore, often the confidence interval of reliability levels alone will not conclude that convergence of the MR solution will be obtained with a given number of realizations.

The number of realizations used to generate the MR model solution in Figure 3.1 was constrained by computational time requirements. Therefore, to test the adequacy of the MR solution with the heuristic algorithm of Morgan et al. (1993), four cost-reliability trade-off relationships, each based on a different set of 100 realizations, are developed and compared in Figure 3.2. The four sets of 100 realizations are referred to as MCS_1, MCS_2, MCS_3, and MCS_4. Figure 3.2 clearly shows that each of the trade-off curves is different from the others, especially at reliability levels greater than 0.7. However, this range of reliability values corresponds to typical reliability levels used by decision makers to derive pollution abatement policies. Furthermore, these high reliability levels and other risk indices, such as the duration (i.e., resiliency) and magnitude (i.e., vulnerability) of water quality violations, determine the potential damage of pollution to aquatic ecosystems and the extent of potential limitation of the beneficial uses of surface water bodies from water quality considerations.

For 100 realizations, the bionomial distribution method, discussed by Morgan (1990), estimates the 95% lower and upper confidence limits at 0.95 reliability level as 0.88 and 0.98, respectively. This confidence interval appears to be small, however, from Figure 3.2, the total waste treatment cost changes significantly within this range (i.e.,

0.88 to 0.98) of reliability levels for each of the four cost-reliability relationships and between the trade-off curves. When compared to one another at all reliability levels, the total waste treatment costs are significantly higher for some of the four sets of 100 realizations than for others. For example, at a reliability level of 0.92, MCS_1 and MCS_3 produce waste treatment costs of $5.04 million/year and $1.48 million/year, respectively. The difference in waste treatment cost for these two sets at this reliability level is $3.56 million/year, which represents an approximate increase of 240% relative to the optimal objective function value for MCS_3. The differences in the trade-off relationships suggest that 100 realizations may not be enough to represent the stochastic water quality management system, especially at the high cost and reliability end of the cost-reliability relationship.

Figure 3.3 shows a plot of the normalized CPU time requirements of the MR model with the heuristic solution algorithm of Morgan et al. (1993) versus number of realizations (up to 100) utilized. The CPU time required to solve the MR model for a given number of realizations is normalized by dividing it by the CPU time utilized for 100 realizations. Figure 3.3 indicates that the CPU time increases slowly as the number of realization is increased from 1 to 40. However, as the number of realizations increases beyond 40, the CPU time increases quickly. In fact, the rate of increase is exponential in nature. From the graph, the time required to solve the MR model for 50 realizations is less than 4% of that required for 100 realizations.

The CPU time used to generate the entire trade-off relationship for 100 realizations is approximately 7 days on the University of Manitoba Sparc2 UNIX stations. Therefore, if the graph in Figure 3.3 is extrapolated for 1000 realizations the estimated CPU time requirement is on the order of several weeks. However, in a typical water quality

management system, an adequate number of realizations is expected to be on the order of thousands.

Therefore, in spite of the superiority of the MR approach over the traditional SIOP method in producing efficient solutions, the computational requirements are significant. Its large computational burden may make the MR method combined with Morgan et al.'s (1993) heuristic solution approach unattractive to water quality management modelers and design engineers, unless parallel processing is implemented or efficient solution techniques are developed. Nevertheless, there is a great potential for reducing this large computational burden. Chapters 4 and 5 present approaches for reducing the CPU time for solving the MR model without sacrificing accuracy of output.

Table 3.1: Solutions for Problems P1 and P2 and the Combined Model

Problem	Total Cost, Z_c ($ million)	Removal Level (%)		Water Quality Index					
				Realization 1 Checkpoints			Realization 2 Checkpoints		
		X_1	X_2	1	2	3	1	2	3
P1	5.1550	44.2	78.5	0.000	0.050	0.000	-0.144	-0.021	0.284
P2	5.7557	77.2	50.6	0.140	0.073	-0.176	0.000	0.000	0.166
Combined	6.9152	77.2	63.3	0.140	0.125	0.000	0.000	0.041	0.287

Table 3.2: Discharger Waste Load Characteristics and Gaging Station Locations

Discharger	Gaging Station, (Station No.)	Location, km	Design Flow, $MGD(l/sec)$	Influent BOD, mg/l
Springfield	Springfield (1)	0.00	6.86 (301.0)	248
	Harrisburg (2)	10.13		
Eugene		10.20	17.16 (752.6)	284
American Can Co.		57.80	5.36 (235.1)	101
Evans Products		79.40	4.14 (181.6)	118
Corvallis		85.80	7.13 (312.6)	121
Albany		105.20	8.71 (382.1)	93
	Albany (3)	108.00		
Western Kraft		108.30	4.23 (185.5)	240
	Salem (4)	160.00		
Boise Cascade		161.20	15.72 (689.5)	130
Salem		163.10	17.69 (775.8)	408
Crown Zellerbach	Portland (5)	254.70	9.47 (415.4)	279

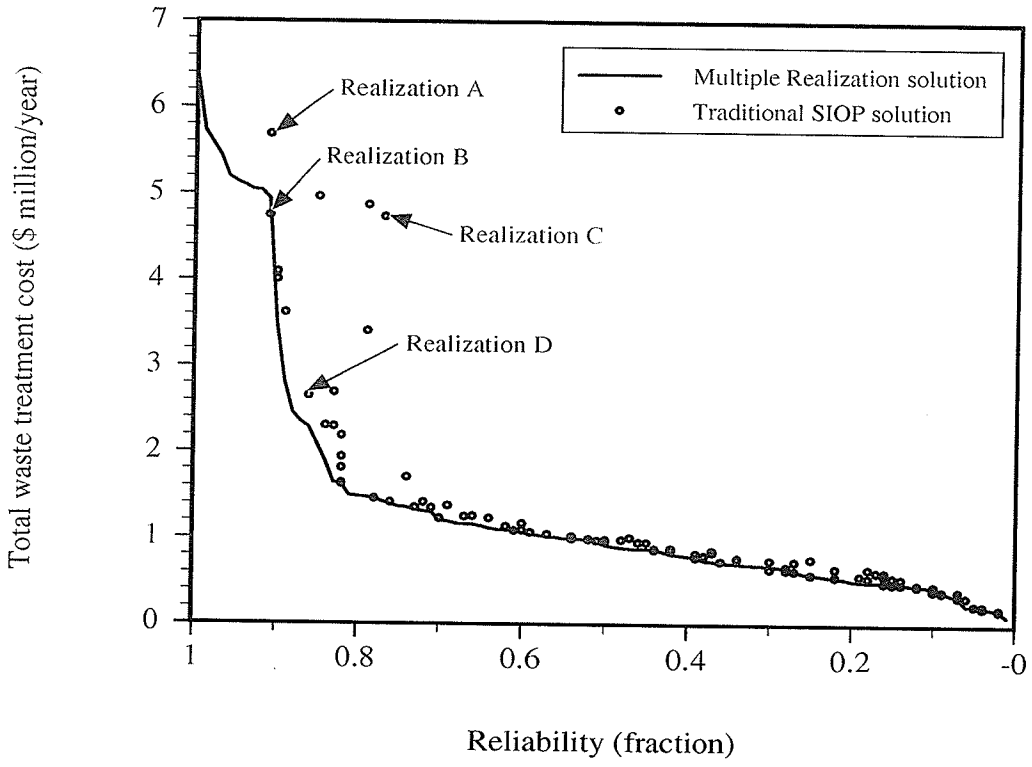


Figure 3.1: Trade-off between cost and reliability

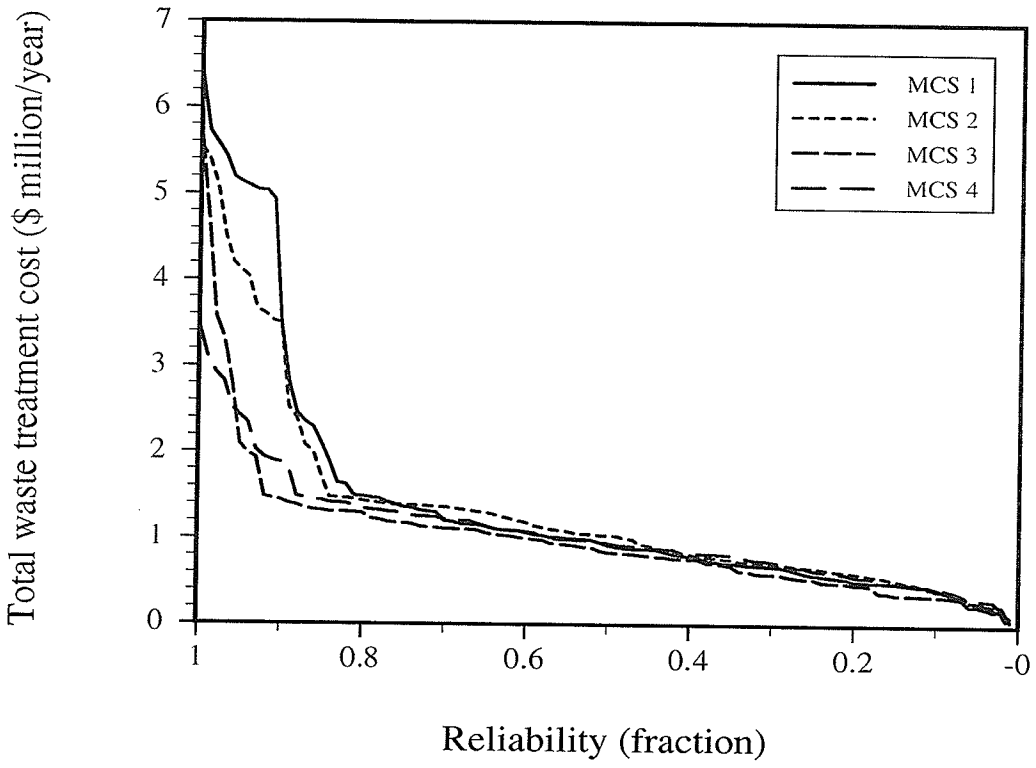


Figure 3.2: Trade-off curves for four sets of 100 realizations each

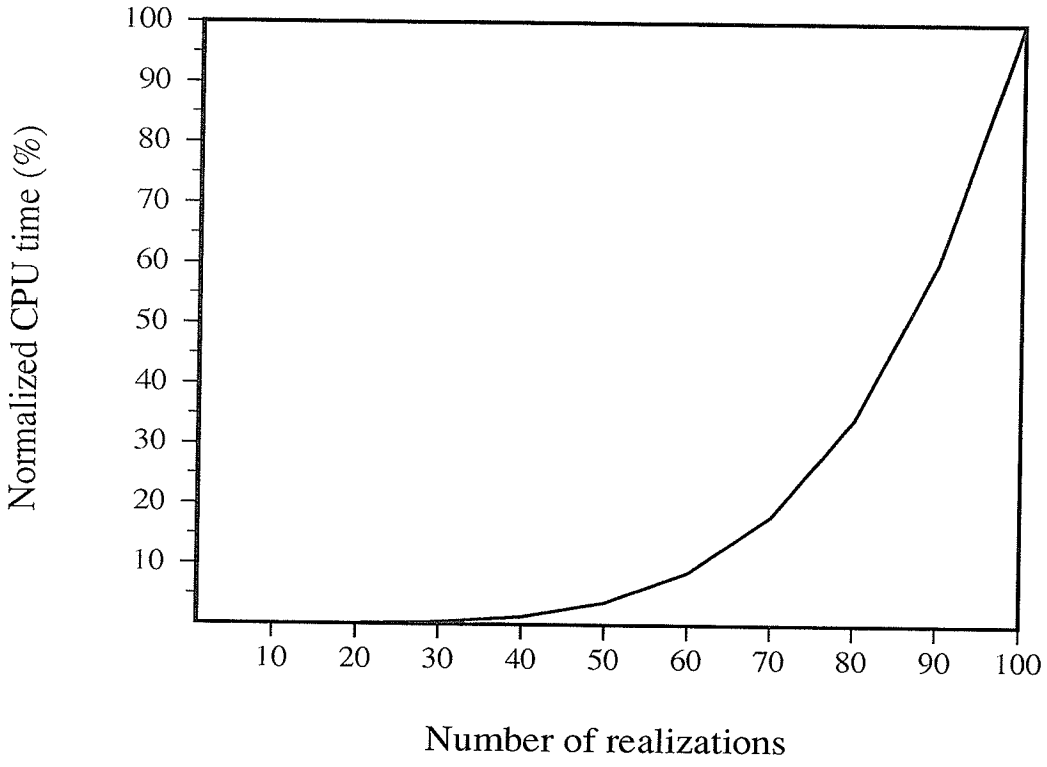


Figure 3.3: CPU time requirements for solving the MR model using Morgan et al.'s heuristic algorithm

Chapter 4

New Algorithms for Solving the MR Model

4.1 Introduction

The ability of the MR model to produce efficient solutions is demonstrated in Chapter 3. It is also shown that large computational resources are needed to obtain these efficient solutions. In this chapter, the successive heuristic algorithm used to solve the MR model is improved to reduce computational time required and to allow a large number of realizations to be incorporated in the stochastic management model.

The large CPU time requirement of the heuristic algorithm developed by Morgan et al. (1993) is due to the large size of the optimization problem created by the MR model and the successive process which typically solves the model more than once at each level of reliability. Therefore, the computational resources required may be decreased by using techniques that reduce the size of the optimization model or reduce

the number of times the optimization model is solved with Morgan et al.'s heuristic solution process. To reduce the size of the optimization models solved at each stage of the heuristic algorithm, three new algorithms are developed in this chapter. In Chapter 5, a neural network approach is developed to reduce the number of stages at which the MR model is solved, and the number of optimization model solutions at each of these stages, to obtain the trade-off relationships.

The general MR water quality management model (*Equations* 3.19 through 3.23) contains M variables and $\{I \times n(\Phi_R) + L + 2M\}$ constraints, where M is the number of decision variables, and the other variables are defined in Chapter 3. The computational effort required to solve this mathematical problem at each reliability level is determined, to a large extent, by the number of constraints. However, even for small water quality systems, the number of constraints could be large for the typical number of realizations required to produce a convergence of the joint probability distribution of the objective function and the corresponding management decisions. The computational capacity of most easily available computers today may be exceeded for this model.

At a given reliability level, most of the constraints that ensure acceptable water quality (*Equation* 3.20) do not influence the model solution. These unimportant constraints are henceforth referred to as non-critical constraints. The non-critical constraints are generally non-binding for the optimal solution and are satisfied by the optimal solution which is obtained based on the other constraints.

Therefore, to reduce the size of the optimization model and the CPU time required, new algorithms have been developed in this research that identify potentially non-critical water quality constraints *a-priori* so that they may be removed from the MR

model before solving it at each reliability level. Eliminating these non-critical constraints will not affect the accuracy of the model solution (see, e.g., Park and Liebman, 1993). Furthermore, since such an approach would significantly reduce the size and the computational time required by the MR model, it would allow a large number of stream background realizations to be included in the model. Two groups of potential non-critical constraints are identified in this work for a stochastic water quality management problem. The first group are those constraints that represent non-critical water quality checkpoints and the second group represent inferior realizations.

The rest of this chapter presents three algorithms developed in this research to identify potential critical constraints, that control the optimal MR model solution at each reliability level. The first algorithm utilizes the traditional SIOP solution, based on a large number of possible stream conditions, to identify constraints that represent potential critical stream locations. The other two algorithms use two different approaches to determine constraint sets that correspond to potential non-inferior realizations in the set of realizations for which water quality violation is not allowed. These algorithms are used to solve the MR model for the example BOD wasteload allocation problem solved with the heuristic method due to Morgan et al. (1993).

4.2 Determination of critical water quality checkpoints

For water quality management models, the checkpoints are chosen such that, if prespecified standards at these checkpoints are maintained, then acceptable water quality

levels are guaranteed along the stream. Often, the checkpoints are selected to include several locations along the river. These checkpoints are generally included in traditional water quality management model formulations because critical locations are typically unknown until the management model is solved. However, most of these locations may be non-critical and unnecessarily increase the size of the management problem.

The critical locations of a management system depend on both the stream background conditions and the water quality management goals, such as the water quality standards. The magnitude and spatial distribution of information that describes stream quality background conditions (e.g., low flows, stream temperature, reaction rates, and background water quality level) influence the critical locations, since this information determines the waste assimilative capacity along the river. Different sets of water quality goals may affect the amount and spatial distribution of waste emissions, and therefore the profile of the water quality index, differently. Thus, the locations with limiting water quality levels may be different for these sets. The critical locations may also change if different standards are specified for different sections of the river.

For a given set of management goals, the water quality constraints representing non-critical locations would consistently be non-binding for all possible scenarios of stream background conditions. Before solving the MR model using the successive heuristic process of Morgan et al. (1993), some of these non-critical checkpoints may be identified with Algorithm 1. This algorithm identifies checkpoints that influence any of the SIOP model solutions for the realizations used to represent the stochastic water quality conditions. The algorithm solves the management model many times, each time using a different realization of design stream conditions. For each solution and the corresponding realization, checkpoints with a high potential of incurring water quality violation,

if there is a small change in the design conditions, are identified as potential critical checkpoints. Such checkpoints will potentially be important in the MR model solution process while the other checkpoints will not significantly affect the solution.

Algorithm 1

Step 1 Randomly generate S possible realizations of water quality background conditions;

Step 2 Solve the deterministic water quality management model based on each generated realization in Step 1, and identify the critical checkpoint(s) for each realization; and

Step 3 Classify all checkpoints identified in Step 2 as potential critical checkpoints and all other checkpoints as non-critical.

In Step 2 of this algorithm, any checkpoint that has a binding water quality constraint, or some specified minimum reserve capacity (i.e., the minimum excess water quality level above the specified standard), or a reserve capacity below a certain threshold, may be considered as a potentially critical checkpoint. Thus, the potential critical checkpoints represent water quality constraints with zero or a small slack variable value, therefore, these checkpoints are expected to influence the solution of the water quality management model based on the S realization generated in Step 1 and under stream conditions similar to those that represent the S realizations. The constraints representing non-critical checkpoints identified in Step 3 are potentially non-binding constraints for the optimal MR model solution at all reliability levels and are therefore removed from

Equation 3.20. *Equation* 3.20 can then be written as:

$$P_{ij}(\mathbf{X}) \quad \geq \quad b_{ij} \qquad \forall \; i \; \in \; \Theta_c; \quad j \; \in \; \Phi_R \tag{4.1}$$

where

Θ_c = {all potential critical water quality checkpoints}.

In this thesis, Θ_c is the set of all checkpoints with zero reserve capacity for, at least, one of the S solutions obtained in Step 2 of Algorithm 1 above. Mathematically, Θ_c is obtained as follows:

$$\Theta_c \; = \; \{\rho_i : \textstyle\sum_{j=1}^{S} \lambda_{ij} \geq 1\} \qquad \forall \; i = 1, \cdots, I \tag{4.2}$$

where

ρ_i = index for water quality Checkpoint i;

λ_{ij} = 1 if $C_{ij} = 0$,

0 otherwise; and

C_{ij} = reserve capacity at Checkpoint i for Realization j if the optimal management solution based on this realization is implemented.

The computational effort required by Algorithm 1 is very small compared with that required by Morgan et al.'s (1993) heuristic successive solution procedure. However, the reduction in the size of the management system and the improvement in the CPU time needed to solve the resulting MR model may be very large when this algorithm is utilized. Furthermore, the algorithm may be used to screen important realizations that could be included at each reliability level of the MR model, as described in Algorithms 2 and 3 in Sections 4.3.1 and 4.3.2, respectively.

4.3 Determination of critical realizations

The superiority of the MR model over the traditional SIOP approach in developing trade-off relationships for a stochastic water quality system is mainly due to the inclusion of non-inferior realizations among the sets of stream conditions used to represent the system. However, at a given reliability level, R, many of the realizations included in the set Φ_R may be inferior to the other realizations. Although there is no analytical expression or procedure available for identifying the inferior realizations, there are certain characteristics of the traditional SIOP solutions that may be explored to determine some potential non-inferior realizations. For a given realization, these characteristics may include but are not limited to (1) the optimal value of the objective function, (2) the number of water quality violations at each checkpoint when the management solutions based on other realizations are invoked, (3) the reliability of the solution when the optimal management decision based on this realization is implemented, (4) the vector of optimal decision variables, and (5) the magnitude of the dual and slack variables. For purposes of this thesis, in this chapter, the first two characteristics are utilized to demonstrate two different approaches for identifying potential critical, non-inferior, realizations at each stage of Morgan et al.'s heuristic successive solution process for the MR model. In other applications, other characteristics may be used to identify the potential critical realizations.

4.3.1 Potential critical realizations based on optimal objective function

The ranks of the optimal objective function values for the traditional SIOP solutions may not give a good estimate of the criticalness of a realization, nevertheless, they give an indication of potentially inferior realizations at each stage of Morgan et al.'s (1993) successive solution process for the MR model. For example, if a sufficient number of realizations is used in an SIOP model, it may be reasonable to assume that the realization corresponding to the smallest optimal objective function value for an LC program would be inferior to that corresponding to the largest optimal objective function value. In fact, for a given system, there exists a parameter γ such that the worst γS realizations in the set Φ_R, based on the optimal objective function values from the traditional SIOP solution, constitute the potential non-inferior realizations. Thus, γ represents a proportion of the total number of realizations that influences the MR solution at all reliability levels. The parameter γ is determined with Algorithm 2. This algorithm obtains γ for a small sample of the total number of realizations and assumes that this γ is the same as that for a larger number of realizations. The value of γ is computed using information obtained by solving the SIOP and the MR models for the small sample, ranking the objective function values of the SIOP solutions, and identifying the binding constraints of each optimal solution of the MR model. The MR solution for the small sample is based on Morgan et al.'s (1993) heuristic approach.

Algorithm 2

Step 1 Generate a small number of realizations, N (about 10% of S), of possible water quality responses for checkpoints that belong to the set of potential critical

checkpoints Θ_c, and solve the deterministic management model based on each realization. The results obtained are the SIOP solutions for N realizations. Rank the optimal values of the objective function for each realization such that the largest (i.e., worst) value is ranked 1 for a minimization problem. Denote the rank for Realization j as D_j. Thus, D_j is the estimated rank of criticalness for Realization j based on the SIOP solutions;

Step 2 Solve the MR model for levels of R (between 1 and 0) for the N realizations generated in Step 1 using Morgan et al.'s heuristic approach and identify the set of binding realizations at each level of R, β_R; and

Step 3 Determine the smallest proportion, γ, of the N realizations, that must be included at all stages of the solution process, to obtain the exact results in Step 2.

The parameter γ can be computed as follows:

$$\gamma = \frac{1}{N}\max_R[\max_{j\in\beta_R}(D_j - (1-R)N)] \tag{4.3}$$

where

$\beta_R = \{$realizations in the MR model with binding constraints at reliability level $R\}$;

and all other variables are already defined.

The term $(1-R)N$ in *Equation* 4.3 represents the number of realizations that have been dropped from the MR model at reliability level R. At reliability level R, the term $\max_{j\in\beta_R}(D_j - (1-R)N)$ represents the maximum positive difference between the estimated level of criticalness of the binding realizations that control the optimal

solution at reliability level R, and the actual level of criticalness of the realization dropped from the model at reliability level $(R + \frac{1}{N})$ (i.e., at the previous stage of solution). For purposes of this discussion, this maximum difference at reliability level R is denoted as d_R. Thus, for a minimization problem, if the set of realizations that have been dropped from the MR model at reliability level R, Φ'_R, is known, then d_R realizations that correspond to the highest objective function values in Φ_R will include all the binding realizations that determine the optimal MR solution at this reliability level. The ratio of the maximum of d_R over all values of R to the total number of realizations N, defines the value of γ.

From *Equation* 4.3 and the discussions above, the value of γ is influenced by those realizations in β_R whose estimated level of criticalness D_j, are higher than $(1 - R)N$. If the estimated level of criticalness for a given realization is less than $(1 - R)N$, the realization will be identified by Algorithm 2 as a potential critical realization earlier in Morgan et al.'s successive heuristic process than at reliability level R. This will not affect the accuracy of the MR model solution.

The value of γ in Algorithm 2 is assumed to be the same for an MR model which uses S realizations to represent the stochastic water quality system. Therefore, at each stage of the successive solution process, the 'worst' γS realizations in the reduced MR model (i.e., in the set Φ_R), are considered as potential non-inferior realizations and the remaining realizations are assumed to be inferior. These 'worst' γS realizations correspond to the most costly γS LC solutions, based on optimal objective function values for each realization of the SIOP solution. Thus, each of the potential non-inferior realizations at a reliability level R, has an estimated rank of criticalness, D_j, not greater than $(1-R+\gamma)S$. The subset containing potential non-inferior realizations

for the general model, *Equations* (3.19) - (3.23), at a reliability level of R is denoted as $\Psi_{R\gamma}$, where $\Psi_{R\gamma}$ is a subset of Φ_R and each element of $\Psi_{R\gamma}$ has an estimated rank of criticalness less than or equal to $(1 - R + \gamma)S$. Mathematically, $\Psi_{R\gamma}$ is obtained as follows:

$$\Psi_{R\gamma} = \Omega_R \cap \Phi_R \tag{4.4}$$

where

$$\Omega_R \quad = \quad \{G_j : D_j \leq (1 - R + \gamma)S \quad \forall \quad j = 1, \cdots, S\} \quad \forall \quad 0 < R \leq 1 \tag{4.5}$$

and all other variables are already defined.

Equation 4.1 can now be written as:

$$P_{ij}(\mathbf{X}) \quad \geq \quad b_{ij} \quad \forall \; i \; \in \; \Theta_c; \quad j \; \in \; \Psi_{R\gamma} \; \subseteq \; \Phi_R \tag{4.6}$$

Using *Equation* 4.6 instead of 3.20 in the MR model, the size of the optimization problems created at the various stages of Morgan et al.'s heuristic solution process would be drastically reduced. The size of the optimization problem at reliability level R, will depend largely on the number of elements in $\Psi_{R\gamma}$, $n(\Psi_{R\gamma})$, which is given as:

$$n(\Psi_{R\gamma}) = \begin{cases} \gamma S & \text{if } R \geq \gamma \\ & \forall\ 0 < R \leq 1 \\ RS & \text{if } R < \gamma \end{cases}$$

Generally, the value of γ from Algorithm 2 will depend on the set of realizations used. By assuming that γ from Algorithm 2 is the same for any set of realizations, over- or under-estimation of the actual value of γ may be produced for the large number of realizations, S, required by the MR model. Overestimated γ will create a larger MR

model than that required if the optimal γ value is used, but will not affect the results of the MR solution based on Morgan et al.'s (1993) algorithm. On the other hand, an underestimated γ may produce an MR solution that differs from Morgan et al.'s solution. This situation arises because some realizations may be excluded from the MR model at reliability levels where they are important and affect the Morgan et al. (1993) solution process. To overcome the problem of an underestimated γ value, a safety factor may be applied to increase the value obtained from Algorithm 2. However, this thesis does not investigate how the safety factor may be obtained. Presently, simulation and the modeller's experience are recommended.

4.3.2 Potential critical realizations based on number of water quality violations

For a given set of management objectives, the spatial variation of stream information, such as low flow, stream velocity, stream temperature, reaction rates, and background water quality index, significantly contributes to the presence of non-inferior realizations. The higher the spatial variation, the larger the number of non-inferior realizations that are responsible for inefficient solutions in an SIOP model. For example, Morgan et al. (1993) show that increasing the heterogeneity of the hydraulic conductivity field of a stochastic groundwater remediation design increases the number of inefficient SIOP model solutions with respect to the cost-reliability trade-off curve obtained with the MR model.

Recall from Chapter 3 that U_j is the overall level of criticalness of Realization j. The value of U_j is obtained by considering all water quality checkpoints along a

stream in Morgan et al.'s heuristic algorithm. In this research, to determine potential critical realizations using the number of water quality violations that may occur at each checkpoint of each realization, U_j is assumed to depend on some measure of the level of criticalness of Realization j at each individual checkpoint $i \in \Theta_c$. For the remainder of this thesis, the level of criticalness of Realization j at Checkpoint i is denoted as w_{ij}.

Generally, U_j may be represented as a function of w_{ij} for $i \in \Theta_c$, in *Equation* 4.7 below:

$$U_j \quad = \quad F(w_{ij}) \tag{4.7}$$

where

$F(w_{ij})=$ a function that relates w_{ij} for all $i \in \Theta_c$ and for Realization j to U_j; and

all other variables are already defined.

Equation 4.7 assumes that the overall level of criticalness for a given realization may be attributed to a combination of its level of criticalness at the checkpoints that belong to the set Θ_c. The level of criticalness of a given realization at Checkpoint i may be approximated with some performance criterion for this realization, such as the number of water quality violations at this checkpoint of the given realization, when all of the SIOP solutions are implemented. However, it may be difficult or impossible to calculate U_j because there are no analytical functions available to represent $F(w_{ij})$. The new heuristic algorithm developed in this section proposes an index for approximating w_{ij}, and uses these estimated values to select a set of potential critical realizations at each stage of Morgan et al.'s heuristic algorithm.

The values of w_{ij} will likely depend on the probability of water quality violation at Checkpoint i of Realization j if the solutions of all of the traditional SIOP model are

invoked. For example, if solutions of 90 out of 100 realizations violate a DO standard at Checkpoint i of Realization F while only those of 50 realizations violate the standard at the same checkpoint of Realization H, then w_{iF} is likely to be a more critical index than w_{iH}. In this thesis, the values of w_{ij} are estimated as:

$$w_{ij} = S - T_{ij} \quad \forall \; i \in \Theta_c; \quad j = 1, \cdots, S \tag{4.8}$$

where

T_{ij} = number of water quality violations at Checkpoint i of Realization j if the other single realization solutions are implemented.

From *Equation* 4.8, the lower the value of w_{ij}, the higher the estimated level of criticalness of Realization j at Checkpoint i.

Based on the estimated values of w_{ij}, in *Equation* 4.8, an improved version of Morgan et al.'s algorithm, referred to as the IM algorithm, is used to select potential critical realizations and to solve a minimization problem. The main difference between Morgan et al.'s heuristic algorithm and the IM algorithm is that, while Morgan et al.'s approach includes all realizations that have not been dropped from the MR model (i.e., the set Φ_R) to obtain the solution at a given reliability level, the IM algorithm selects a small number of potentially important realizations that belong to the set Φ_R to produce an efficient solution. For each Checkpoint $i \in \Theta_c$, the IM algorithm obtains potentially important realizations by selecting a small number of the elements of Φ_R that have the lowest values of w_{ij}. The realizations selected for the checkpoints are initially assumed to include all the important realizations needed to obtain an efficient solution at the given reliability level. However, for a given reliability level R, the selected realizations

may exclude some important background stream conditions. These excluded stream conditions correspond to realizations that adversely affect the IM solution for reliability levels lower than R. This situation often results in a more expensive LC solution for a lower reliability level than that for a higher reliability. The IM algorithm identifies such anomally and the important realizations missed at an earlier solution stage, due to the assumption about which realizations are important, and steps back in the solution process to include these realizations in the MR model at the appropriate reliability levels. While the potential non-inferior realizations in Algorithm 2 are obtained based on the most costly LC SIOP solutions corresponding to realizations in the set Φ_R, the important realizations in the IM algorithm are selected on the basis of the number of non-violations of water quality at each checkpoint (under stream conditions represented by realizations that belong to Φ_R), if the SIOP solutions of other realizations in $\Phi_{1.0}$ are implemented in turn. The IM algorithm is given in Figure 4.1. The symbols for this algorithm are defined as:

k_i = variable indicating the upper bound on w_{ij} for realizations to be included in the MR model at a given reliability level;

$ISTART$ = heuristic integer parameter indicating the number of the most critical realizations at each checkpoint included at the initial stage of the IM algorithm ($ISTART \ll S$);

$JCOUNT$ = integer variable > 1 if candidate problems for the optimal solution at a given stage are created at the previous stage of the solution process,
≤ 1 otherwise;

H_j = integer variable equal to 0 if water a quality violation is not allowed for Realization j at a given reliability level, 1 otherwise;

$\mho_R$ = {potential critical realizations at reliability level, R};

$LARGE$ = large positive number relative to expected objective function values;

$OBJS$ = objective function value for the most current optimization problem solved;

$\mathbf{X}$ = vector of decision variable values corresponding to $OBJS$;

OBJ = current best objective function value during the solution process at a given reliability level;

$\mathbf{Y}$ = vector of decision variables values corresponding to OBJ;

$ICRIT_R$ = most critical realization at a reliability level R;

$OBFUN_R$ = optimal objective function value at reliability level, R;

$\mathcal{X}_R$ = vector of decision variables corresponding to $OBFUN_R$;

M = total number of binding realizations at current stage of solution process;

Q_j^m = 2 if G_j is the m th binding realization at the previous level of reliability $R + \frac{1}{S}$, 0 otherwise; and

the other symbols are already defined.

Description of the IM algorithm

Step 1 of the IM algorithm initializes parameters required by the algorithm. At the second step, the objective function value at the beginning of each stage (i.e., at each reliability level) of the solution process is set to a large positive number to assist in identifying the smallest objective function value among candidate solutions. In Step 3, the binding realizations of the optimal solution at the previous reliability level, $R + \frac{1}{S}$, are removed from the model in turn to create candidate optimization problems that will be solved at reliability level R, in Step 4. However, if the counter, $JCOUNT$ is equal to 1, no candidate solutions are created for the present stage, therefore the algorithm proceeds to Step 4 to create one optimization problem. In Step 4, potential critical realizations are selected for one of the optimization problems created for the current stage, and the optimization model is solved. The optimal (best) objective function value and the corresponding decision variables at a reliability level, R, are selected in Step 5. For cases in which $JCOUNT$ is not equal to 1, the most critical realization at the previous stage (i.e., at reliability level $= R + \frac{1}{S}$) is also identified in Step 5. Step 6 ensures that all binding realizations of the solution at the previous stage have been considered as candidates for being the most critical realization at the current stage. This step is ignored if $JCOUNT$ is 1. Step 7 ensures that at the first stage of the solution process (i.e., $R = 1.0$), the algorithm proceeds to Step 10 to record the optimal solution, since Steps 8 and 9 are not relevant if $R = 1.0$. Because the w_{ij}'s are approximations of the level of criticalness at the checkpoints, some of the important or non-inferior realizations at a given reliability level R, may not be included in the set $\mho_R$. Therefore, at Step 8, the solution obtained at every reliability level $R < 1$, is compared with those

at the previous levels to ensure that the optimal objective function value is an increasing function of R. Also, in Step 8, if the solution at reliability level R, is less than that at level $R+\frac{1}{S}$, the most critical realization identified at level $R+\frac{1}{S}$ (in Step 5) is dropped from the model and the parameter k_i is adjusted. This adjustment is needed to include inferior realizations that may become non-inferior due to the elimination of the most critical realization at level $R+\frac{1}{S}$ from the MR model. In Step 9, if unsatisfactory solutions (in terms of producing an increasing functional relationship between objective function and reliability) are identified in Step 8, the stage of the successive process is moved back in order to re-solve the model, using higher k_i values for reliability levels where the unacceptable functional relationship occurs. In Step 10, the algorithm is directed to Step 4 to re-solve the model, if an unsatisfactory functional relationship was observed at Step 8. Otherwise, the optimal solution at reliability level R is saved in Step 10. All binding realizations of the optimal solution recorded in Step 10 are identified in Step 11. These binding realizations are used to create the problems for the next stage of the successive solution process. At the final step, Step 12, the solution is terminated if the reliability level is zero or there are no binding realizations at a given stage. The latter condition for termination of the algorithm (i.e., $M = 0$) assumes that the management system is water quality limited. If conditions for termination are not met, the reliability level is adjusted to the value corresponding to the next stage, the counter for the binding realizations, m, is initialized, and the algorithm returns to Step 2.

4.4 Discussion of results

The algorithms developed in this chapter for determining potential critical checkpoints and realizations, and for solving the MR model, are demonstrated for the LC management program presented in Chapter 3. Out of 35 water quality checkpoints only 7 were identified by Algorithm 1 as potentially critical. These checkpoints were distributed throughout the entire length of the river as shown in Table 4.1.

The remaining 28 non-critical checkpoints were not considered in Algorithm 2 and in the IM algorithm. However, the reliability of the solutions obtained using these two algorithms is evaluated by considering all of the 35 checkpoints for each realization included in the MR model.

For Algorithm 2, the computed values of γ for the four sets of Monte Carlo simulation (MCS_1, MCS_2, MCS_3, and MCS_4) used in Chapter 3 are shown in Table 4.2. This table shows different values of γ for four applications of Algorithm 2 based on 100 realizations each. The value of γ equal to 0.21 for MCS_2 indicates that at any reliability level, a maximum of 21 realizations, based on the ranks of the optimal objective function values of the traditional SIOP solutions of MCS_2, constitute the potentially critical realizations. Therefore, the values of γ obtained from Algorithm 2 may require the addition of a factor of safety to ensure that the set $\Psi_{R\gamma}$, in *Equation* 4.6, contains all non-inferior realizations at each reliability level, for a case in which a large number of realizations are incorporated in the MR model. Alternatively, more stable values of γ may be obtained by increasing the number of realizations N, used in Experiment 2. However, larger values of N will have a corresponding exponential increase in the CPU time required by Algorithm 2 to estimate the value of γ.

Table 4.2 also shows the values of γ_a which are computed values of γ if the ranking of criticalness of the realizations based on Morgan et al.'s algorithm, U_l, in each of the four sets of Monte Carlo simulations, are used instead of D_l, in *Equation* 4.3. As expected, the values of γ_a are generally smaller than the corresponding values of γ. The error in using the rank of the optimal objective function value for a given realization to approximate the criticalness of that realization is responsible for γ being greater than γ_a in some cases. The values of γ_a suggest that, even if the correct ranking of the criticalness of the realizations in the MR model are known *a-priori*, several realizations must be included to obtain efficient solutions.

To compare the computational efficiency of the two methods for identifying potential critical realizations, solutions of the MR model are obtained based on Algorithm 2 and on the IM Algorithm. One hundred stream background realizations were utilized for this comparison. The maximum of the values of γ for the four sets of Monte Carlo simulation, MCS_1 - MCS_4, (i.e., $\gamma = 0.21$) is used in *Equation* 4.6, and the heuristic integer parameter $ISTART$ for the IM algorithm is arbitrarily set at 2. Each of these two approaches for solving the MR model produced the exact same values of the optimal objective function and decision variables at all reliability levels as those obtained by Morgan et al.'s heuristic algorithm. This implies that the IM algorithm may screen out unimportant or inferior realizations from the MR model and create optimization models that include the same set of binding realizations as Morgan et al.'s algorithm at each reliability level. The CPU time for solving the MR model on the University of Manitoba Sparc2 Unix stations was 483 and 256 seconds for Algorithm 2 and the IM algorithm, respectively. Recall from Chapter 3 that, using the same computers, Morgan et al.'s algorithm requires approximately 7 days to solve the MR model for a case in

which 100 realizations are utilized. The CPU time for Algorithm 2 did not include the computational time used to obtain γ. The computational time for solving the MR model by the two methods is also compared for a case in which 500 realizations are used. For this case, while the CPU time utilized by Algorithm 2 is approximately 26 hours, that by the IM Algorithm is only 0.3 hours.

Thus, although both methods may achieve substantial savings in computational time compared with Morgan et al.'s approach, the IM Algorithm appears to perform better than Algorithm 2 for the example LC programs analyzed in this research. However, Algorithm 2 may perform well for cases in which γ values are small. For such small values, the number of realizations included in the MR model at each reliability level may be small.

Because the computational time requirement of the IM Algorithm is less than that of Algorithm 2 for solving the example problems based on 100 and 500 realizations, the IM Algorithm is further investigated here. The IM algorithm is used to solve the MR model based on MCS_1, MCS_2, MCS_3, and MCS_4. The cost-reliability trade-off curves obtained are the same as those presented in Figure 3.3. This suggests that the IM Algorithm is adequate in terms of producing efficient solutions. Recall from Chapter 3 that 100 realizations are not enough to represent the stochastic water quality system analyzed in this thesis. The number of realizations in the MR model was gradually increased from 100 until significantly improved convergence characteristics of the model output was observed for 2000 realizations. For the improved model output convergence, the differences between the cost-reliability trade-off curves, based on different sets of 2000 realizations, is smaller than those based on MCS_1, MCS_2, MCS_3, and MCS_4.

Figure 4.2 shows the trade-off relationships between the total waste treatment cost and reliability obtained by the IM Algorithm for the LC MR model that utilizes four different sets of 2000 realizations each, SIM_1, SIM_2, SIM_3, and SIM_4. The four curves are developed to assess the convergence characteristics attained by the solution of the MR model that incorporates 2000 realizations. Small differences between these curves indicate good convergence of the MR model solution based on 2000 realizations, while large differences imply that the number of realizations included is inadequate. Thus, the four trade-off curves in Figure 4.2 indicate very good convergence of the output from the model at reliabilities less than 0.82. Although convergence of these curves at high reliability levels (i.e., reliabilities greater than 0.82) is not as good as that at low levels, it is significantly better than that observed for the case in which 100 realizations were used by the MR model (see, Figure 3.2).

The maximum absolute difference between the waste treatment cost for the curves in Figure 4.2 is $ 0.96 million/year. This maximum difference, which occurs between SIM_1 and SIM_3, represents approximately 64% change in cost at the corresponding reliability level. Although this maximum difference is high, the differences between the total waste treatment cost for these curves are small at several reliability levels. If better convergence is required than that shown in Figure 4.2, a larger number of realizations must be used.

The CPU time used by the IM algorithm to solve the MR model that utilizes 2000 realizations is approximately 10.5 hours. Therefore, if the number of realizations is increased above 2000, the computational resources needed could be very high, since the CPU time required to solve the MR model increases exponentially with the number of realizations.

The sensitivity of the computational resource requirements of the IM Algorithm to the heuristic integer parameter that indicates the number of the most critical realizations at each checkpoint at the initial stage of the IM algorithm, $ISTART$, is investigated by solving the MR model for two cases in which 100 and 2000 realizations are used. This parameter is varied from 1 to 6 for each case. Figure 4.3 shows a plot of CPU time utilized versus the heuristic parameter. The figure shows that, for a given number of realizations, the CPU time is high when $ISTART$ is equal to 1. It decreases to a minimum when $ISTART$ is 2, and then increases slowly with $ISTART$ values up to 6. The slow rate at which the CPU time changes with $ISTART$ values between 2 and 6 indicates that the parameter is not sensitive to computational resources within this interval. Figure 4.3 also suggests that an $ISTART$ value of 2 may be adequate for the IM Algorithm, in terms of computational efficiency, for the example management system analyzed.

Table 4.1: Spatial distribution of potentially critical checkpoints

Critical checkpoint	Distance from Springfield (km)
1	57.80
2	85.80
3	163.03
4	216.10
5	241.80
6	254.70
7	296.59

Table 4.2: Values of γ and γ_a for four Monte Carlo experiments

MCS Experiment	γ	γ_a
MCS_1	0.16	0.11
MCS_2	0.21	0.11
MCS_3	0.15	0.15
MCS_4	0.17	0.12

Step 1 Initialize parameters by setting:

$R = 1.0$;
$k_i = ISTART \quad \forall \quad i \in \Theta_c$;
$H_j = 0 \quad \forall \quad j = 1, \cdots, S$; and
$JCOUNT = 1$;

Step 2 Set the objective function value at the beginning of each stage, OBJ, to a large positive number $LARGE$, to assist in searching for the minimum objective function value among candidate solutions in Step 5;

Step 3 If candidate problems for this stage were not created at the previous stage of the solution process (i.e., If $JCOUNT = 1$), there is only one optimization problem, therefore, go to Step 4, otherwise identify the candidate optimization problem to solve by setting:

$m = m + 1$; and
$H_j = 1$ if $Q_j^m = 2 \quad \forall j = 1, \cdots, S$;

Step 4 Obtain the set of potential critical realizations for the optimization problem identified in Step 3, $\mho_R$, as follows:

$\mho_R = \Phi_R \bigcap \Omega_R$;
where,
$\Phi_R = \{G_j : H_j = 0 \quad \forall j = 1, \cdots, S\}$;
$\Omega_R = \bigcup_{i \in \Theta_c} \alpha_i$; and
$\alpha_i = \{G_j : w_{ij} \leq k_i \quad \forall j = 1, \cdots, S\} \quad \forall i \in \Theta_c$;
Replace the water quality constraint set, *Equation* (3.20), by:
$P_{ij}(\mathbf{X}) \geq b_{ij} \quad \forall i \in \Theta_c; \quad j \in \mho_R$
and solve the resulting MR model.
Denote the optimal objective function value by $OBJS$ and the corresponding vector of decision variables by $\mathbf{X}$;

Step 5 If only one optimization problem is identified in Step 3 (i.e., $JCOUNT = 1$), the solution obtained in Step 4 is the optimal solution at the current stage, therefore store this optimal solution by setting:

$OBJ = OBJS$;
$\mathbf{Y} = \mathbf{X}$; and
go to Step 7.

Otherwise, compare the candidate solution just obtained from Step 4 with the other candidate solutions that have already been obtained for this stage. If the candidate solution just obtained indicates a better objective function value than the others (i.e., $OBJS < OBJ$), store the current candidate solution as the current optimal solution at the present stage by setting:

$OBJ = OBJS$;
$\mathbf{Y} = \mathbf{X}$;
$ICRIT_{R+\frac{1}{S}} = G_j$ if $Q_j^m = 2 \quad \forall j = 1, \cdots, S$;
$H_j = 0$ if $Q_j^m = 2 \quad \forall j = 1, \cdots, S$;

Step 6 If all of the candidate solutions have not been obtained, (i.e., $m < M$), go to Step 3;

Step 7 If this stage of solution includes all of the realizations generated for the MR model (i.e., $R = 1.0$), increase the value of $JCOUNT$ by 1, so that $JCOUNT > 1$ to indicate that candidate solutions for the next stage are created at this stage, and go to Step 10;

Step 8 Compare the optimal objective function value obtained at the current stage, OBJ, with that at the previous stage, $OBFUN_{R+\frac{1}{S}}$. If $OBJ > OBFUN_{R+\frac{1}{S}}$, then some important critical realizations were not included at the previous solution stage, therefore, go to Step 9. Otherwise, increase the value of $JCOUNT$ by 1. Identify the most critical realization at the previous solution stage and drop it from the MR model by setting:

$H_j = 1$ if $ICRIT_{R+\frac{1}{S}} = G_j \quad \forall j = 1, \cdots, S$;
Denote the critical realization just dropped from the MR model, $ICRIT_{R+\frac{1}{S}}$, by Realization l
and revise the k_i's as follows:
$k_i = k_i + 1$ if $w_{il} \leq k_i \quad \forall i \in \Theta_c$;
and go to Step 10;

Step 9 Set back the stage of the solution process by one as follows:

$OBJ = OBFUN_{R+\frac{1}{S}}$;
$H_j = 0$ if $ICRIT_{R+\frac{1}{S}} = G_j \quad \forall j = 1, \cdots, S$;
$R = R + \frac{1}{S}$;
$JCOUNT = 0$;
and go to Step 8;

Step 10 If the solution process was set back, in Step 9 (i.e., $JCOUNT = 1$), go to Step 4, otherwise save the optimal solution at reliability level R as follows:

$OBFUN_R = OBJ$;
$\mathcal{X}_R = \mathbf{Y}$;

Step 11 Identify all realizations that have one or more binding water quality constraints for the optimal solution in Step 10 and refer to each of them as a **binding realization.**

If G_j is the m th binding realization, set:
$Q_j^m = 2 \quad \forall m = 1, \cdots, M, \quad \forall j = 1, \cdots, S$;
otherwise set:
$Q_j^m = 0 \quad \forall m = 1, \cdots, M, \quad \forall j = 1, \cdots, S$;

Step 12 If there are no binding realizations (i.e., $M = 0$) or the reliability level is zero ($R = 0$), STOP, otherwise, set the counter for binding realizations to zero (i.e., $m = 0$), the reliability level of the next solution stage to $R - \frac{1}{S}$, and go to Step 2.

Figure 4.1: The IM Algorithm

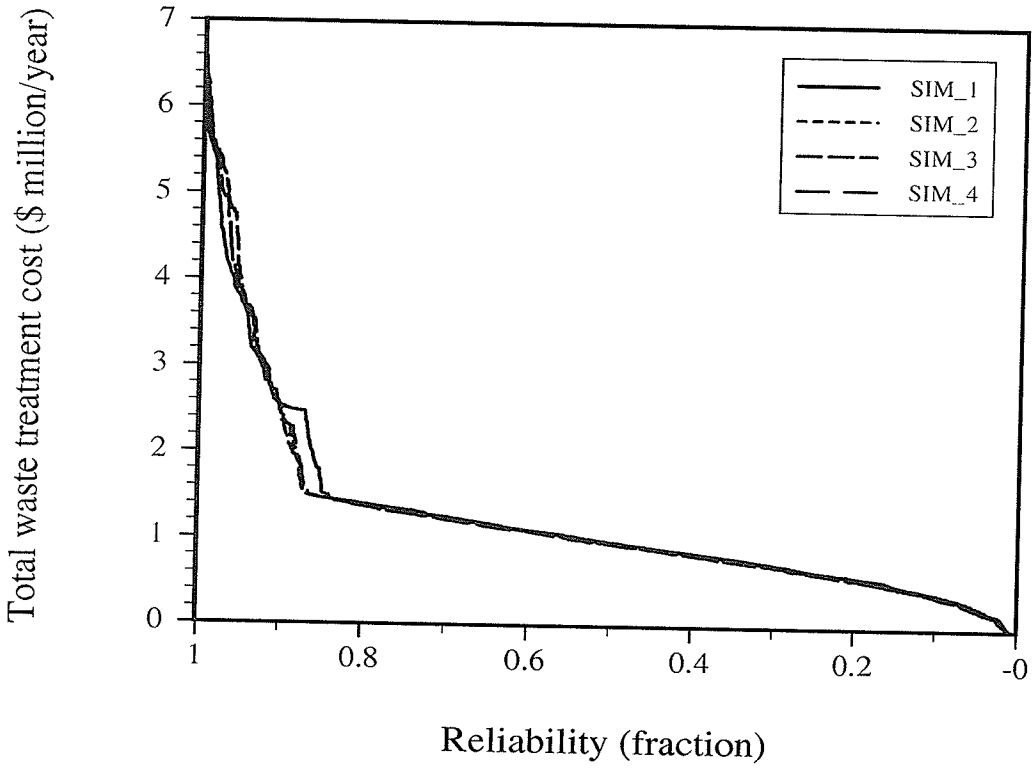


Figure 4.2: Trade-off curves for four sets of 2000 realizations each

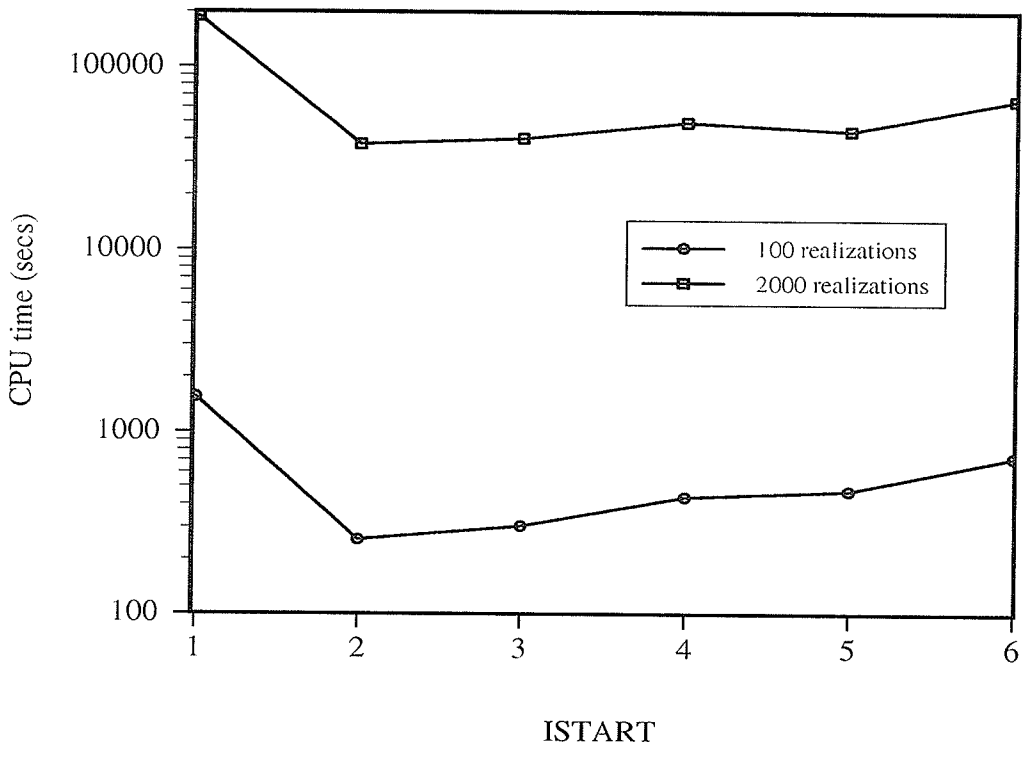


Figure 4.3: Sensitivity of CPU time to heuristic parameter $ISTART$

Chapter 5

Using Neural Network to Reduce the Computational Burden of the MR Model

5.1 Introduction

This chapter develops and demonstrates an artificial neural network (ANN) method for reducing the computational time requirements of the IM Algorithm presented in Chapter 4. Generally, the heuristic algorithm of Morgan et al. (1993) and the IM algorithm solve many optimization models at several solution stages to generate cost-reliability relationships. The large number of optimization solutions significantly increases the CPU time required by these algorithms to solve the MR model. However, a large proportion of the optimization models are necessary only because the level of criticalness of the realizations in the MR model (i.e., the order in which the realizations included

in the MR model are dropped from the model when the IM and Morgan et al. (1993) algorithms are used to obtain a solution) are not known prior to solving the model. The goal of the ANN approach is to estimate the level of criticalness of all realizations included in an MR model so that the number of optimization models to be solved is significantly reduced. The ANN method estimates the level of criticalness of each of the S realizations included in the MR model based on training examples obtained using the IM Algorithm to solve the model for a relatively smaller number of realizations than S.

The ANN approach is derived from the work of Ranjithan et al. (1993) who utilize an ANN to screen important realizations required to solve an MR model at high reliability levels for a groundwater reclamation design. However, to improve the prediction performance of the ANN for the surface water quality system presented in this thesis, the detailed ANN design, analysis, and application differ significantly from those of Ranjithan et al. (1993).

As indicated in Chapter 4, there is no analytical expression for obtaining the level of criticalness of the realizations included in an MR model. Even if an analytical expression exists, it is expected to be quite complex due to many components of a realization that may be related to the level of criticalness. Thus, to describe the relationship between the level of criticalness and the components of a realization, it is attractive to use ANNs which have been shown to be capable of approximating practically any function if adequate network architecture and training are provided (Hornik et al. 1989). For the MR model presented, the important components of a realization may be the coefficients of the constraints and objective function or the solution (i.e., optimal objective function and decision, dual, and slack variables) of the management model based on this realization. Ranjithan et al. (1993) use the log-hydraulic conductivity at several

locations within a hypothetical aquifer as the important information required by an ANN to select a few realizations needed by Morgan et al.'s (1993) heuristic algorithm to develop the trade-off relationship between cost of aquifer remediation and reliability, at reliability levels greater than 80%.

To use an ANN to represent a given relationship, satisfactory training and testing, or generalization, performance of the network is required. Once an adequate ANN architecture is obtained and sufficient training is provided to the network to produce acceptable generalization performance, the network is adapted to estimate the criticalness of each of the S realizations included in the MR model. Computational time efficiency of the ANN based approach for solving the MR model is gained in the following two ways. (1) The MR model does not have to be solved several times at each stage in order to obtain the optimal solution among candidate solutions, and to create the optimization problems for the next stage of the solution process. This is because the level of criticalness of each realization will be "known" *a-priori*, therefore, the realization to be dropped at each stage of the solution process will also be "known" before solving the model. (2) A number or group of critical realizations may be removed from the MR model at a time, instead of one critical realization at each stage for the IM algorithm, in order to develop the trade-off relationship between optimal objective function value and reliability. Because of the efficiency of the ANN training mechanism and the parallel computer processing capability of ANNs, the methodology developed in this chapter may achieve substantial computational time savings for water quality systems that require a large number of realizations to represent the stochastic nature of the system.

The next section describes general ANN modelling. A prototype ANN for estimating

the level of criticalness of each realization in an MR model is then presented. Finally, an ANN-based algorithm for solving the MR model is described and demonstrated for the BOD wasteload allocation problem solved in Chapters 3 and 4.

5.2 ANN modelling

5.2.1 Basic theory

Computers are known to be much faster and more efficient than human brains in doing arithmetic computations. However, there is a class of problems such as speech generation, speaker recognition, image identification, pattern classification, character recognition, and handwriting analysis, among others, which the brain is much better at solving than the fastest supercomputer. The biological structure and information processing mechanism of the brain are responsible for the ease with which the above tasks are performed several times each day by humans without consciously thinking about what is being done. Some of the important features of human brain that make it so powerful in performing the class of problems listed above include:

- Its robustness and fault tolerance. Nerve cells in the brain die every day without affecting its performance significantly;
- Its flexibility. It can easily adjust to a new environment by "learning;"
- Its ability to deal with information that is fuzzy, probabilistic, noisy, or inconsistent;

- Its highly parallel and interconnected structure. It relies on a large network of simple highly interconnected computing elements (neurons) for its computation; and

- Its small size, compact nature, and very low power dissipation.

ANN models are motivated by the capability of human brains to easily perform some tasks that are exceedingly difficult, if not impossible, to solve using standard computer architectures and algorithms. Therefore, ANN research has produced a simplified model which is believed to represent the important features of biological computation in the brain. Although, presently there is very little known about biological systems, and the ANN parallels drawn between these systems may be oversimplifications of actual biological neural models, nevertheless, ANN is commonly considered as a novel mathematical computational scheme for many nonlinear and extremely complicated systems.

Generally, an ANN may be considered as a parallel processing-based computational model consisting of several massively interconnected simple processing elements, and having the ability to adapt or learn, to generalize, or to cluster or organize data through the learning of training patterns submitted to the model. The learning, generalization, or clustering process is performed by gradually adjusting internal parameters of the network to the point where the model can produce the 'correct' or 'acceptable' answer in response to each input vector. After the learning procedure is successfully completed for a given set of training examples, the model performance is tested on different examples (testing examples) which were not used in the training process. If the network performance is adequate, the final internal parameters may be used to predict answers corresponding to other input examples or patterns that are presented to the network.

Classical logic in a rule-based, artificial intelligence system is replaced by vague conclusions and associative recall (i.e., an exact match versus a best match) in ANNs. This is an advantage in situations where no clear set of logical rules can be given. The main difference between an ANN and an expert system approach of artificial intelligence is that, while the former is based on empirical factual knowledge of symbolic processing, the latter involves programming a system to apply a hierarchy of explicit rules. When describing the difference between ANNs and expert systems, Roger and Dowla (1994) wrote:

> "One might view expert systems as organizing behaviour by description, whereas neural networks attempt to imitate the behaviour. One might make a case that, instead of using a set of rules, human experts often apply intuition or deeper insight to the behaviour that they have learned. ..."

5.2.2 Architecture

The basic structure of an ANN (Figure 5.1) is based on several simple building blocks often called processing units or neurons. These units which perform the processing are similar and communicate by sending signals to each other over a large number of weighted and directed connections. The connection between each pair of units is called a link or synapse. The basic form of a processing unit and link are shown in Figure 5.1. Information processing within a unit is performed with an activation function. This activation function uses the net input to a unit from the weighted output values of prior connected units to compute a new activation or output for the unit.

Processing units (neuron)

Each unit performs a simple task by receiving input information from neighbouring units or external sources and using this information and some transformation function (i.e., an activation function) to compute an output signal (i.e., an activation level) which is transmitted to other units or outside the network. A unit often has many inputs, but produces only one output signal. In addition to the above task, the units perform weight adjustments of the links between them during the process of learning the training patterns submitted to the network. The information processing procedure is basically parallel since many units can perform their computations concurrently.

Depending on their function in the network, three types of units - input units, hidden units, and output units may be defined. The units that receive their inputs from a source which is external to the network, and therefore whose activation levels are the inputs for the network are called input units. Those units that drive external signals and whose activations represent the output of the network are known as output units. Because they are not visible from outside the network, the remaining units are referred to as hidden units. The input and output signals, or activations, of hidden units remain within the network itself. The input and output units constitute an input layer and an output layer, respectively, while the hidden units may be grouped into one or more hidden layers. The network in Figure 5.1 consists of three input units, four hidden units, and two output units and organized into three layers. The number of units in each layer of the network, and the number of layers depend on the application of interest and the type of network being used.

Links (Synapses)

In general, the input layer is connected to the hidden layer which is in turn connected to the output layer. For networks with more than one hidden layer, these layers would also be connected sequentially. For example, if there are two hidden layers, the input layer will be connected to the first hidden layer which will be connected to the second hidden layer, and finally the second hidden layer will be connected to the output layer. The links between the processing units are directed connections and the direction indicates the direction of transfer of information, signals, or activations. The units from which a connection starts, the source unit, transmits its output activation to the other unit, the target unit. Multiple connections between one unit and the same input port of another unit are not allowed.

Each connection has associated with it a weight which represents the contribution of the output, or activation, of a particular source unit towards the output, or activation, signal of the target unit. In most cases, simple weighted summation of the activation signals of all source units of a given target unit (a propagation rule) is performed to obtain a single aggregated input to this target unit. Often, a bias or offset term is added to the aggregate input to improve the performance of the network. Thus, the net aggregate input to unit m is computed as:

$$\iota_m = \sum_n \mathcal{W}_{mn} a_n + \theta_m \tag{5.1}$$

where

ι_m = the net input to unit m;

$\mathcal{W}_{mn}$ = the weight of the connection from source unit n to target unit m;

a_n = the activation value of source unit n of the network;

θ_m = the bias input to target m.

If the value of $\mathcal{W}_{mn}$ is negative, then this weight decreases the activity of the target unit and the connection between units m and n is considered as an inhibitory link. On the other hand, positive values of $\mathcal{W}_{mn}$ enhance the activity of the target unit and are considered as excitatory. In some cases more complex rules for combining inputs are used, in which a distinction is made between excitatory and inhibitory inputs. The units that follow the propagation rule in *Equation* 5.1 are called sigma units. There are other less commonly used propagation rules such as the sigma-pi given as:

$$\iota_m = \sum_n \mathcal{W}_{mn} \prod_k b_{nk} + \theta_m \tag{5.2}$$

where

b_{nk} = the activation value contribution of source unit k towards the activation of target unit n; and

all other symbols are already defined.

5.2.3 Activation functions

To obtain an output signal for a given unit, the net unit input computed with *Equation* 5.1 or 5.2 is transformed using a squashing function as follows:

$$a_m = \mathcal{F}_m(\iota_m) \tag{5.3}$$

where

$\mathcal{F}_m(.)=$ the activation function associated with unit m.

Often, $\mathcal{F}_m(.)$ is a nondecreasing function of the net input ι_m, into a target unit, although activation functions are not restricted to nondecreasing functions. Generally, some form of threshold function is used. Figure 5.2 shows some commonly used activation functions.

In this thesis, the sigmoid function is used to compute activation values of target units since it the most popular activation function used in the literature for an ANN with non-discrete outputs. This function is given as:

$$\mathcal{F}(\iota_m) = \frac{1}{1 + \exp^{-\iota_m}} \tag{5.4}$$

Substituting for ι_m on the right hand side of *Equation* 5.4 from *Equation* 5.1, $\mathcal{F}_m(\iota_m)$ is computed as:

$$\mathcal{F}(\iota_m) = \frac{1}{1 + \exp^{-(\sum_n w_{mn} a_n + \theta_m)}} \tag{5.5}$$

The output signal of a unit may be a stochastic function of the net input into the unit. In such a case, the net input is used to determine the probability p, that the unit's activation value is high using the following equation:

$$p(a_m \leftarrow 1) = \frac{1}{1 + \exp^{-\iota_m / g}} \tag{5.6}$$

where

g = a parameter which determines the slope of the probability function.

5.2.4 A simple classification of ANNs

The pattern of connections between links and the method of data transmission within networks classify ANNs into two broad groups, namely, feed-forward and recurrent networks. In feed-forward networks, the direction of information propagation is strictly input unit $\rightarrow$ hidden unit $\rightarrow$ (hidden unit) $\rightarrow$ output unit. Information can also be transmitted directly from an input unit to an output unit. For feed-forward networks with multiple hidden layers, units in hidden layers that are closer to the input layer can send information to other hidden layers farther away from the input layer or to the output layer. If the architecture is considered to be a structured layout with sequential numbering of the layers from inputs to outputs, then feed-forward networks allow information transfer from units in lower numbered layers to those in higher numbered layers. No feedback connections are permitted, that is, connections extending from higher numbered layers to the same or lower numbered layers are not allowed. Thus, feed-forward networks have unidirectional connections. Examples of feed-forward networks include Adaline and Perceptron ANNs.

Networks that are not strictly feed-forward, but include direct or indirect loops of connections from units of a higher numbered layer to those of the same or lower numbered layer, are often called recurrent networks. Kohonen and Hopfield ANNs are examples of recurrent networks.

5.2.5 ANN learning

Unlike standard computers which perform their task by following a sequence of steps laid out by a human programmer in some form of algorithm, an ANN learns to solve

the desired task by studying a set of example patterns (training examples) submitted to it. Information needed to solve a given problem is stored in the form of connection weights $\mathcal{W}_{mn}$. Through the modification of the $\mathcal{W}_{mn}$ values, the network is able to learn to solve problems that are similar, but not necessarily identical, to the example patterns used to train it. Thus, an ANN has to be designed such that the application of an input pattern, which lies within the bounds of the problem to be solved, produces output signals that are the same as or similar to the target output.

Many weight modification iterations are performed in order to obtain the 'correct' connection weights that represent the system. During each iteration, the connection weights and activations of the processing units are used to compute a new activation value for each unit in the network. The output signal of the network for each training example is compared with the expected solution. The error in predicting the correct solutions for the example training patterns is used to update the weights or the unit activations. There are several modes of updating the unit activations, but they can be categorized into synchronous and asynchronous updating. With synchronous updating, the units change their activation all together after each iteration, while, with asynchronous updating, each unit has a probability of modifying its activation during a given iteration.

A successfully trained network has two main characteristics. First, by using the activation functions, it is capable of transforming the input patterns of the training set to output signals that match corresponding target outputs of the training examples. Second, it is capable of producing 'correct' output for new patterns that belong to the class of the training examples. The former characteristic is referred to as the learning ability while the latter is often called the generalization ability.

The learning and generalization performance of a network are often measured by a function of the difference between target outputs and computed network outputs for training and testing example sets, respectively, submitted to the network. The smaller the value of the function the better the training or generalization capability.

The success of an ANN at learning and generalizing depends on its architecture, the training examples and data preparation, the type of learning algorithm employed, the technique used to optimize the learning performance index, and the extent of training the network undergoes. As shown in Figure 5.1, the architecture of a given network consists of the number of input, hidden, and output units and their interconnections. In order to learn to solve a given type of problem, the salient characteristics of the input vectors, or patterns, and sometimes of the output signals, have to be included in the network. If important features of the input and output signals are not included, the ANN may not be successful at learning and generalizing. On the other hand, if unimportant features are incorporated in the network, these features could serve as noise terms which may degrade the network's learning and generalization performances. Although ANNs are often tolerant against noise, increased noise may affect the network's performance. Also, noise terms may increase the size of the network unneccessarily.

The appropriate number of hidden units is problem-dependent. There are techniques in the literature for efficiently determining the appropriate number of hidden units (see, e.g. Ash, 1989; Fahlman and Lebiere, 1990; Mozer and Smolensky, 1989; Hirose et al, 1991; Karnin, 1990; Frean, 1990; Karunanithi et al, 1994). However, often a trial and error procedure is used to obtain this number (see, e.g., Ranjithan et al, 1993; Rogers and Dowla, 1994; Tang and Fishwick, 1993, Saad et al, 1994; Rizzo and Dougherty, 1994).

In general, an increase in the number of hidden units increases the number of memory locations in the network for memorizing the information needed to solve problems within the bounds of interest. As the number of units and the memorizing capacity increases, the generalization performance also increases up to a point, after which further increase in the number of units decreases the network's generalizing capability. Also, even small increases in the number of hidden units increases the size of the network quickly. Increased network size often increases the time needed to train the ANN. For a large number of hidden units, the feasible space of the ANN training process often has a flat surface with several local optima, therefore, the optimization process of the performance index has a high potential of getting trapped in a local optimum. Furthermore, increasing the number of hidden units increases the adequate number of training examples. Therefore, determining the best number of hidden units is an important step of the network development process.

The number of training examples and their representativeness of the problem space also influence the performance of the network. Generally, as the number of training examples and the representativeness of the task to be performed increase, the generalization performance improves, since in such cases the network recognizes more patterns that describe the system represented by the network. However, the learning time required increases with the number of training examples. Often, there are some sets of input patterns that allow the network to achieve better generalization performance than that due to other sets of input signals. Such good training sets are more representative of the problem space and should be investigated and utilized to improve network performance. Like most types of black box modelling, some form of data normalization or transformation may enhance the network training process and the performance

characteristics.

The best learning algorithm by which the connection weights may be modified, and the most appropriate technique for optimizing the learning performance index, depends on the type of ANN architecture and the specific problem to be solved. Selecting the learning algorithm and the optimization technique often depends on experience and sometimes on initial simulation trials.

The extent of learning the network undergoes depends on the tolerance level of the difference between the best possible and the acceptable level of learning performance index. This tolerance level may be explicitly specified for a given network or implicitly incorporated by specifying the number of training cycles or iterations to be performed. In general, if the tolerance level is small or the number of iterations is large, the learning and generalization performances improve. However, for some networks, if the index is set too small or the number of iterations is set too high, overtraining may occur which can deteriorate the generalization performance.

5.2.6 Learning algorithms

ANN learning algorithms differ in the way in which information is represented in the weights. However, for all these algorithms, knowledge encoded in a network is distributed among a number of weights or a number of units after the learning process is complete. Storing the learned information in a large number of links or inputs is an important feature, since it allows the network a level of tolerance to individual component failure. Learning algorithms may be classified as supervised or unsupervised.

Supervised learning

For supervised or associative learning, the network is trained by supplying it with input vectors and corresponding output patterns. These input-output patterns may be provided from outside or within the network. The input vectors are propagated forward through the various layers until it reaches the outputs of the final layer. The generated outputs are then compared with the desired or target output values, and a function of the differences, termed as error function, is used to characterize the network's learning error. The value of this error is used as a basis for adjusting the weights. This process is repeated many times (of the order of hundreds or thousands) until an acceptable value of the error function is obtained or the specified number of iterations is reached. The objective of this training procedure is to minimize the value of the error function or to ensure that its value is below a certain threshold level. After successfully learning the training patterns, the network may be used to map input vectors to output patterns. The process by which the network error is used to modify the connection weights distinguishes the various forms of supervised learning.

The Backpropagation (BP) Algorithm (Appendix A) is the best understood and most commonly used supervised learning algorithm. Traditionally, it uses the generalized delta rule which performs steepest gradient descent on the total quadractic error $\mathcal{E}$, given as:

$$\mathcal{E} = \frac{1}{2P} \sum_{p} \sum_{k} (t_k^p - O_k^p)^2 \tag{5.7}$$

where

t_k^p = kth element of the desired output for training pattern p;

O_k^p = kth element of the network output for training pattern p; and

P = total number of training patterns.

Despite its apparent success, the BP learning algorithm has some weaknesses which include a long training process, network paralysis, and local minimum entrapment of the error minimization procedure. The most important deficiency of the BP learning algorithm is the long training process. Consequently, a great amount of research has concentrated on attempts to modify the algorithm to speed up the learning process by replacing the Steepest Descent Method with other direction search methods of minimization such as the Conjugate Gradient Method. Network paralysis occurs especially when sigmoid activation functions are used. For such cases, the total input of a hidden or output unit can reach very high positive or negative values, and the activation may be very close to one or zero (see, Figure 5.2a and *Equation* 5.4). Since the weight adjustments are proportional to $a_n(1 - a_n)$ (see, *Equations* A.12 and A.13), the training process can come to a virtual standstill. Transforming the input data can eliminate this problem. The problem of local minimum entrapment occurs because the error surface of a complex network is full of hills and valleys. This problem may be removed by (1) choosing random weights so that the magnitude of a typical net input to a given unit is a little less than one, (2) using a unit activation update mode that selects training patterns in a random order, (3) using probabilistic techniques, or (4) searching around an optimal point to ensure that it is not a local minimum.

Unsupervised learning

Unlike supervised learning, unsupervised (self-organizing) learning is used for problems which have training sets consisting of input vectors, but no corresponding desired out-

put signal. For such problems, the relevant information has to be found within the input vectors of the training examples. Examples of such problems include dimensionality reduction (e.g., principal component analysis), feature mapping, clustering, and vector quantization. These types of problems do not explicitly indicate to the network what the correct outputs should be for a given input vector. Instead, the network must identify the important characteristics (statistics) of the training examples and must adjust the weights in order to extract these characteristics. The procedure used to achieve this depends on the learning algorithm, which is based on some type of internal error measure. When a pattern is presented to the network, it is propagated to the outputs, and the error measure is then applied to extract the relevant statistics from these outputs. The connection weights are modified in order to minimize this error measure via the learning rule. The Hebbian learning rule is an example of an unsupervised learning algorithm. Details relating to unsupervised learning and the appropriate learning algorithms are not given in this thesis since the ANN application presented does not belong to the class of networks that requires self-organizing learning.

5.2.7 ANN testing

After a pre-specified training process for an ANN is over, the predicted outputs for the training examples are compared with the known targets. Often, this comparison is based on the quadractic error computed in *Equation* 5.7, and sometimes based on visual inspection of linear plots of the predicted versus target values. A small value of the error measure or a linear plot that is close to the 45^0 line indicates a good training performance.

If satisfactory training performance is obtained, the information stored in the links and units after the training process, are used to predict the output for testing example patterns that were not included in the training set. The predicted output for the testing examples are compared with known targets for these examples. The testing performance index is measured in a similar way as the training performance. Good testing performance indicates good generalization, or interpolation, capability of the ANN.

5.3 Estimating the level of criticalness of realizations in the MR model using ANN

This section presents a feed-forward ANN approach for estimating the level of criticalness of each of the realizations used to represent a stochastic water quality management system. The training example set for this ANN approach is based on the MR model solution using the IM Algorithm developed in Chapter 4. Each example, or pattern, j, in the training set represents the characteristics of Realization j and the target level of criticalness of this realization obtained from the MR model solution with the IM Algorithm. The number of realizations included in the MR model to obtain the training examples is selected to be smaller than that expected to be adequate for representing the stochastic management system. The goal is to utilize the learning and genaralization capabilities of ANNs in order to rank the criticalness of the realizations incorporated in an MR model that includes an adequate number of realizations S before solving the model. If satisfactory generalization performance of the ANN is obtained, the to-

tal computational time required to solve the MR model can be significantly reduced without deteriorating the results, especially for cases in which S is very large.

5.3.1 Architecture of the ANN

As discussed earlier in this chapter, the architecture of an ANN greatly influences its learning and generalization performance. In the prototype example for estimating U_j, it is necessary to identify and exclude the unimportant features, since these features could be sources of noise that may deteriorate the performance of the network. Although all important characteristics have to be included in the training examples, such information should be presented to the network in a form that will not unnecessarily increase the size of the network. To do this, it may be useful to process the important characteristics into lumped features, or to use dimensionality reduction techniques, such as principal component analysis.

For ranking the criticalness of realizations in a surface water quality management problem, the important information may include the waste treatment cost coefficients of the polluters and the following data or parameters in each reach of the river system: (1) low flow; (2) stream velocity; (3) stream temperature; (4) reaction rates; and (5) background water quality. Including all this information in even moderately sized systems could result in very large networks. Such large networks may require a large number of training examples to learn how to predict the U_j's and a large CPU time. These requirements are not favourable in terms of computational burden reduction of the MR model solution. Therefore, the solution of the management program based on a given realization is used to lump together the important characteristics of that real-

ization. Here, the optimal management decision variables and the level of criticalness at each Checkpoint $i \in \Theta_c$, w_{ij}, defined in *Equation* 4.8, are used to represent the input vectors of the training examples.

For this ANN approach, it is assumed that the value of U_j depends on the optimal management decisions of polluters upstream of the most critical checkpoint of Realization j and on the value of w_{ij} for this critical checkpoint. Thus, stream conditions downstream of the most critical checkpoint of a given realization are considered unimportant for determining the criticalness of this realization. Consequently, each training or testing pattern for this ANN may have a different number of active input units (i.e., inputs that represent stream conditions upstream of the most critical checkpoint), since the most critical checkpoint is generally not expected to be the same for all realizations. The most critical checkpoint for Realization j, MCC_j, is defined as that checkpoint with the smallest number of non-violations of water quality, under stream conditions represented by this realization, if all SIOP solutions based on other realizations included in the MR model are implemented in turn. Mathematically, MCC_j is given as:

$$MCC_j = i \qquad if \qquad w_{ij} = \min_{k \in \Theta_c} w_{kj} \qquad \forall \quad i \in \Theta_c \tag{5.8}$$

If MCC_j has more than one value (i.e., more than one critical checkpoint), the most downstream checkpoint among these critical locations is chosen as the most critical because the downstream critical location includes stream conditions for upstream checkpoints.

Since the number of examples in a training set may not adequately represent the joint probability distribution of the important features of the stochastic water quality system, the same example pattern in two different training sets may have significantly

different levels of criticalness. For example, if a training pattern has three active input components or units, a typical input vector may be represented as (7.5, 0.2, 3.7), where 7.5, 0.2, and 3.7 represent the quantitative indices for Features 1, 2, and 3, respectively. Assume this input vector constitutes one of 10 training patterns and has a normalized level of criticalness (i.e., expected output signal of the ANN) of 0.6. The 0.6 value of the normalized level of criticalness indicates that the vector (7.5, 0.2, 3.7) is the fifth to be dropped from the IM solution process. If this input vector now belongs to another set of 10 training patterns and represents the second realization to be dropped from the IM solution, its normalized level of criticalness is equal to 0.9. Therefore, to reduce training difficulties that may be presented by such a situation and to minimize the prediction error of the ANN, the rank (where the smallest value is assigned a rank of 1) of each component of each input vector, relative to the corresponding component of all other input vectors, is also included as an important feature in identifying the level of criticalness of the realizations in a given set. Thus, the relative importance of an input pattern in a given training set is included as an important feature in determining the level of criticalness with the ANN. Each of the important features for each training example is normalized as follows:

$$\mathcal{X}_{sj} = \frac{X_{sj} - X_{s,min}}{X_{s,max} - X_{s,min}} \tag{5.9}$$

$$\mathcal{Y}_{sj} = \frac{Y_{sj} - Y_{s,min}}{Y_{s,max} - Y_{s,min}} \tag{5.10}$$

where

$\mathcal{X}_{sj}$ $= s$ th normalized feature for training example j;

X_{sj} $= s$ th feature for training example j;

$X_{s,min}$ = the minimum value of feature s within the set of training examples;

$X_{s,max}$ = the maximum value of feature s within the set of training examples;

$\mathcal{Y}_{sj}$ = normalized rank of feature s for training example j;

Y_{sj} = rank of feature s for training example j;

$Y_{s,min}$ = the minimum rank of feature s within the set of training examples; and

$Y_{s,max}$ = the maximum rank of feature s within the set of training examples.

To reduce the number of input units, for each training example, each important feature and its corresponding rank are combined into a single feature. These new combined features are transformed to lie between -2 and 2 in order to enhance the ANN training process. The range for the transformed values is obtained through a trial and error process with the objective of minimizing training error and time. The transformed values constitute the components of the input vectors of the training examples and are computed in *Equation* 5.11 for active inputs that represent features upstream of the most critical checkpoint, MCC_j, of the j th realization.

$$\mathcal{Z}_{sj} = 4\sqrt{\mathcal{X}_{sj}\mathcal{Y}_{sj}} - 2 \tag{5.11}$$

where

$\mathcal{Z}_{sj}$ = the value of the input for feature s for training example j.

For inactive input units that correspond to stream characteristics that are downstream of MCC_j, the components of the input vectors are set to a large negative number.

The ANN architecture shown in Figure 5.3, PANN1, classifies the input units into 7 groups. For a given training example, Input_s represents the active units if MCC_j is the s th most upstream checkpoint in the set Θ_c. For example, if the index s, of the most

critical checkpoint for Realization j, is equal to 1, only two BOD waste polluters on the Willamette River are upstream of this checkpoint, therefore three input units, that represent the characteristics of removal levels (i.e., magnitude and relative importance among the set of realizations used for training or testing) for these two polluters and the value of w_{ij} for this critical checkpoint, constitute the active inputs. The units of each group of active input units are represented by zeros in the lower rectangular boxes shown in Figure 5.3. For the same reason for incorporating the relative importance of the components of input vectors as input features, an output signal for each training example is obtained based on (1) the optimal objective function of the MR model (using the IM Algorithm) just before the realization which represents this training example is removed from the model, and (2) the rank of the objective function value obtained in (1) above (where the smallest value is assigned a rank of 1). The values from (1) and (2) are normalized to lie between 0 and 1 using equations similar to *Equations* 5.9 and 5.10. The output signals are then computed as:

$$t^j = 0.8\sqrt{{}_o\mathcal{V}_{jr}\mathcal{V}_j} + 0.1 \tag{5.12}$$

where

t^j = output signal for training example j;

${}_o\mathcal{V}_j$ = normalized optimal objective function value of the MR model when example j is the most critical realization;

${}_r\mathcal{V}_j$ = normalized rank of ${}_o\mathcal{V}_j$.

Although the outputs of the sigmoid squashing function used for this research are in the range [0,1], the target output signals in the training examples (*Equation* 5.12) are transformed to be in the range [0.1,0.9]. The coefficients 0.8 and 0.1 in *Equation* 5.12

transform the output signals from the range [0,1] to [0.1,0.9]. This transformation is used to allow for both interpolation and extrapolation if the trained network is applied to example patterns that are not included in the training set. Since the maximum and minimum output signals from an ANN are one and zero, respectively, the extrapolation capability is necessary to allow the level of criticalness to be predicted for patterns that are more or less critical than each of the training examples. The transformation also improves the network performance. This may be explained by the fact that for the output range [0.1,0.9], the sigmoid activation function is sensitive to inputs, while outside this range it is insensitive to inputs. Experience obtained through analyzing the ANN network confirms the improved performance when this transformation is used.

The set of hidden units connected to the active input units is henceforth referred to as active hidden units. The number of active hidden units varies with the location of the most critical checkpoint for a given training example. In this thesis, 10 hidden units are utilized by each of Hidden_1, Hidden_3, Hidden_4, and Hidden_7, while only 1 hidden unit is specified for each of the remaining groups of hidden units. These numbers are obtained through insight gained into the water quality system regarding the approximate proportion of training examples that correspond to the different groups of active input units and through a trial-and-error process. This insight was obtained by examining the training example set. If a small number of training examples correspond to a given group of active units, then for these types of examples, a small number of hidden units and links may be used to store information regarding the functional relationship between the characteristics of a realization and the level of criticalness. The factors considered in selecting the number of hidden units are the network's learning and generalization performances and its training time.

5.3.2 Training and Testing

A BP Algorithm that uses the Conjugate Gradient (CG) Method with restarts, instead of the traditional Steepest Gradient Descent technique, is utilized to minimize the training error. Although the CG Method is known to perform better than the Steepest Gradient technique, in terms of speed of convergence and insensitivity to instabilities and oscillatory problems (Johansson et al. 1992), the possibility of the optimization process becoming stuck in a local minimum exists for the CG Method. Therefore, to reduce the likelihood of the above problem, after a minimum point is identified, other neighbourhoods are searched to ensure that the identified point is not a local minimum. If a local minimum point is located, the search process leaves the valley that contains this point and starts searching for the global optimum again.

One thousand patterns are used to train the network and another 1000 patterns are used to test the generalization performance of the network. Each of these sets of patterns is obtained by solving the MR model that incorporates 1000 realizations using the IM Algorithm. Based on a critical observation of the characteristics (e.g., optimal cost, decision and dual variables, and reliability) of the patterns of the example sets, and the cost-reliability plots of these sets, each training or testing example set is divided into two classes before learning or generalization begins. The optimal objective function value of the management program for each realization (or pattern in an example set) is used as the index for this classification. In a given example set, all patterns with the value of the above index greater than or equal to \$0.91 million/year form one class, while the others constitute the second class. This classification is a very simple approach of clustering and ensuring that similar patterns are included in a given training or testing

process. Such an approach may improve the training time significantly, especially for cases in which the number of training patterns is large.

To prevent overtraining the network, a tolerance level for the maximum absolute difference between the target output of training example j, t^j, and the corresponding output signal from the ANN, O^j, is set to 0.0005. This tolerance level gives the interval of acceptance for agreement between a target output and the output signal of the network or the degree to which a 'near miss' will count as a 'hit'. Thus, if an output signal lies in this interval, the error in the signal is propagated back as zero. For example, if the target output for training example j is 0.80 and the tolerance level is 0.0005, then any output signal for this example in the interval [0.7995,0.8005] is regarded as 0.80.

There are two ways in which the termination criterion for the network training process may be specified: (1) it can be given in terms of the acceptable maximum value of the error function (i.e., the value of $\mathcal{E}$ from *Equation* 5.7) or (2) it may be defined in terms of the maximum number of training iterations or cycles to be performed. Criterion (2) is used in this thesis, since the time required to train the network could be estimated *a-priori* based on the CPU time used by the first few iterations. On the other hand, for Criterion (1), the time needed to train the network, which is a very important factor for determining the architecture and the number of training examples to utilize in this prototype network, cannot be predicted. This time is obtained only after the training process converges or terminates. In such cases, for the acceptable maximum value of $\mathcal{E}$ specified, the training may proceed for a very long time. Furthermore, for the case of learning to estimate the level of criticalness, the error value $\mathcal{E}$ may not be the best measure of the network performance. In this case, the goal is to obtain output signals that have the 'same' order of ranking as the corresponding target outputs.

To test the PANN1 network, the learning and generalization performance of the network are assessed based on Kendall's rank correlation coefficient and on visual inspection of plots between the predicted and actual level of criticalness. The higher the value of Kendall's rank correlation coefficient and the closer the predicted-actual level of criticalness plot is to the 45^0 line, the better the learning or generalization performance. Although Kendall's rank correlation coefficient is more nonparametric, and uses a weaker property of data, than Spearman's rank correlation coefficient, it is a better representation of the performance of the network in terms of correctly identifying the elements in Φ_R at each reliability level. While Spearman's coefficient is based on the numerical difference of ranks, Kendall's coefficient uses only a relative ordering of ranks which determines whether a given data point (i.e., level of criticalness for a realization) has a higher or lower estimated rank than the actual rank. This is appropriate for assessing the ANN performance in determining whether or not a given realization belongs to the set Φ_R, at all reliability levels.

For a given training or testing case, Kendall's rank correlation coefficient is calculated for all patterns that belong to the corresponding case. For example, if the testing case consists of 1000 patterns, Kendall's rank correlation coefficient is calculated based on the actual and predicted levels of criticalness of each of the 1000 patterns. However, for some water quality management systems, the performance of the ANN in predicting the highly critical realizations is needed, since these realizations are utilized to obtain the cost-reliability relationship, often required by decision makers, at high cost and high reliability levels. To assess the performance of the ANN in estimating the level of criticalness of these realizations, the Kendall rank correlation coefficient may be computed for patterns that represent such realizations.

5.4 Performance of prototype ANN in predicting level of criticalness Uj

The learning and generalization performance of the ANN network, PANN1, described in Section 5.3, after 1000 training iterations, are shown in Figures 5.4(a) and 5.4(b), respectively. Visual inspection of these graphs and Kendall's rank correlation coefficient between the estimated and actual level of criticalness suggests that the PANN1 architecture, the number of training examples, and the extent of training may be adequate for estimating U_j. For these graphs, most of the points plot very close to the 45^o line and Kendall's rank correlation coefficient is high. Therefore, the connection weights $\mathcal{W}_{mn}$ obtained after 1000 training iterations may be adapted and used to estimate the level of criticalness of realizations included in the MR model.

As expected, the learning performance of the network is better than the generalization or testing performance. While Kendall's rank correlation coefficient between estimated and actual values of U_j is 0.9823 for the training example set, its value is 0.9763 for the testing set. Furthermore, there are few testing example patterns whose estimated level of criticalness are very different from the actual level of criticalness, as indicated by the number of points that lie very far from the 45^o line. The poor generalization performance with respect to some testing examples, e.g., Realization A in Figure 5.4(b), suggests that the training example set includes patterns that do not have similar characteristics as those of the poorly performing testing patterns. In such a case, one way of improving the generalization performance is to include the poorly performing testing patterns in the training example set. However, for the application in this research, such an effort to improve the generalization performance is not pur-

sued, since the overall network performance is assumed to be adequate for identifying the important realizations needed to obtain efficient solutions and develop objective function-reliability trade-off relationships.

To illustrate the superiority of the ANN application compared to the use of optimal objective function values from the SIOP solution in estimating the level of criticalness, the rank of the optimal objective function value for each realization in the SIOP model, D_j, is plotted against the actual level of criticalness obtained from the IM algorithm, U_j, in Figures 5.5(a) and 5.5(b). These may be compared with the results shown in Figures 5.4(a) and 5.4(b). For both training and testing example sets, the ANN method produces plots that are generally closer to the 45^o line (see, Figures 5.4(a) and 5.4(b)) and have higher Kendall's rank correlation coeficient than those based on the ranks of the optimal objective function values from the SIOP solutions. The correlation coefficients of the training and testing example sets for the PANN1 network are 0.9823 and 0.9751, respectively, while those for the method based on optimal objective function values for single realizations are 0.9171 and 0.9150.

As indicated earlier in this chapter, the performance of an ANN generally varies with the extent of learning or the number of iterations the network undergoes. Most of the salient input patterns in the training set may be learnt by the network during the training process well before the error criterion, which is normally assigned a very low value, attains its desired value or the training process converges. Thus, there may not be significant advantage in network performance by allowing the training process to continue after a certain number of training iterations. Figure 5.6 shows Kendall's rank correlation coefficient and the value of the error criterion $\mathcal{E}$, for training iterations of 10 to 26,091 for the PANN1 network. The training process converges after 26,091

iterations. From the graph in Figure 5.6, Kendall's rank correlation coefficient and $\mathcal{E}$, increase and decrease, respectively, quickly as the number of iterations increases from 10 to 1000 for both the training and testing example sets. After 1000 training iterations, these values remain fairly constant until the training process converges. The CPU time for the training process varies approximately proportionally to the number of iterations.

The observations in the previous paragraph suggest that after 1000 training iterations the performance of the network does not gain any significant improvement from further training, although substantial computer time resources may be required. Linear plots of the performance of the network for 2000 and 26,091 training iterations are shown in Figure 5.7(a) and 5.7(b) for the training set and in Figures 5.8(a) and 5.8(b) for the testing set. A critical visual comparison of these plots with the corresponding plots for 1000 training iterations (Figures 5.4(a) and 5.4(b)) supports the assertion that 1000 iterations is adequate for training the PANN1 network.

The performance of a three layer fully connected network (i.e., each input unit connected to each hidden unit and each hidden unit connected to each output unit) is investigated and compared with that of the PANN1 network. This fully connected network is henceforth referred to as PANN2. The PANN2 network uses the values and ranks of the optimal cost, decision variables (waste removal levels for 10 dischargers), and dual variables (for constraints that ensure acceptable water quality at the seven checkpoints that belong to Θ_c) from the SIOP solution, and w_{ij} for all seven checkpoints to represent the important input characteristics of a given realization j. The values of the input units are computed using *Equations* 5.9, 5.10, and 5.11, and the target output values are obtained as in *Equation* 5.12. The architecture of the PANN2 network consists of 25 input units, 15 hidden units and one output unit. The number of hidden units was

optimized using a trial-and-error method. The number of hidden units was gradually increased from one until no significant improvement in the training performance was observed.

Figures 5.9(a) and 5.9(b) show the plots of estimated values of criticalness from the PANN2 network versus the actual values for the training and testing example sets, respetively, for a case in which 1000 training iterations are performed. Kendall's rank correlation coefficient for the training and testing example sets are 0.9773 and 0.9718, respectively. For the same training and testing exaple sets, when these coefficients are compared with the corresponding coefficients for the PANN1 network (0.9823 and 0.9751, respectively) the PANN1 network appears to have a slightly better training and generalization performance ability. Visual comparison of the training and testing performance plots for the PANN2 (Figures 5.9(a) and 5.9(b)) and PANN1 (Figures 5.4(a) and 5.4(b)) also suggests that the PANN1 network is better for predicting the level of criticalness U_j. The observation of good generalization performance of the PANN1 network when compared with the generalization capability of the PANN2 network suggests that, for a given realization, the optimal decision variables upstream of the most critical checkpoint and the probability of water quality violations at this critical checkpoint are important characteristics for predicting the level of criticalness.

The CPU time used to generate each of the training and testing example sets is approximately one hour on the University of Manitoba Sparc2 UNIX system and the training time for 1000 iterations was approximately 13 minutes for the PANN1 network and 11 minutes for the PANN2 network. The training time for the networks may be reduced since the learning process can be implemented on a parallel computer system. Thus, the time used to generate the training and testing patterns for the networks may

be a significant proportion of the total time needed by the ANN approach to estimate the level of criticalness of realizations in the MR model. Because the CPU time needed to generate a training set increases exponentially with the number of examples in this set, the possibility of using a number of smaller training sets instead of one large set is investigated. For example, the performance of a network that utilizes two training sets each containing 500 example patterns is compared with that of one training set of 1000 example patterns.

Figure 5.10 shows Kendall's rank correlation coefficient and the error measure $\mathcal{E}$ plotted for the number of training sets that constitute the 1000 training examples used in the PANN1 network. In this graph, the same 1000 realizations are used to generate the training sets for each scenario. For example, in the scenario with four training sets, the 1000 realizations are divided into four equal groups and the MR model is solved for each group of 250 realizations to obtain the four training sets. The total time for generating these small multiple training sets is significantly less than that required to generate the patterns for a single large set. For the multiple training set scenarios analyzed, the four set case required a total of 20 minutes to generate the training sets compared to one hour for the single set of 1000 training examples. Figure 5.10 shows that both training and generalization performance of the PANN1 network generally worsen as the number of multiple sets increases. However, the deterioration in performance is small when the number of training sets is not greater than four.

Linear plots of the estimated level of criticalness versus the actual level of criticalness for both training and testing example sets for the two, four, and ten multiple training set scenarios are shown in Figures 5.11 - 5.13. Visual inspection of these graphs clearly shows the extent to which the network performance degenerates with increase

in the number of multiple training sets. This suggests that important information is increasingly lost as the realizations used to obtain the training examples are divided into a larger number of smaller sets.

5.5 Using estimated levels of criticalness to solve the MR model

After adequate learning and generalizing network performance is achieved, the connection weights can be adapted and used to estimate the level of criticalness of each realization included in a given MR model. If the estimation is exact, the set Φ_R in *Equation* 3.20 can be obtained accurately as:

$$\Phi_R = \{G_j : \hat{U}_j \geq (1-R)S + 1\} \quad \forall \;\; 0 < R \leq 1; \;\; R \times S \in \{interger\} \qquad (5.13)$$

where

$\hat{U}_j$ = estimated level of criticalness for Realization j, based on the PANN1 network, and

all other symbols are already defined.

However, in a practical situation, the estimation of the level of criticalness is often not perfect. Therefore, a successive solution algorithm is used, that screens the elements of Φ_R obtained by *Equation* 5.13 to remove any group of realizations that might have been identified as the set of binding realizations at some earlier solution stage. This screening process helps to prevent situations in which realizations with underestimated levels of criticalness (i.e., $\hat{U}_j > U_j$) may continue to appear in the set Φ_R obtained from

Equation 5.13, even though these realizations should have been dropped from the MR model at an earlier solution stage. By screening out sets of realizations that control the optimal MR model solution at a higher reliability level than R, the likelihood of obtaining inefficient solutions is minimized.

The method developed is referred to as a Neural Network based IM (NNIM) Algorithm. This algorithm is given below:

The NNIM Algorithm

Step 1 Use adapted matrix of ANN connection weights to obtain $\hat{U}_j \quad \forall \quad j = 1, \ldots, S$;

Step 2 Set $R = 1.0$ and $\Phi_{1.0} = \{G_j \quad \forall \quad j = 1, \ldots, S\}$;

Step 3 Obtain the set of potential critical realizations $\mho_R$ from Φ_R by selecting a small number, κ, of the most critical realizations at each Checkpoint $i \in \Theta_c$. Mathematically, $\mho_R$ is given as: $\mho_R = \bigcup_{i \in \Theta_c} \alpha_i$; where $\alpha_i = \{G_j : w_{ij} \leq \Lambda_{iR}^{\kappa} \quad \forall \quad j \in \Phi_R\}$ and Λ_{iR}^{κ} is the κ th highest value of $w_{ij} \quad \forall \quad j \in \Phi_R; \quad i \in \Theta_c$. Replace the water quality constraint set in *Equation* 3.20 by $P_{ij}(\mathbf{X}) \geq b_{ij} \quad \forall \quad i \in \Theta_c; \quad j \in \mho_R$ and solve the resulting MR model. Identify the set of binding realizations, β_R, for this solution;

Step 4 Evaluate the reliability of the solution obtained in Step 3 with respect to all realizations included in the MR model. This is necessary since the value of R in this algorithm may not be equal to the actual reliability of the solution, if the estimated levels of criticalness for all realizations are not the same as the corresponding actual levels of criticalness. If the reliability level obtained is less than a pre-assigned minimum threshold $TREL$, STOP;

Step 5 Obtain a new value of R at a new stage of solution as: $R = R - \Delta R$ where ΔR is a small positive number and $\Delta R \times S$ is an interger;

Step 6 Obtain Φ_R using *Equation* 5.13. Screen Φ_R to remove sets of binding realizations that have already been identified in Step 3. Go to Step 3.

Unlike Morgan et al.'s heuristic and IM Algorithms, the NNIM Algorithm does not have to solve the MR model at S stages in order to generate cost-reliability trade-off relationships for reliability levels between 1 and 0. Depending on the value of ΔR used at a given stage, a number of realizations, $\Delta R \times S$ are dropped from the model at this stage. For example, if the cost-reliability relationship at all reliability levels is of interest and 1000 realizations are incorporated in the MR model, the IM and Morgan et al.'s (1993) algorithms will solve the model at 1000 stages reducing the size of the model by one realization at each solution stage. However, if the sensitivity of cost to reliability and other factors require the accuracy of reliability values to be on the order of 0.005, then ΔR can be set to 0.005 so that 5 realizations can be removed from the MR model after each NNIM solution stage. In this case, only 200 solutions need to be produced by the NNIM algorithm to generate the cost-reliability relationship at reliability levels between one and zero. Also, the NNIM Algorithm does not require solving the MR model several times at a given stage in order to create candidate solutions for the next stage. However, to produce efficient solutions, good generalization performance of the ANN used to predict the level of criticalness of the realizations included in the MR model is required. Furthermore, the value of ΔR should not be too large. This thesis does not investigate any generalized procedure for selecting the best value of ΔR. Although a high generalization performance (i.e., high Kendall rank correlation) is attained by

the ANN used for the case study in this research, the threshold of ANN generalization performance that is acceptable in the NNIM Algorithm is not examined.

5.6 Discussion of results

The NNIM Algorithm is used to solve the MR model for the LC program for four different scenarios in which 2000 realizations are incorporated in each scenario. These four scenarios are the same as the sets SIM_1, SIM_2, SIM_3, and SIM_4 that were generated in Chapter 4. Based on its performance presented in Section 5.4, the PANN1 network that uses a single training set of 1000 example patterns and 1000 training cycles is utilized to obtain the matrix of connection weights adapted to estimate the level of criticalness of realizations incorporated in the MR model. The PANN1 network is tested with another set containing 1000 example patterns. There may be a best value of κ to use in the NNIM Algorithm to obtain efficient solutions while minimizing the computational time requirement of the NNIM Algorithm. This best value of κ may be obtained using a trial-and-error procedure, but this is not investigated in this thesis. The value of κ is arbitrarily set to 7 for this research. Thus, to solve the MR model at a given stage, the NNIM Algorithm selects and incorporates a maximum of 49 potentially important realizations (i.e., 7 realizations for each of the 7 critical checkpoints) in a single optimization model to obtain an efficient solution.

Figure 5.14 compares the trade-off relationships between cost and reliability obtained using the IM and NNIM Algorithms to solve the MR model that incorporates each of the sets SIM_1, SIM_2, SIM_3, and SIM_4. The graphs in this figure suggest that, in general, the two algorithms produce similar cost-reliability trade-off relationships.

In particular, for SIM_2, SIM_3, and SIM_4 (Figures 5.14(b) - 5.14(d)), the NNIM Algorithm generates solutions that lie on the curve produced by the IM Algorithm. However, the solutions of the two algorithms for SIM_1 (Figure 5.14(a)) are not as close as those for the other sets at reliability levels greater than 0.85. For SIM_1, at reliability levels less than 0.95, the solutions based on the NNIM Algorithm are slightly less efficient than those based on the IM Algorithm, but at reliability levels between 0.90 and 0.85, the NNIM solutions are more efficient than the IM solutions. Even for SIM_1, the difference observed between the solutions obtained using the IM and NNIM Algorithms is small enough for the two solutions to be considered similar.

The total CPU time used to produce the cost-reliability trade-off graphs, based on the IM and NNIM Algorithms is significantly different. On the University of Manitoba Sparc2 Unix system, while the IM algorithm utilizes approximately 42 hours to produce the four trade-off curves, the NNIM Algorithm uses about 5.5 hours to obtain the cost-reliability relationships. The total time of 5.5 hours utilized by the NNIM Algorithm is comprised of 2.0 hours used to generate the training and testing example sets for the PANN1 network, 0.3 hours used to train the network, and 0.8 hours used to solve the MR model for each of the four cases in which SIM_1, SIM_2, SIM_3, and SIM_4 are incorporated.

The process of obtaining the matrix of connection weights (i.e., generating training and testing examples sets and training the ANN) used to estimate the level of criticalness constitutes a significant percentage (about 42%) of the total CPU time required by the NNIM Algorithm to generate the four cost-reliability relationships in Figure 5.14. This percentage will be higher (about 74%) if only one trade-off relationship is developed. The analyses of stochastic water quality management systems often require several

simulation runs. The matrix of connection weights adapted after training the network may be used for each of these runs if the NNIM Algorithm is utilized. In such cases, the computational time efficiency of the NNIM Algorithm may be enormous when compared to the CPU time required by the IM Algorithm.

To obtain a better convergence of the output from the MR model than that shown in Figure 4.1 for 2000 realizations, four sets of 4000 realizations each are incorporated in the MR model in turn. These sets are referred to as SIMUL_1, SIMUL_2, SIMUL_3, and SIMUL_4. Figure 5.15 shows plots of four cost-reliability trade-off relationships produced by the NNIM Algorithm for the LC MR model that incorporates SIMUL_1, SIMUL_2, SIMUL_3, and SIMUL_4. The connection weights adapted from the PANN1 network are utilized in the NNIM Algorithm. The four cost-reliability trade-off plots virtually lie on the same curve. This indicates a very good convergence of the output from the MR model. The CPU time used to produce the results for each of these trade-off relationships, excluding the time utilized to obtain the matrix of connection weights, is approximately 2.9 hours. Even if the time utilized to obtain the connection weights (i.e., 2.3 hours) is added to the CPU time utilized by one set of 4000 realizations, a total time of 5.2 hours is obtained. This total time is less than the CPU time of 10.5 hours used by the IM algorithm to solve the MR model that incorporates 2000 realizations.

All the cost-reliability trade-off relationships developed so far in this thesis are for an LC program. However, in order to develop politically acceptable and implementable pollution control plans, many water quality management modelling attempts have considered equity options since fairness among waste dischargers, in terms of their pollution reduction effort, has historically been considered as an important management objective. Also, because a trade-off exists between waste treatment cost and equity, equity

options may be used to generate alternative water quality management decisions and to gain insight into the management system. Therefore, trade-off relationships between cost and reliability are developed for cases in which equity goals are included in the MR model described in *Equation* 3.19 - 3.23.

An acceptable measure of the equity among waste dischargers for a given water quality management program has been the focus of some discussion (Brill et al., 1976 and Chadderton and Kropp, 1985). Differences in water quality impacts, total waste discharged into a river system, effluent concentration, fraction waste removal levels, and the distribution of the of waste treatment cost among the polluters have been used to calculate equity indices (Chadderton and Kropp, 1985). Measures of equity based on waste removal levels are the easiest to apply and the most commonly used in direct regulation water quality management models (Brill et al. 1976). Equity measures based on waste removal levels assume that a uniform fraction of waste removal among the dischargers is fair. For such measures, UT model solutions, which specify the same removal effort for all polluters, are considered to be the most attractive from the point of view of equity.

In this research, the difference between the maximum and the minimum waste removal levels of point source polluters is used as a measure of equity. This simple measure of equity is used to demonstrate the capability of the MR model and the solution approaches, to include an equity objective in a water quality management model. The smaller the measure of equity, the "fairer" the pollution abatement effort of the polluters. This equity goal is represented in a management model as:

$$x_k - x_{min} \geq 0 \tag{5.14}$$

$$x_k - x_{max} \leq 0 \tag{5.15}$$

$$x_{max} - x_{min} \leq EQUITY \tag{5.16}$$

where

x_k = fraction waste removal level for polluter k;

x_{max} = maximum fraction waste removal level performed;

x_{min} = minimum fraction waste removal level performed; and

$EQUITY$ = desired equity level.

To include equity goals in the MR model, *Equation* 3.21 is replaced by *Equations* 5.14-5.16.

Figure 5.16 shows plots of the trade-off between cost and reliability for different equity levels. All the solutions in this figure incorporate 4000 realizations. Apart from the UT solution which is based on the SIOP model, all the other solutions are obtained using the NNIM Algorithm to solve the MR model. Recall from Chapter 3 that for an UT program, the solution of the traditional SIOP model is the same as that produced by the MR model. The MR model that produces the LC solution does not have any equity constraints, *Equations* 5.14-5.16, while the UT solution has $EQUITY = 0$. The LC and UT solutions represent the upper and lower limits of the equity measure. The 0.5 and 0.3 equity solutions ensure that at each reliability level, the difference between the highest and lowest waste removal levels does not exceed 0.5 and 0.3, respectively.

At a given reliability level, the difference between the waste treatment cost of the solutions, for any two equity scenarios, represents the cost of maintaining fairer pollution control decisions for the more equitable solution (i.e., the cost of equity). It is observed from Figure 5.16 that, in general, the cost of equity decreases with a decrease in the

reliability level of the solutions. This is expected since higher waste removal levels are required to meet the water quality goal at higher reliability levels. These higher removal levels correspond to steeper portions of waste treatment cost curves, where the marginal cost of waste treatment is higher.

Figure 5.16 also shows that the cost of equity at a given reliability level may be nonlinear with respect to changes in equity. For example, at a reliability level $R = 0.8$, the cost of equity between the LC ($EQUITY = 0.63$) and 0.3 equity solutions is \$ 1.58 million/year, while that between the LC and UT ($EQUITY = 0$) solutions is \$ 7.41 million/year. Thus, the cost of equity for a 0.63 improvement in the index for fairness is about 4.7 times that for a 0.33 improvement in the index for fairness. The information in Figure 5.16 about cost, reliability, equity, and the relationships among them is important to decision makers who often have to assess such information and other unmodelled issues in order to implement pollution control decisions for complex water quality systems.

Although the measure of equity used in this research may not be a true reflection of fairness among polluters in a real system, it can be used to gain some understanding of the physical water quality management system. For example, if the cost of equity for using a UT program instead of an LC program is high at the reliability level of interest, then the distribution of waste treatment cost among polluters has a high variability due to large differences in the efficiency of the waste treatment operations of the polluters. A high variance of the distribution of waste treatment cost may also be caused by stream conditions that are significantly different for the sections of the river where the polluters are located. While some polluters may be situated in a high assimilative capacity section of the river, others may be located in critical reaches where a high level

of treatment effort is required to achieve acceptable water quality standards.

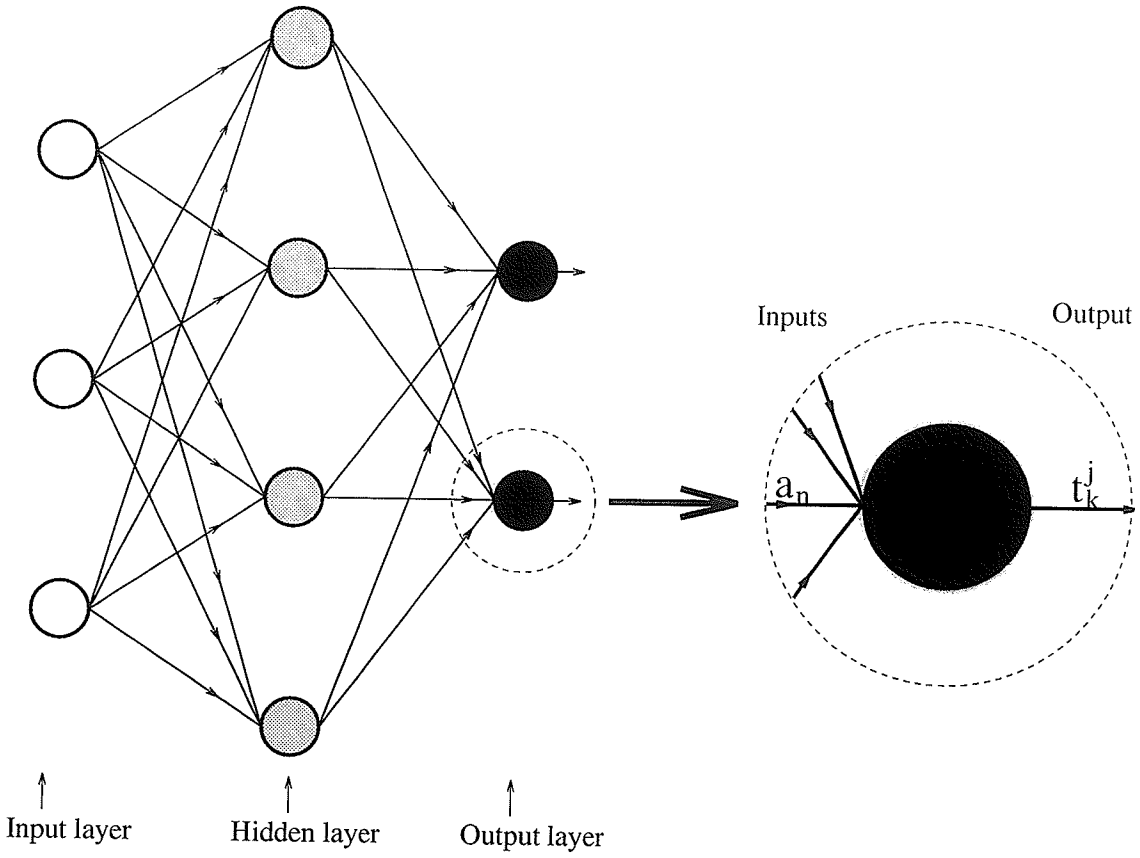


Figure 5.1: Basic ANN architecture

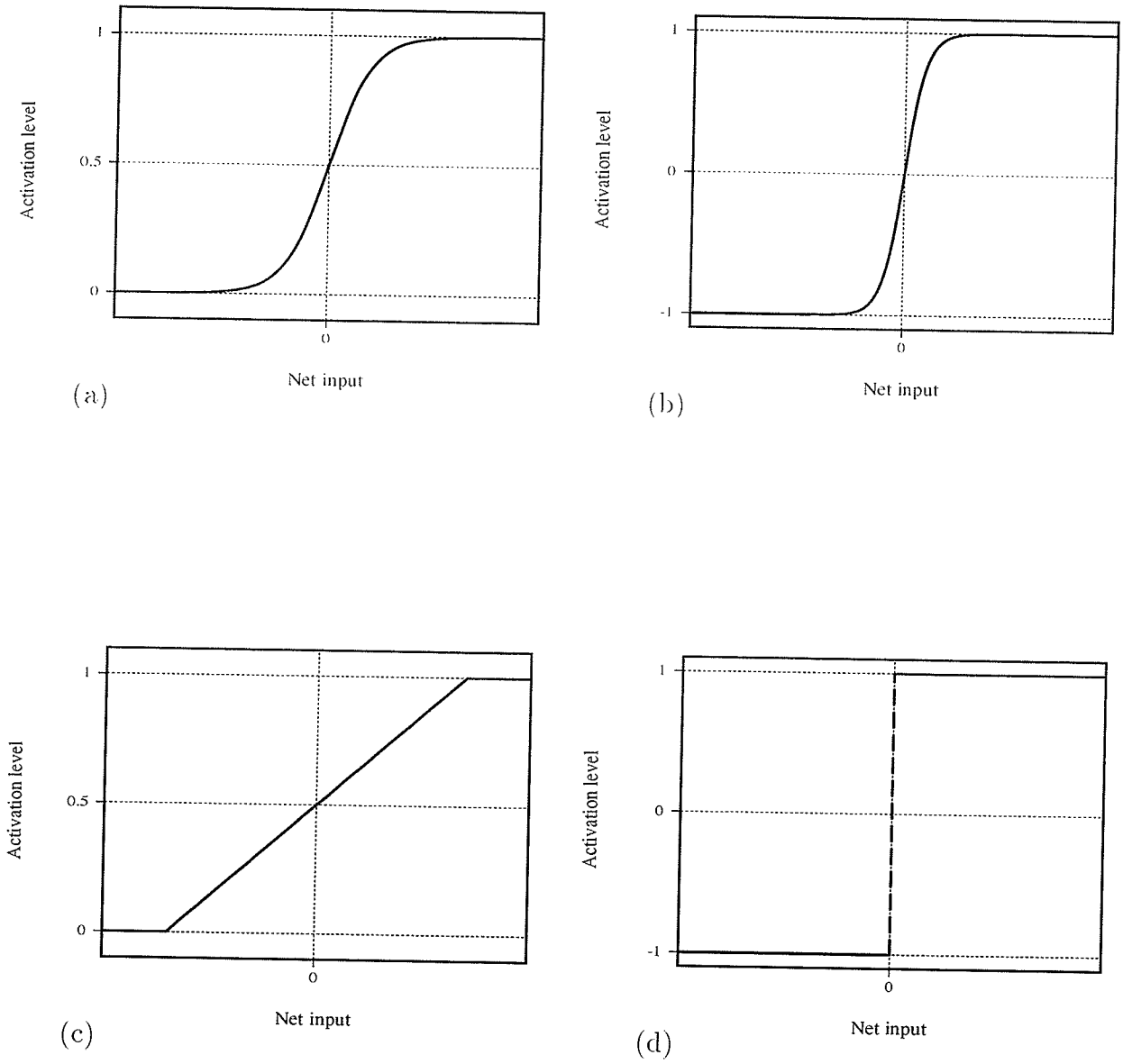


Figure 5.2: Typical activation functions: (a) sigmoid; (b) hyperbolic; (c) semi-linear; (d) sgn

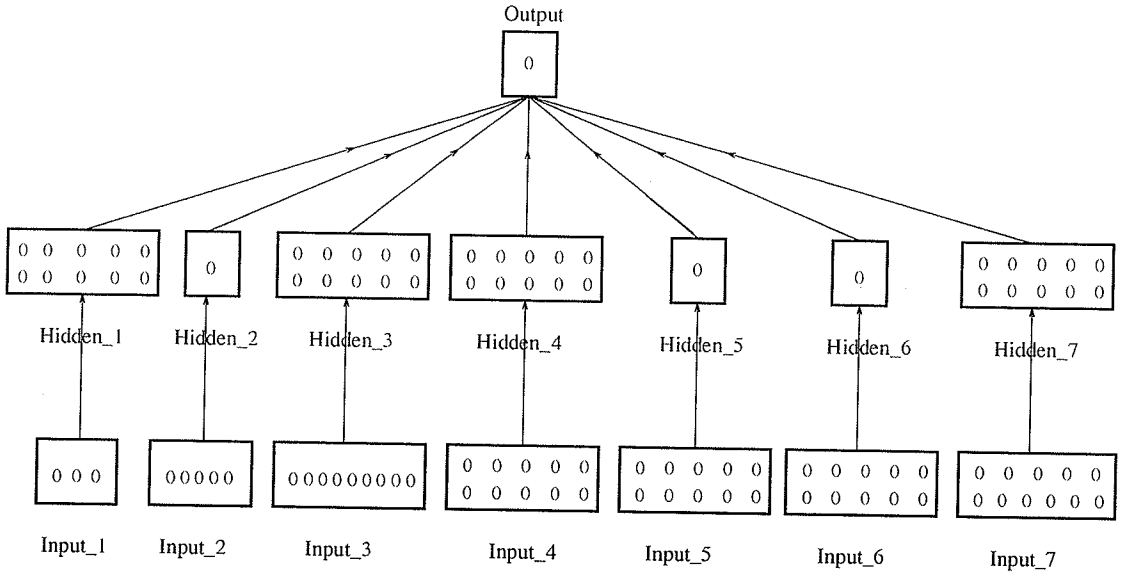
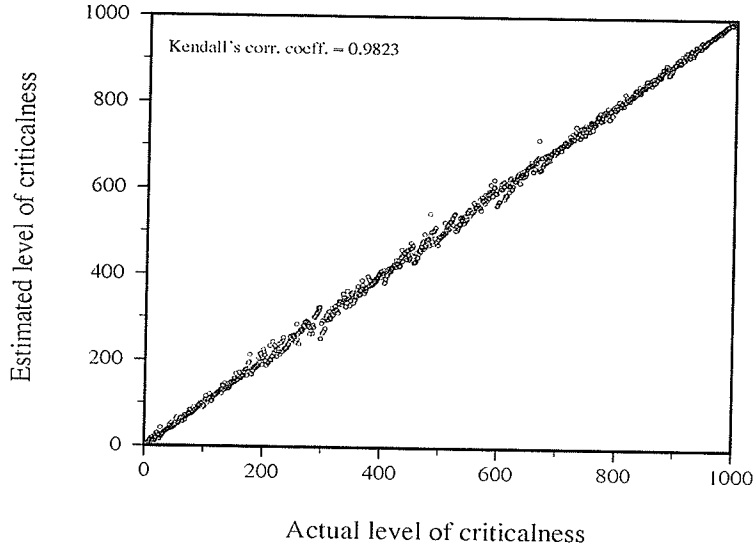
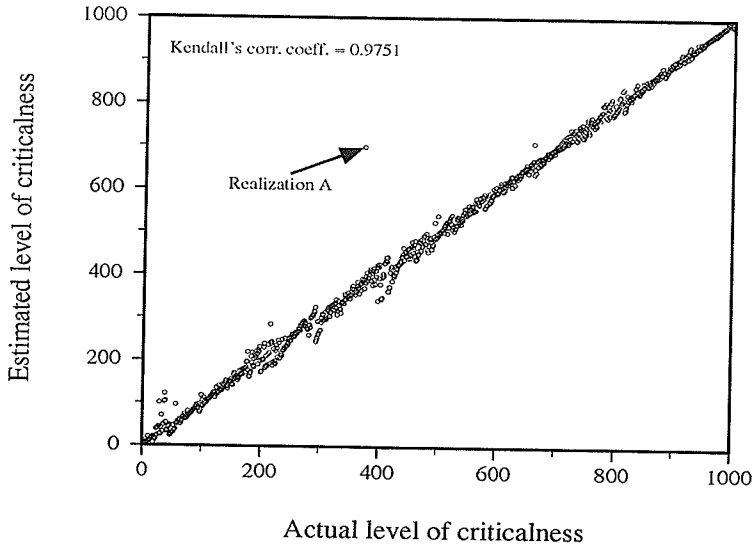


Figure 5.3: Prototype ANN architecture (PANN1). Each unit in Input_s is connected to every unit in Hidden_s, and all hidden units are connected to the output unit

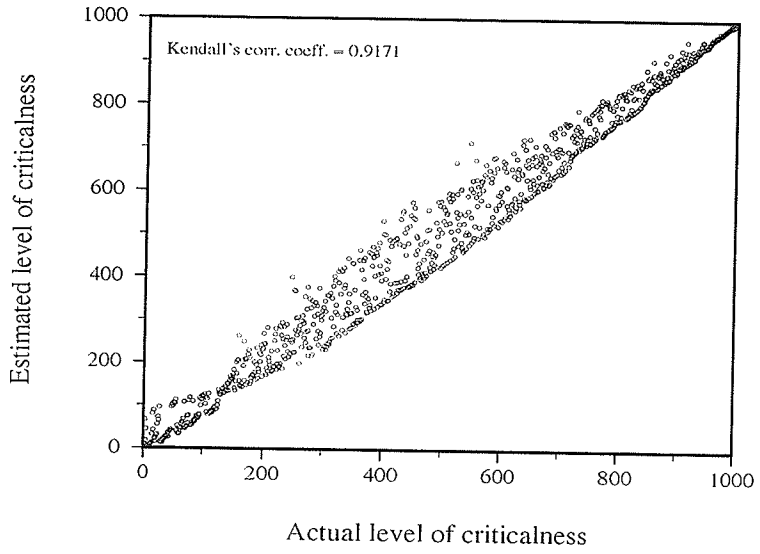


(a)

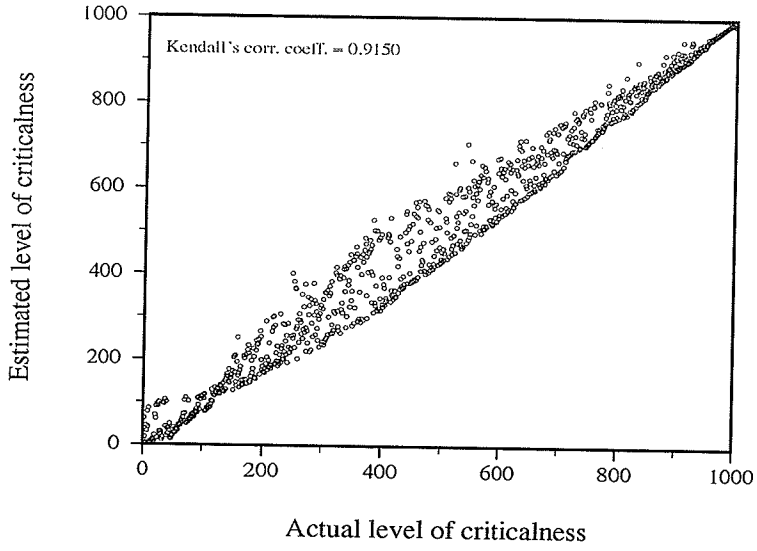


(b)

Figure 5.4: Performance of PANN1 network after 1000 iterations: (a) Training set; (b) Testing set



(a)



(b)

Figure 5.5: Plot of estimated level of criticalness from the traditional SIOP solution versus actual level of criticalness for 1000 example patterns: (a) Training set; (b) Testing set

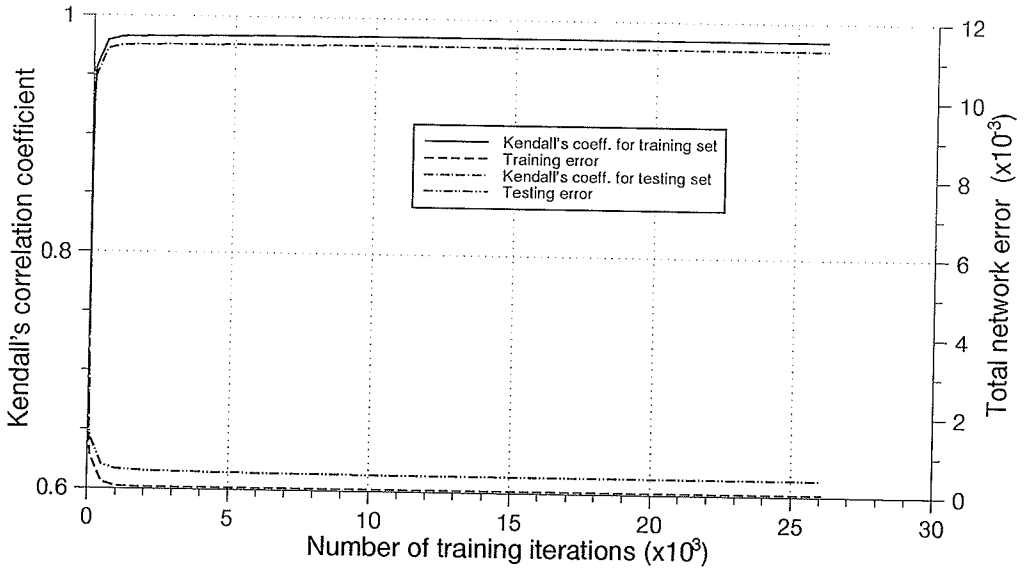
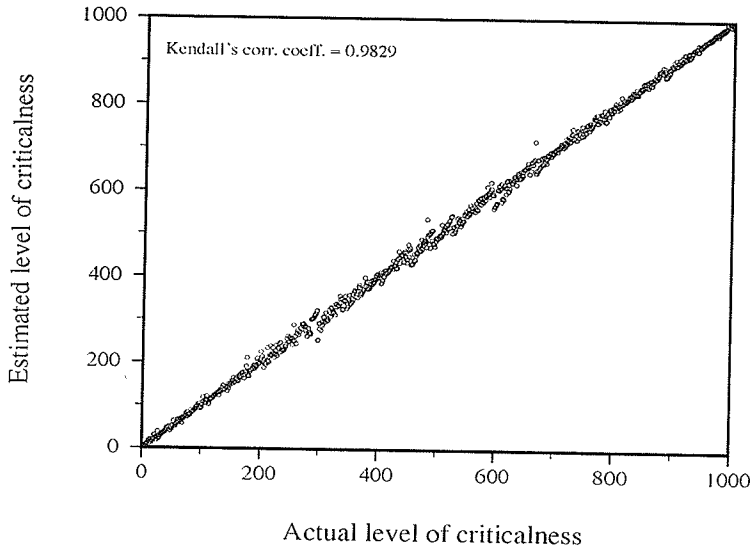
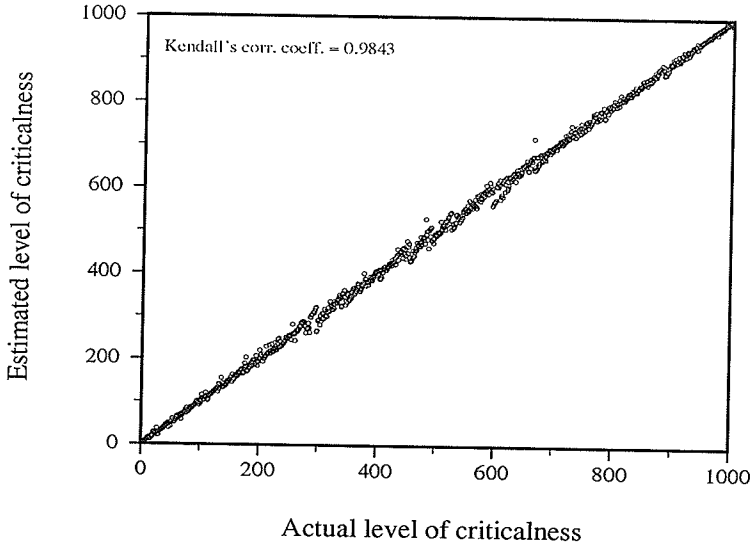


Figure 5.6: Performance of PANN1 network at different levels of training

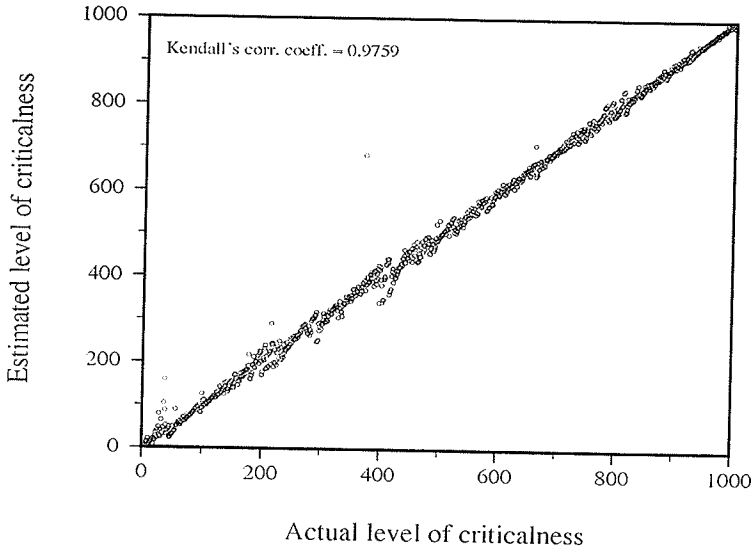


(a)

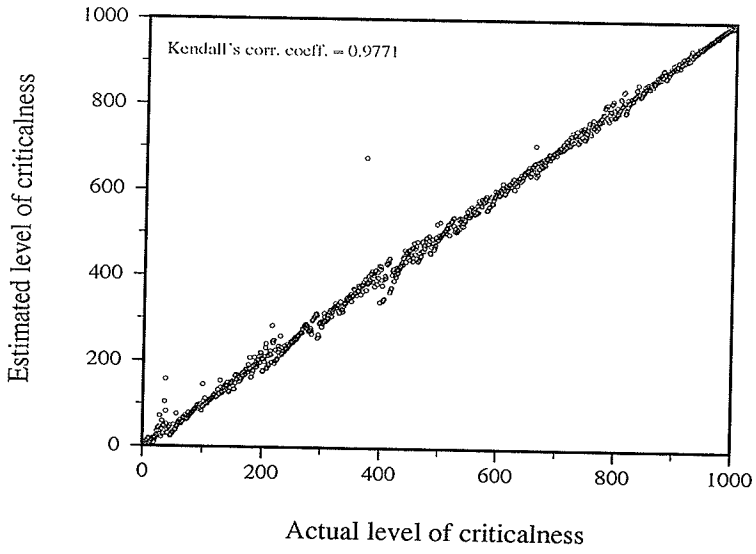


(b)

Figure 5.7: Training performance of PANN1 network at two different levels of training: (a) 2000 iterations; (b) convergence (26091 iterations)

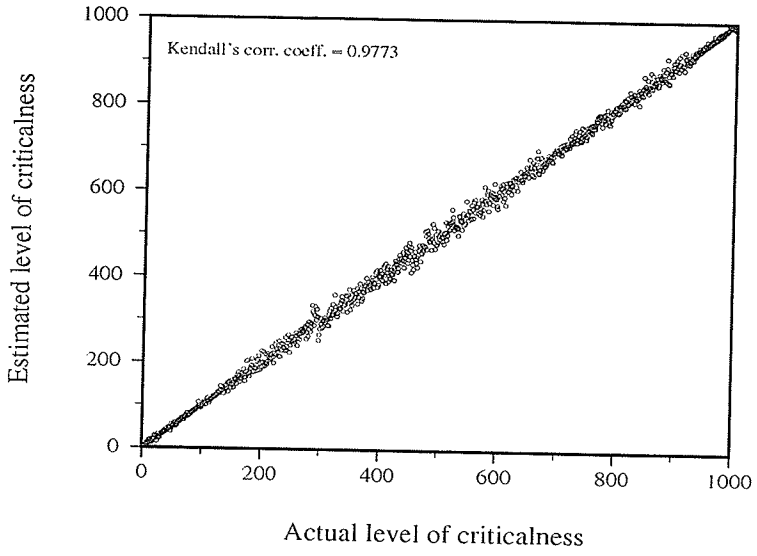


(a)

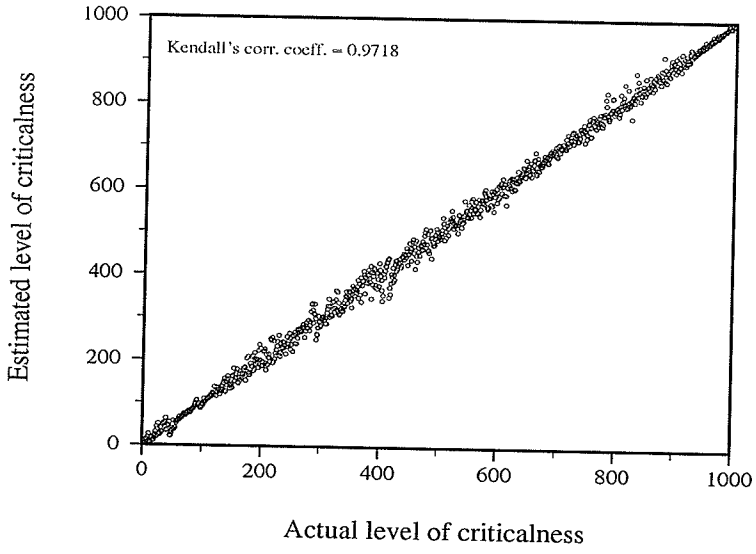


(b)

Figure 5.8: Generalization performance of PANN1 network at two different levels of training: (a) 2000 iterations; (b) convergence (26091 iterations)



(a)



(b)

Figure 5.9: Performance of PANN2 network after 1000 iterations: (a) Training set; (b) Testing set

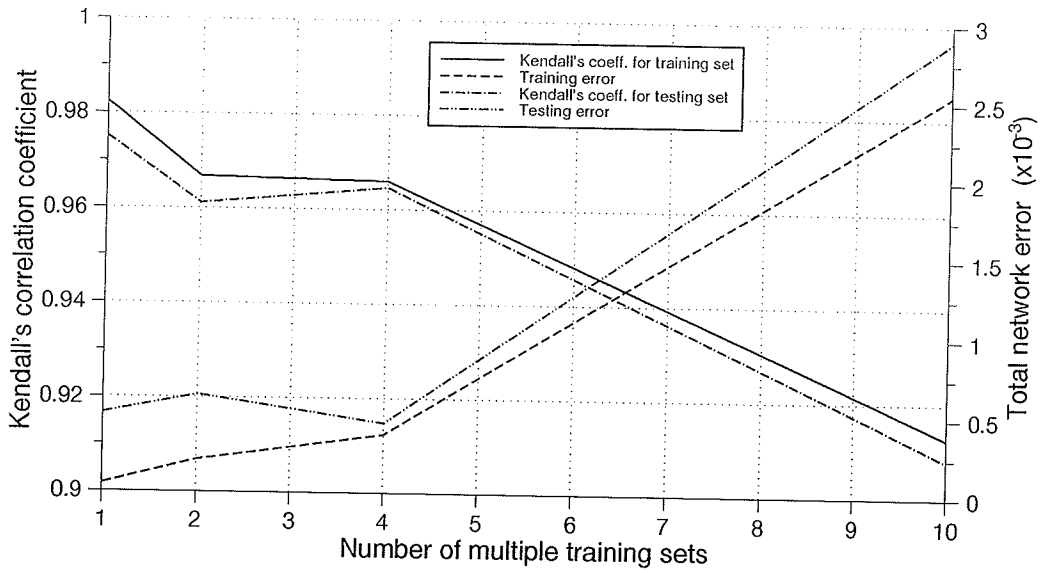
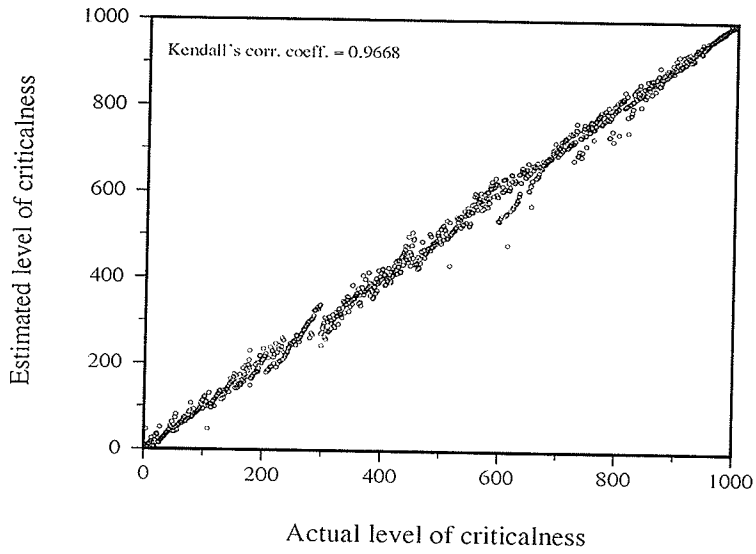
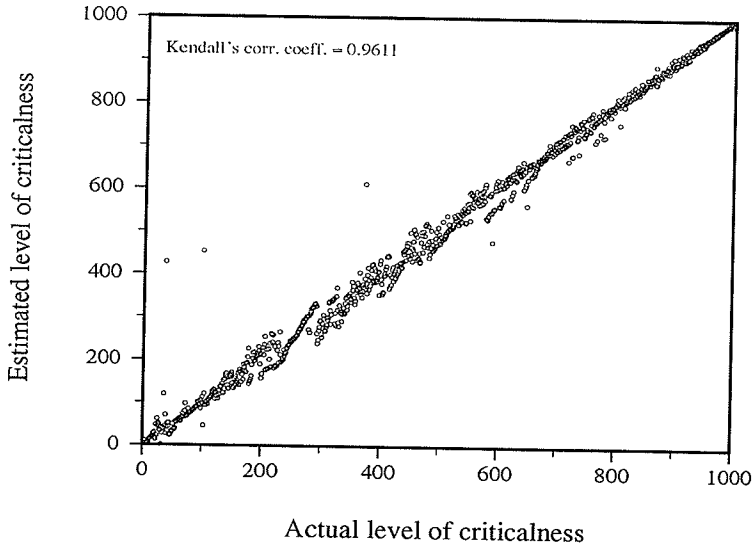


Figure 5.10: Performance of PANN1 network for different numbers of multiple training sets after a total of 1000 iterations

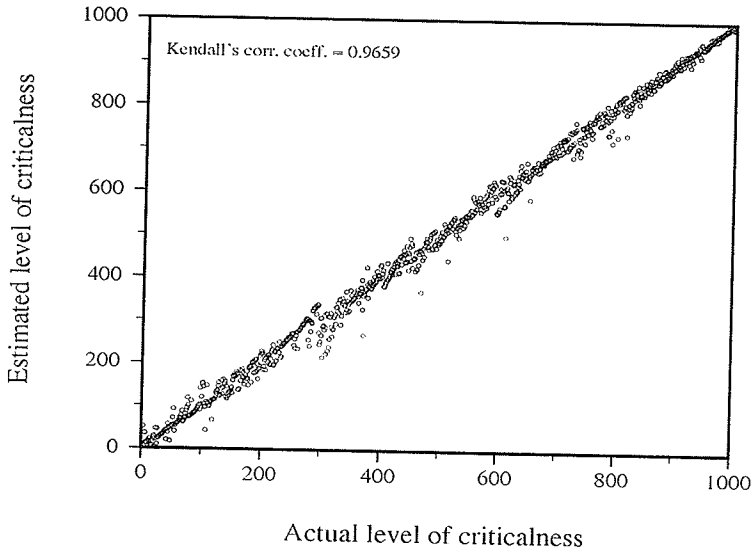


(a)

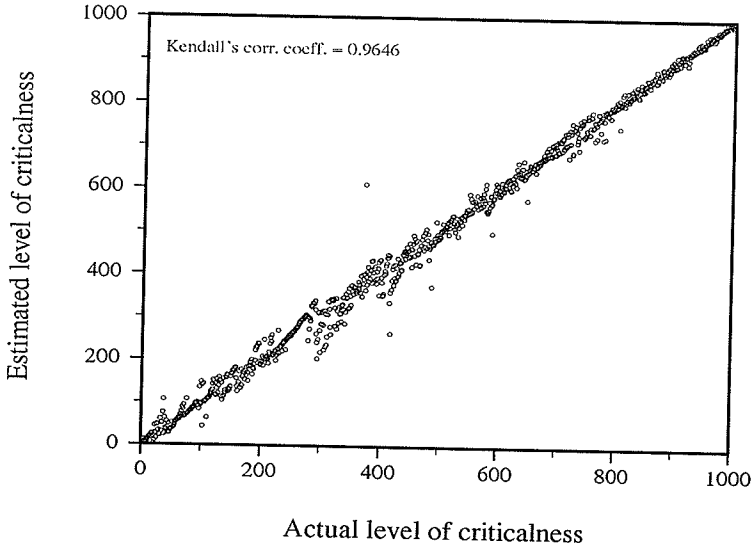


(b)

Figure 5.11: Performance of PANN1 network with two training example sets of 500 patterns each after 1000 iterations: (a) Training set; (b) Testing set

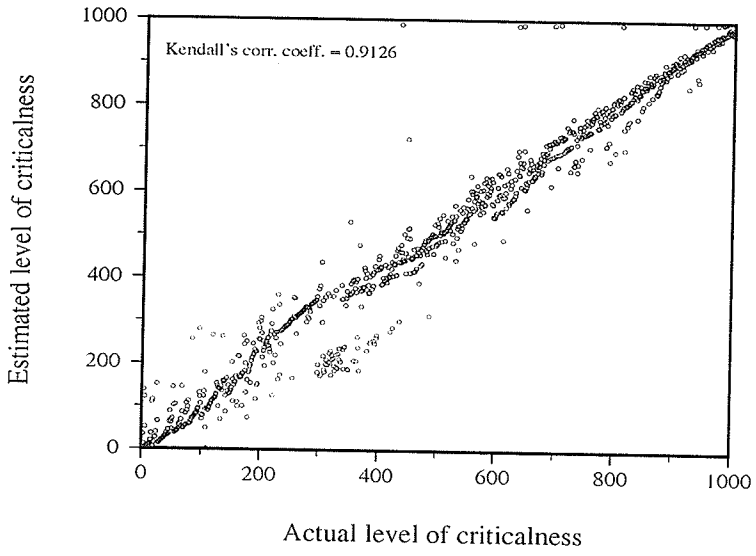


(a)

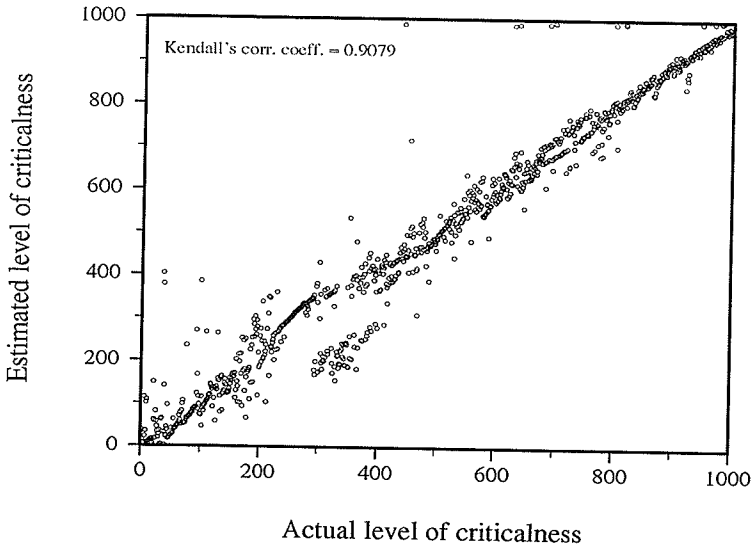


(b)

Figure 5.12: Performance of PANN1 network with four training example sets of 250 patterns each after 1000 iterations: (a) Training set; (b) Testing set



(a)



(b)

Figure 5.13: Performance of PANN1 network with ten training example sets of 100 patterns each after 1000 iterations: (a) Training set; (b) Testing set

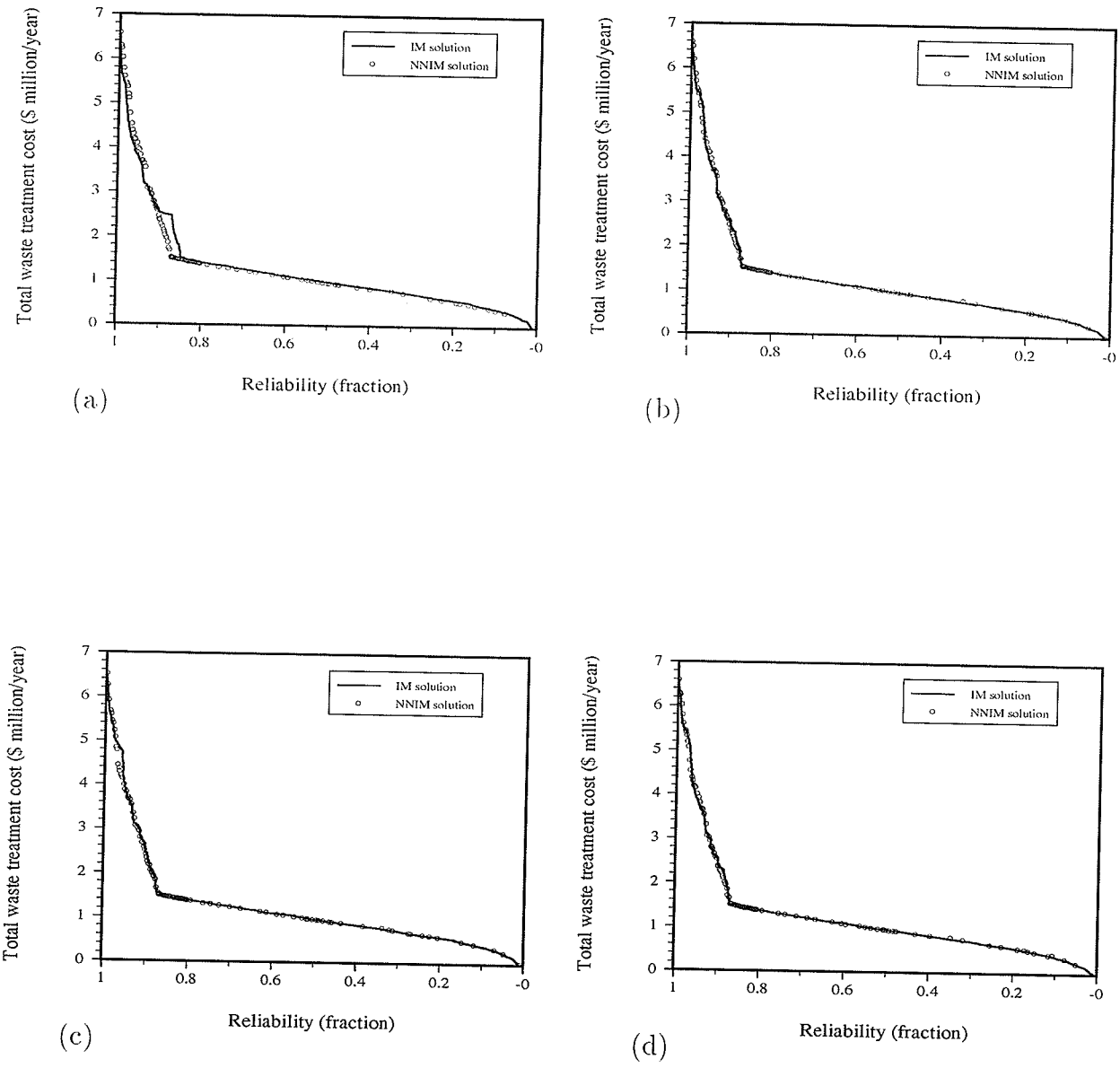


Figure 5.14: Comparison of IM and NNIM solutions based on 2000 realizations: (a) SIM_1; (b) SIM_2; (c) SIM_3; (d) SIM_4

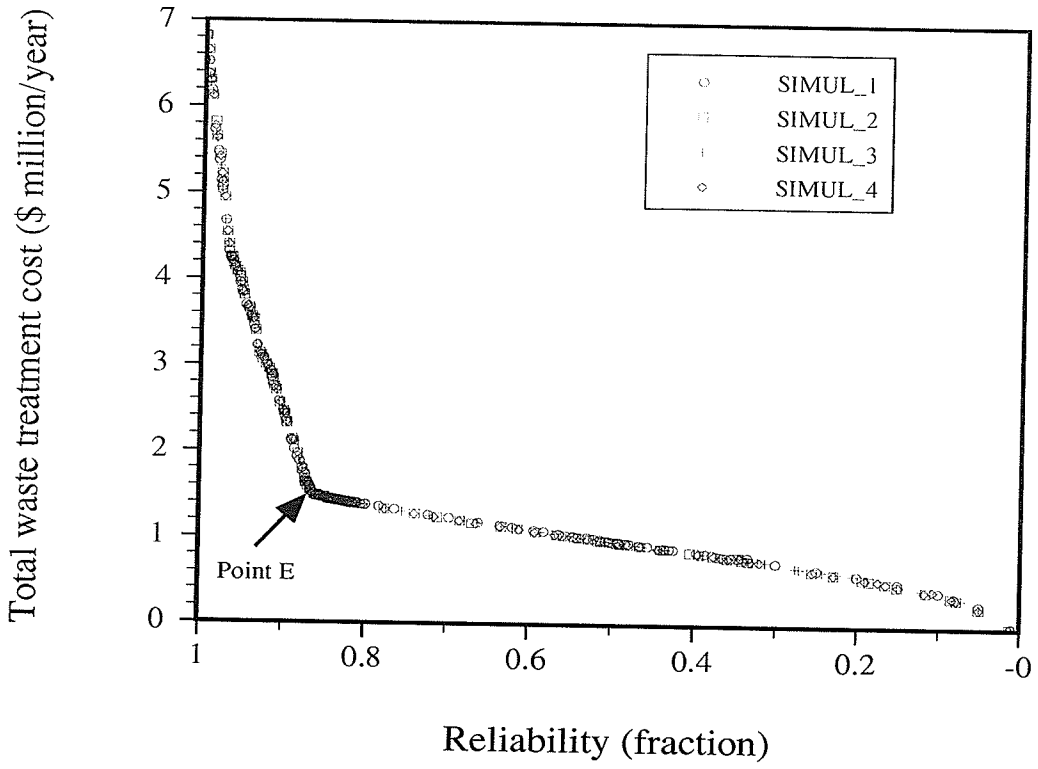


Figure 5.15: Trade-off plots for four sets of 4000 realizations each

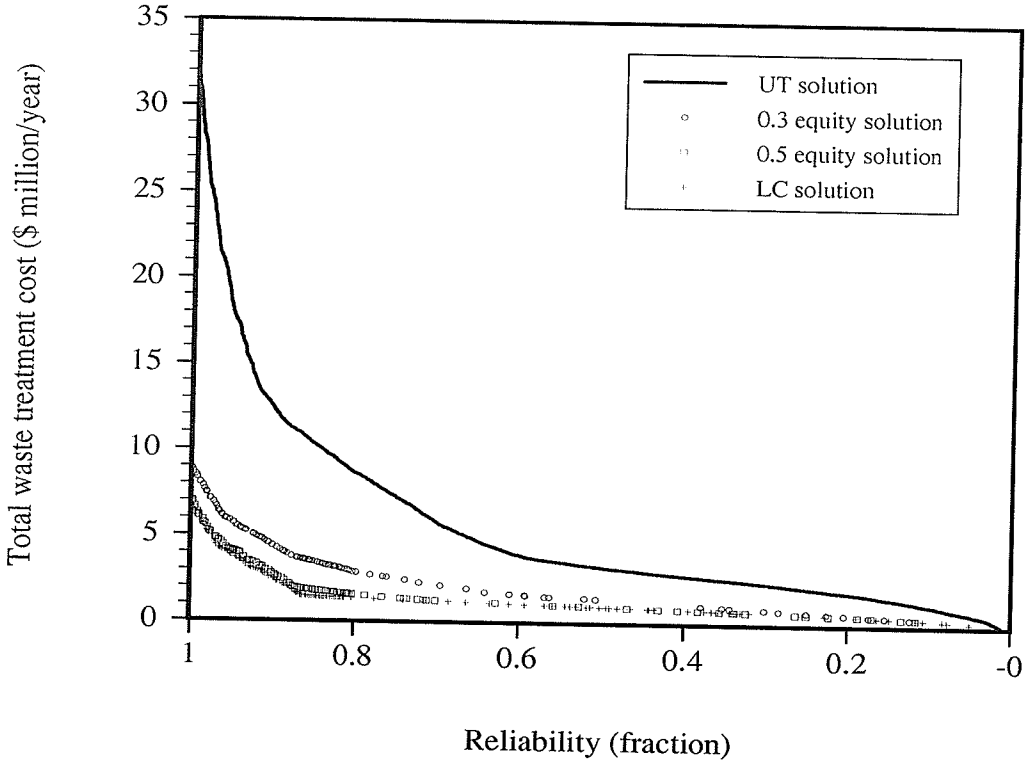


Figure 5.16: Trade-off plots for equity scenarios

Chapter 6

Seasonal Water Quality Management Using a Markov Chain Model

6.1 Introduction

This chapter presents an approach for incorporating the stochastic nature of stream background conditions and the spatial dependence of these conditions in the design of robust seasonal waste management programs using Markov chain modelling. Such robust designs are necessary for seasonal programs, because seasonal water quality information is often sparse or limited for many river systems. Furthermore, techniques for incorporating input information uncertainty in non-seasonal programs may not be easily used to address concerns of limited historical information which may be of particular importance to seasonal management models.

For the Markov chain approach, transition probabilities between possible states of the water quality system for a given season are estimated based on simple regression relationships between low flows at adjacent flow stations. The probability of water quality violation is determined based on the transition probabilities, on a selected water quality management model, and on Bayes' theorem of conditional probability. The Markov approach is demonstrated for a seasonal uniform program that maintains a pre-assigned risk of water quality violation. The sensitivity of the solutions of the Markov model to limited water quality background information is compared with that of the seasonal Minimum Average Uniform Treatment (MAUT) model developed by Lence et al. (1990). The MAUT model is a comprehensive seasonal model for controlling multiple waste discharges in terms of considering the stochastic nature of input water quality information. Furthermore, this model ensures equity among waste dischargers, achieves cost savings relative to equivalent non-seasonal programs, and is significantly less complicated to analyze than other non-seasonal programs (see, e.g., Lence et al., 1990 for a discussion of the complicated nature of the modelling process and the analyses of a seasonal least cost program). The MAUT program is reviewed in the next section.

6.2 The Seasonal Minimum Average Uniform Treatment (MAUT) program

The MAUT program divides the year into seasons and specifies the uniform fraction of waste removal efficiency required for each season so that the total annual cost of waste treatment is minimized and the number of years in which water quality violations occur

during the recorded period are limited to a pre-assigned value. The solution approach is a two stage process. The first stage solves a linear programming (LP) model for each season s, of each year y, in the historical record. Each of these LP models minimizes the uniform fraction waste removal required UT_{sy}, given strict water quality constraints that maintain the desired water quality goal (e.g., the dissolved oxygen, DO, standard at all locations in the stream). Impact or transfer coefficients for the water quality constraints are based on results from simulation models for steady-state stream conditions for each period.

During the second stage of the MAUT program a non-linear programming (NLP) model selects a design set of seasonal uniform fraction waste removal levels η_s. This model minimizes the annual cost of waste treatment subject to a risk equivalency constraint. The risk equivalency constraint limits the number of years, on average, in which at least one water quality violation would occur. One or more water quality violations occurs in a given year if the design uniform fraction removal level in either or both seasons is less than the historical value UT_{sy}, obtained from the first stage. The NLP model also has constraint sets that limit the design uniform fraction waste removal in each season η_s, to remain between the technological maximum and minimum levels of waste removal.

Although a novel approach to designing seasonal waste discharge programs, the MAUT program has several drawbacks. The probability distribution of waste removal levels for each season (i.e., the distribution of UT_{sy} for each s) is based on the limited amount of discrete data available and as such is not smooth. This short-coming is particularly important at distribution extremes. In addition, the risk equivalent constraint in the NLP model allows only a few discrete points in the feasible solution space to

be considered. The discrete nature of the MAUT program, given the limited historical input data normally available and the definition of risk on which it is based, could result in the over- or under-design of waste removal levels. More importantly, the MAUT program assumes that no combination of stream conditions that is more critical for water quality management than the observed data exists. Thus, the MAUT program disregards other possible, and perhaps worse stream background conditions. Finally, to produce results that are not sensitive to uncertainties in stream conditions, the MAUT program relies on a large amount of measured data, which may be unavailable for many river systems.

6.3 Non-stationary Markov Chain (MARK) Model

The MARK model developed in this research differs from the MAUT program in the method used to calculate the probability of water quality violation. First, the transition probability matrices for low flows at adjacent gauging stations are developed, and are used to determine the probability of all possible critical states of the water quality system for each season of the year (i.e., the combination of low flows for gauging stations throughout the stream in that season). For each possible state of the system in a given season, a uniform treatment management solution is obtained. Finally, an NLP model is applied, which selects the design seasonal uniform fraction removal levels η_s, that minimize the total cost of waste treatment subject to a risk constraint. The risk constraint limits the joint probability of violation of the water quality standard for all seasons, given that the probability of violation in each season is the sum of the probabilities of all states that require removal levels higher than η_s.

In the MARK model, the range of potential values for each stochastic input datum (e.g., the 7-day averaged low flows at each gauging station) is divided into discrete intervals indexed by different integers. Discrete representative values are used to approximate the continuous stochastic datum in each interval. Each unique combination of intervals in the water quality system's data defines a state of that system. If there are I number of elements in an input data set and N discrete intervals for the values of each input datum, the maximum number of possible system states will be N^I. Such a system is said to be in state $t = (a_1, a_2, \ldots, a_i, \ldots, a_I)$ if datum 1 is in interval a_1, datum 2 is in interval a_2, ..., datum i is in interval a_i, ..., and datum I is in interval a_I, where $a_1, a_2, \ldots, a_i, \ldots, a_I$ are integers between 1 and N inclusive.

The number of discrete intervals N, and the ranges of these intervals may be chosen in a heuristic fashion. Smaller intervals may be selected for the range of data values expected to produce the design output. Yapo et al. (1993) caution that the discrete intervals should be chosen so that a reasonable number of historical data points (at least $J^{\frac{1}{3}}$, where J is the total number of data points) is observed within each interval. To eliminate subjectivity, cluster analysis may be used to determine discrete intervals for each datum. The K-means clustering algorithm is used by Yapo et al. (1993) to select discrete streamflow ranges for a Markov chain model for flood forecasting. The size of the intervals determines accuracy in the results of the Markov model, especially if the historical record is short. For short historical records and small intervals, the maximum likelihood method (ML), which is often used to determine transition probabilities, may result in poor estimates (see, e. g., Bogardi et al., 1993). On the other hand, if the intervals are large, the discrete values used to approximate the continuous stochastic data may be inappropriate, and may produce inaccurate output for the Markov chain

model.

Matrices of conditional or transition probabilities are estimated for pairs of station flow data that are physically dependent. The 7-day averaged low flow data at two adjacent gauging stations may be considered to be physically related since flow at the upstream station influences flow at the downstream station. The probability of a given system state t, given that the 7-day averaged low flow level at station i is known, is equal to the conditional probability of the occurrence of low flow at station $i+1$, given the low flow at station i , where low flows at station i have the strongest physical effect on those at station $i+1$. The probability of occurrence of a given low flow interval at station $i+1$ is assumed to be entirely dependent on the low flow interval at station i, but independent of low flow levels at all other stations. Thus, a one-stage Markov process is used to model the 7-day averaged low flow data. Although a simple probabilistic analysis can be used to model the low flow data, a Markov chain is utilized because it lends itself to good physical representation and interpretation of the stochastic relationships between flows at adjacent stations on the same river. If the low flows at a given station do not depend on the low flows at only one station, but on low flows at q number of stations, a q-stage Markovian model may be used. For example, a two-stage Markov process may be used to model a dendritic stream in which there are two gauging stations that are directly upstream of a third station, on separate branches of the stream and significantly influence the flow at the third station.

In general, the larger the number of stages q, for the Markov process, the better the probabilistic relationships among the low flow data at flow stations are represented. However, in such cases, the size of the transition probability matrices increases, increasing the dimension of the mathematical problem. Further, accurate estimation of these

larger transition probability matrices require a longer historical record.

If the stochastic low flow data are spatially distributed throughout the stream, a non-stationary Markov model is required in which the transition matrices are different for low flow data at different pairs of stations. When low flows at a given station have no physical dependence on those at any other station, a suitable probability distribution can be used to determine the unconditional or initial probability of occurrence of any low flow interval for that station.

The solution approach of the MARK model is a two step process. The first step solves uniform treatment optimization models for each possible state of the system for each season of the year and the second step selects design seasonal waste removal levels that minimize cost and maintain an acceptable risk or probability of water quality violation. For each season s, and system state t, the following uniform treatment (UT) model is solved:

$$\min UT_{st} \tag{6.1}$$

subject to

limits on water quality impacts:

$$\sum_{k=1}^{K} R_k(1 - UT_{st})a_{khst} \leq B_{hs} \qquad h = 1, \ldots, H \tag{6.2}$$

limits on the allowable treatment level in any state of the system:

$$UT_{st} \leq \mu U_{ks} \qquad k = 1, \ldots, K \tag{6.3}$$

$$UT_{ds} \geq \mu L_{ks} \qquad k = 1, \ldots, K \tag{6.4}$$

where

UT_{st} = uniform fraction waste removal level required, based on water quality design conditions when the system is in state t in season s, where state t represents a given combination of low flows throughout the stream in season s;

H = total number of water quality checkpoints;

K = total number of waste dischargers;

R_k = influent waste load for discharger k (e. g., mg/l BOD);

a_{khst} = impact on water quality at point h in the stream, caused by a unit of waste discharge released from discharger k in season s and system state t (e. g., mg/l DO depleted per kg/day BOD released);

B_{hs} = allowable water quality impact at checkpoint h in season s (e. g., mg/l DO);

μU_{ks} = upper limit on fraction waste removal for discharger k in season s; and

μL_{ks} = lower limit on fraction waste removal for discharger k in season s.

The first constraint set ensures that the water quality impact at each checkpoint is less than the allowable water quality impact based on the water quality goal. The second and third constraint sets represent the technological maximum and minimum limits of waste treatment, respectively. Although any linear programming algorithm can be used to solve *Equations* 6.1 - 6.4, an accelerated search technique is found to be efficient for this problem.

If there is a small number of seasons (e.g., two) and the months possessing simi-

lar seasonal background water quality characteristics are correctly grouped, the spatial dependence of the extreme water quality conditions is more important than seasonal persistence in predicting such conditions. Consequently, the extreme water quality conditions (i.e., the low flows) may be assumed to be independent from one season to the next in such cases, and the uniform treatment management solutions based on the seasonal extreme conditions may also be considered independent. Therefore, Bayes' theorem of conditional probability and the probability law of independent events can be applied to evaluate the probability distribution of uniform fraction removal levels for each season, based on the unconditional probabilities of intervals for the independent data, the transition probability matrices, and the uniform treatment optimization model. The probability that a random uniform fractional removal level for season s, ξ_s, is greater than some value E_s is given by:

$$P[\xi_s > E_s] = \sum_{l \in F_t} P[G_l^s] \tag{6.5}$$

where

$P[G_l^s]$ = probability that the system is in state l in season s;

F_t = set of all system states that result in ξ_s greater than E_s; and

$l \in F_t$ = system state that produces a fraction waste removal level ξ_s that is greater than E_s.

For a system state $l = (b_1, b_2, \ldots, b_i, \ldots, b_I)$ in season s, the probability of occurrence $\mathrm{P}[G_l^s]$, is given as:

$$P[G_l^s] = P[b_1] \prod_{i=1}^{I-1} P[b_{i+1}|b_i] \tag{6.6}$$

where

$P[b_1]$ = probability that the 7-day averaged low flow at station 1 is in interval b_1; and

$P[b_{i+1}|b_i]$ = transition or conditional probability that the 7-day averaged low flow at station $i+1$ is in interval b_{i+1} given that the low flow at station i is in interval b_i.

The probability of the occurrence of a flow interval at a station is conditioned on the low flow level at the adjacent upstream station because, for a natural stream, the flow at a given location is physically dependent on that at a close upstream station. Further, flows at an upstream station are not physical dependent on flows at a downstream station. A method developed in this research to estimate the transition probabilities is presented in the following section.

The second step of the MARK model solves an NLP model that selects the most cost efficient waste removal levels for each season s, η_s, which maintain an acceptable probability of water quality violation. The NLP uses the probability distribution for the uniform fraction removal level UT_{st}, obtained from step one, and limits the joint probability of non-violation for all seasons. Let P_s denote $P[\xi_s > \eta_s]$, where η_s is the design fractional removal level for season s. The formulation for the second step of the MARK model is:

$$\min \sum_{s=1}^{S} day_s C_s \eta_s \tag{6.7}$$

subject to

limit on the probability of water quality violation:

$$\prod_{s}(1-P_s)+R-1\geq 0 \tag{6.8}$$

limits on the allowable treatment level for a given season:

$$\eta_s \leq \eta U_s \qquad \forall \quad s=1,\ldots,S \tag{6.9}$$

$$\eta_s \geq \eta L_s \qquad \forall \quad s=1,\ldots,S \tag{6.10}$$

where

day_s = number of days in season s;

C_s = total daily cost per unit of fraction removal for all dischargers in season s (e. g., dollars);

R = maximum acceptable probability of water quality violation in a given year;

ηU_s = upper limit on design fraction waste removal level in season s;

ηL_s = lower limit on design fraction waste removal level in season s; and

all other variables are defined above.

The objective function minimizes the total annual cost. The total cost consists of the annual amortized fixed cost of the various waste treatment unit processes and the variable cost of operating these processes throughout the year. In general, the optimal cost of waste treatment in a given season may depend on the level of waste treatment performed in other seasons of the year, consequently, the optimization problem can be complicated, especially if different waste treatment processes are used by the dischargers (Lence et al., 1990). However, for the MAUT and MARK models, it is assumed

that all waste dischargers treat their waste at the same efficiency and use the same waste treatment processes. Therefore, possible complications involved in modelling the seasonal dependence of waste treatment cost are eliminated.

Equation 6.8 ensures that the probability of water quality violation in any given year does not exceed an assigned maximum value. The second and third constraint sets, *Equations* 6.9 and 6.10, represent the technological maximum and minimum limits of waste treatment, respectively. Constraint *Equation* 6.8 is non-convex. Therefore, a direct search procedure is used in this thesis to solve the NLP model. The search is limited to η_s values that correspond with P_s values less than R. In practice, the feasible solution space is small since typical values of R are small (usually less than 0.20).

6.3.1 Estimation of transition probabilities

This section discusses some existing methods for estimating transition probabilities, and presents a new technique developed to obtain these probabilities when historical records are limited. A common approach to estimating the transition probability matrices for cases in which historical records exist is by the maximum likelihood (ML) method. This method determines the probability of transition from interval m of datum i to a interval n of datum $i+1$ in a one-stage Markov chain process as $\wedge_{mn}/\sum_n \wedge_{mn}$, where $\wedge_{mn}$ is the number of historical transitions from interval m of datum i to interval n of datum $i+1$. Yapo et al. (1993) presented proof of the ML estimates of transition probabilities. A major disadvantage of the ML approach is the need for a large number of concurrent historical data to obtain good estimates of transition probability matrices. To overcome the problem of limited data, long synthetic records may be generated based

on the historical records. The ML approach, based on the synthetic records may then be used to estimate the transition probabilities. However, such an approach may require fitting a multivariate distribution, which also needs a long historical record to give a good estimate of the distribution parameters.

For seasonal water quality management programs, the stochastic input data that are typically most important (especially for BOD-DO control programs) are the low flow and high temperature values since seasonal differences in low flow and temperature data are the main driving force for the design of seasonal programs and since most other information, such as the stream velocity and reaction rates, depends on flow and temperature data. The transition probabilities for low flow data can be quantitatively estimated using historical data. This research developed a method for estimating transition probabilities for low flow data at two adjacent gauging stations along a stream. This method can also be applied to any pair of data that are related, such as high temperatures at adjacent stations, BOD runoff and streamflow, and stream velocity and streamflow. It may also be extended to estimate transition probabilities for multi-stage Markovian processes if the data fit a linear regression model.

The transition probabilities for each season are estimated based on the range of low flow data (e.g., the 7-day average low flows for that season) and on simple regression relationships between low flows, transformed or untransformed, at locations along the stream. It is assumed that the low flows at a downstream station depend solely on the observed low flows at the adjacent upstream station. If the gauging stations are close enough and there are no large tributaries between them, linear regression models for adjacent stations are adequate for relating the low flows, transformed or untransformed. Although the residuals of a linear regression between low flows at adjacent gauging

stations are expected to increase with the magnitude of the upstream low flows, especially if the range of upstream flows is large, appropriate transformations can often be obtained to provide constant variance residuals (Neter et al. 1989).

The method for estimating transition probabilities developed here is based on the results of simple linear regression for the case in which the independent datum is a random variable (Neter et al. 1989). If Y and X are the dependent and independent variables, respectively (e.g., upstream and downstream low flows at two adjacent stations in a given season), and if these data fit a linear regression model, then the conditional distribution of Y_r given X_r is given as $Y_r \sim N(\overline{Y}_r, \sigma_r^2)$, where $\mathrm{N}(\overline{Y}_r, \sigma_r^2) =$ normal distribution with a mean of $\overline{Y}_r$ and a variance of σ_r^2; $Y_r =$ the r th random value of the dependent variable, e.g., the r th random value of the low flow at the downstream station in a given season; and $X_r = r$ th random value of the independent variable, e.g., the r th random value of the low flow at the upstream station in a given season. The mean $\overline{Y}_r$ and the variance σ_r^2 are given as:

$$\overline{Y}_r = \beta_0 + \beta_1 X_r \tag{6.11}$$

$$\sigma_r^2 = MSE[1 + \frac{1}{J} + \frac{(X_r - \overline{X^*})^2}{\sum_{y=1}^{J}(X_y^* - \overline{X^*})^2}] \tag{6.12}$$

where

$\overline{Y}_r$ = the mean of the r th estimated value of the dependent variable, e.g., the mean of the r th estimated low flow at the downstream station in a given season;

σ_r^2 = variance of the r th estimated value of the dependent variable, e.g., variance of the r th estimated low flow at the downstream station

in a given season;

β_0, β_1 = regression parameters;

MSE = mean square error of the regression equation;

J = total number of pairs of known data, e.g., the number of years in which low flow data are available at both upstream and downstream stations in a given season;

$\overline{X^*}$ = mean value of the independent known datum, e.g., the mean value of the observed low flows at the upstream station in a given season; and

X_y^* = y th value of the independent known datum, e.g., observed low flow value at the upstream station in year y in a given season.

For each gauging station, flow intervals are defined and the mean of the minimum and maximum flows for each interval is selected as the representative flow for that interval. The representative flow for a given interval of an independent station (i.e., the upstream station) is used to compute a mean response flow for the dependent station (i.e., the downstream station). The mean response of the regression model is the mean of the flow at the dependent station for a given flow interval at the independent station, and the variance of flow predictions about this mean is the variance of the normally distributed flow at the dependent station given the flow at the independent station. The mean and the variance can be estimated using *Equation* 6.11 and 6.12, respectively. The probability of being in flow interval n at the dependent station, given that the system is in flow interval m at the independent station P_{mn}, is computed as:

$$P_{mn} = P[Y_n|X_m] = \Phi[\frac{Y_n^U - \overline{Y}_m}{\sigma_m}] - \Phi[\frac{Y_n^L - \overline{Y}_m}{\sigma_m}] \qquad (6.13)$$

where

$P[Y_n|X_m]$ = probability of being in flow inteval n at the dependent station, given that the system is in flow interval m at the independent station;

Y_n^U = upper value of flow interval n at the dependent station;

Y_n^L = lower value of flow interval n at the dependent station;

$\overline{Y_m}$ = mean response at the dependent station when flow at the independent station is in interval m;

σ_m = standard deviation of the flow at the dependent station when the flow at the independent station is in interval m; and

$\Phi[Z]$ = probability that the normalized flow is less than or equal to Z.

Finally, it is assumed that the mean 7-day averaged low flow for a given interval at a downstream station is greater than or equal to that of an upstream station. If there is no major consumptive use of water from a stream flow interval combinations that have flow at an upstream station greater than that of a downstream station are impossible.

The conditional probabilities are normalized by dividing the values in each row by the sum of the values in that row. This normalization results in the following properties for the transition probability matrices:

$$0 \leq P_{mn} \leq 1 \qquad \forall \quad m = 1, \ldots, M, \quad n = 1, \ldots, N$$

$$\sum_{n=1}^{N} P_{mn} = 1 \qquad \forall \quad m = 1, \ldots, M$$

where

M = total number of flow intervals for the independent station; and

N = total number of flow intervals for the dependent station.

It is required that an equal number of intervals are defined for the flows at all stations (i.e., $M = N$).

6.4 Application to the Willamette River Basin

The MARK model is demonstrated and compared with the MAUT approach for a two season BOD waste management program for the Willamette River in Oregon described in Chapter 3. Seasonal low flow data (7-day averaged low flows) at five stations, Springfield (Station 1), Harrisburg (Station 2), Albany (Station 3), Salem (Station 4), and Portland (Station 5) are used for this analysis. The temperature data for Station 2, Harrisburg, is used to represent the stream temperature for the entire system. The highest monthly mean temperature at Station 2, for a given season, is used in this analysis.

Solutions of the MARK model are evaluated for three water quality goals: the 6.00, 7.00, and 7.50 mg/l DO standards. These DO standards represent the feasible range of DO for the stream background conditions. Six winter and summer season length combinations are analyzed for each of the three DO standards. Table 6.1 shows a summary of the six two-season length combinations used in this study. Each combination of season lengths and water quality goal represents one scenario of the analysis. These 18 scenarios (i.e., three DO standards and six season length combinations) were analyzed to be consistent with Lence et al. (1990) and to allow complete comparison of the results from the MARK and MAUT models. The probability of water quality violation is set at 10%, i.e., R = 0.10 in *Equation* 6.8.

Although mean streamflow persistence between seasons is significant for many large

river systems, winter and summer 7-day averaged low flows often do not show such persistence if all months with summer flow characteristics are separated from those possessing winter characteristics. Low flow persistence within a season may be ignored, since seasonal water quality management models typically use a single design low flow value (e.g., the 7-day averaged low flow) for a given station during a given season. The 7-day averaged low flow statistic for a significantly long period of time (e.g., greater than 6 months) does not generally reflect the hydrological condition for that period. Figure 6.1 shows graphs of 7-day averaged low flows of summer against those of winter and the corresponding R^2-statistic for linear regression models for the scenario with a 6 month summer season. Visual inspection of these graphs and the R^2-statistics, which indicate the percentage of the variation of predicted winter flows associated with summer flows suggests that the 7-day averaged low flows for winter and summer may be assumed as independent in this scenario; therefore, seasonal persistence of these low flows may be neglected. Although the significance of the R^2-statistics, depends on the number of data points and how well the possible range of values of the independent variables are represented by the sample data, the 30-year flow record utilized is assumed to be adequate for determining the R^2-statistics and for characterizing the relative importance of the linear correlation between temporal and spatial low flow data. For cases in which seasonal 7-day averaged low flow persistence may not be negligible, the importance of spatial dependence would be expected to dominate seasonal persistence significantly.

Consequently, the probability of the low flow magnitude at Station 5 is assumed to be entirely determined by the low flow at Station 4. Similarly, it is assumed that the low flows at Stations 3 and 2 determine the probability of low flow magnitudes at Stations 4 and 3, respectively. Each pair of low flows at adjacent stations, from Station

2 to 5, fit a linear regression model. Figures 6.2 and 6.3 show linear plots for the winter and summer low flows, respectively, at adjacent stations for the scenario with 6 month summer season. The R^2-statistics for the linear regression models are also indicated in these figures. The low flows at Stations 1 and 2, did not fit a simple linear regression model. However, the difference between the flows at Stations 1 and 2, and the flow at Station 2 fit a simple linear regression model. This implies that the low flow of the tributary between Stations 1 and 2 hydrologically influences the low flow at Station 2 more significantly than does the low flow at Station 1. Conditional probabilities were determined between the low flows for the inflow between Stations 1 and 2 and for Station 2. Takyi (1991) and Lence and Takyi (1992) fit a lognormal distribution to the 7-day average low flows at the five stations in this system. Therefore, a lognormal distribution was assumed for the low flows of the inflow between Stations 1 and 2, and used to estimate unconditional probabilities for these flows.

Seventeen flow intervals were arbitrarily assigned to the low flow data for the inflow between Stations 1 and 2 and for Stations 2 to 5. For these 17 flow intervals and five gauging stations, 1,419,857 (i.e., 17^5) uniform treatment optimization models, *Equations* 6.1 - 6.4, must be solved for each season. In practice, one does not solve this large number of optimization problems since the probability of occurrence of most of these 1,419,857 system states is zero. In the Willamette River example presented, if 17 flow intervals are used, less than 40,000 uniform treatment models are solved for each season.

6.5 Discussion of results

By assigning the probabilities of occurrence for a given set of input data, the MARK model produces a smooth representation of the probability distribution of the uniform treatment levels for each season. Unlike the MAUT program, the MARK model has the ability to explicitly incorporate the probabilistic relationships between low flow data at adjacent gauging stations and to consider possible water quality input data that may be more critical to water quality management than the historical data set. These critical water quality conditions are predicted from statistical relationships between the historical streamflow records. Therefore, the extremes of the probability distributions of the uniform treatment levels may be better represented in applications of the MARK model than in applications of the MAUT program. In contrast to the MAUT program, the risk constraint in the MARK model, (*Equation* 6.8), allows for the consideration of a large number of feasible solutions and results in a more accurate representation of the feasible solution space.

Figure 6.4 shows the probability of the DO standard violation for the range of waste removal levels for summer for the MARK and MAUT models, and the scenario with the 7.50 mg/l DO standard and the 4 month summer season. A similar distribution was obtained for the winter season for this scenario. The probability of the DO standard violation for a given uniform fraction removal level ξ_s, in season s, is the sum of the probabilities associated with all water quality system states that produce higher uniform fraction removal levels than ξ_s. This figure shows that the two models give similar overall probability distributions of waste removal levels. However, it demonstrates that the tails of the distributions may be different for the models. The MARK

model produces smoother distributions than the MAUT model, especially at the tails. A smooth probability distribution of the waste removal levels is expected to have a better convergence to the actual parent distributions (i. e., the distribution for a long historical record) for the given data than is expected from a lumpy distribution. Thus, the rate of convergence may be faster for the MARK model than for the MAUT model.

The difference between the two distributions at the higher end of the range of waste removal levels is important since the maximum acceptable probability of water quality violation used in practice typically corresponds to this portion of the distributions. Figure 6.4 shows that for a probability of DO standard violation less than 0.5, the MARK model results in higher waste removal levels than the MAUT model. This indicates that the decisions obtained from the MARK model will be more conservative and may have better performance characteristics with respect to uncertainties in low flow data than those from the MAUT model.

For the Willamette River basin example analyzed, the highest waste removal level required for a given season is always larger for the MARK model than for the MAUT model. These results may be attributed to the amount and range of water quality conditions included in the analysis of the MARK model. The first stage of the MAUT program applies an optimization model based on historical records only. In the absence of a long series of observed data, the MAUT approach may exclude possible worse scenarios of stream conditions, which would require higher waste removal levels.

The total annual cost predicted by both the MAUT and MARK models for the 18 scenarios analyzed are shown in Figure 6.5. Generally, the MAUT and the MARK models have similar costs at the 10% risk level for the Willamette River basin. Total cost

decreases as the summer season length increases from 1 month to 3, 4, and 6 months for the 6.00, 7.00, and 7.50 mg/l DO standards, respectively. The cost then increases as the summer season length increases. Figure 6.5 also shows that, although the two analyzed models give similar results, the MARK model typically designs a slightly higher cost program than the MAUT program. The relatively lower total annual cost obtained with the MAUT program corresponds with lower removal levels, higher total waste discharged into the river system, and potentially unacceptable probabilities of water quality violation.

Table 6.2 shows the seasonal uniform fraction waste removal levels for the 6 season length combinations for the 6.00, 7.00, and 7.50 mg/l DO standards. The results obtained for the MARK and MAUT models indicate that the MARK model is a comparable approach for designing seasonal risk equivalent water quality management strategies. The solutions for both models are close because the low flow data for the Willamette River appears to characterize reasonably the joint probability distributions of station flows. However, as shown in Figure 6.4, the small differences in removal levels (from the solutions of the two models) may result in significantly different probabilities of water quality violation.

With the exception of the 1 month summer scenario, the summer uniform fractional removal levels for the MARK model are always equal to or greater than those obtained for the MAUT program. The winter uniform fraction removal levels, however, show cases in which either model results in higher removal levels. Takyi (1991) and Lence and Takyi (1992) demonstrate that the waste treatment designs under the MAUT model for the Willamette River are more sensitive to uncertainties in summer flows than to those in winter flows. Therefore, one advantage of the MARK model, in this case study,

may be that the waste treatment designs under this approach are more conservative during the sensitive periods of the year.

Apart from the 1 month summer season scenario, there are no scenarios for the MAUT program that have both winter and summer uniform fraction removal levels that are simultaneously equal to, or greater than, the corresponding removal levels under the MARK model. However, in 8 of the 18 scenarios, the MARK model results in removal levels that are simultaneously equal to or greater than those under the MAUT program. This further indicates that, for the Willamette River basin, although differences in removals levels are small, solutions based on the MARK model may result in more conservative removal levels than those based on the MAUT program.

The scenario with the 7.50 mg/l DO standard and the 6 month summer season produces interesting results. Not only are the removal levels (required at the 10% probability of DO standard violation level) for both seasons greater for the MARK model than for the MAUT program, but the winter uniform fraction removal level for the MARK model is 0.05 greater than that for the MAUT. This is a case in which the MAUT program may result in a significantly different waste treatment strategy. A 0.05 difference in the winter uniform fraction removal level for this system would result in 4,200 kg/day of additional waste discharged.

The sensitivity of the MARK and MAUT management models to systems with short historical flow records was examined by investigating the management model solutions for the case in which only 10 years of continuous flow data exist. Twenty-one scenarios, which represent 10 consecutive years of continuous stream flow data each, were selected from the 30 years of historical low flow record. The MARK and the MAUT models

were applied to each of these scenarios for the case with the 7.50 mg/l DO standard and the 4 month summer season. The probability of DO standard violation, R, was set at 0.10.

The winter and summer waste removal levels for the 10-year low flow scenarios and for the case in which the entire 30 years of historical flow data were used (referred to as the Actual Flow Scenario) are shown in Table 6.3. For all scenarios, the MARK model solutions are more conservative than those of the MAUT model. The solutions for some of these scenarios show that the removal levels for both the MARK and MAUT models do not change significantly from the MARK and MAUT solutions for the Actual Flow Scenario. On the other hand, the solutions for Scenarios 14 - 21 indicate a significant difference between the two models. In 16 cases, the MARK model solution based on 10 years of flow data would underestimate the winter or summer waste removal levels, relative to those in the Actual Flow Scenario. However, the summer or winter waste removal levels, in 13 (Scenarios 2 - 5 and 11 - 19) of these 16 cases are higher, and therefore more conservative, than those required under the Actual Flow Scenario. On the other hand, in 13 cases, the MAUT model solutions under the 10 years of flow data would underestimate the winter waste removal level, relative to those in the Actual Flow Scenario without increasing the summer removal level above that of the Actual Flow Scenario. Thus, solutions to the MARK model, using 10 years of continuous low flow data, may be less susceptible to failure to maintain an acceptable DO standard than those of the MAUT model.

To represent the overall sensitivity of the model results to short flow records, the potential probability of the DO standard violation for the waste removal levels in the MARK and MAUT models under the 10-year low flow scenarios were evaluated. The

potential probability of the DO standard violation is evaluated by comparing the MARK and MAUT waste treatment solutions for the 10-year low flow scenarios with the cumulative probability distributions of the waste removal levels derived from the 30 years of historical low flow data. The potential probability of the DO standard violation is the joint probability of DO standard violation that would occur both in winter and summer, for the 10-year low flow uniform treatment solutions, if the correct probabilities of stream flow conditions were the same as those using the 30 years of known historical data. Thus, the potential probability of the DO standard violation gives an index of the likelihood of a given solution to maintain an acceptable water quality. The MARK model results were compared with the cumulative probability distributions of the actual streamflow data under the MARK model, and the MAUT model results were compared with the cumulative probability distributions for the actual streamflow data under the MAUT model.

The potential probability of the DO standard violation for the 10-year low flow scenarios are listed in Columns 4 and 7 of Table 6.3 for the MARK and MAUT models, respectively. The range of potential probability of the DO standard violation is from 0.04 to 0.32 and from 0.03 to 0.64 for the MARK and MAUT models, respectively. The low potential probability of the DO standard violation for the MAUT model in Scenarios 1- 8 is an indication of the limited amount of feasible solutions for the MAUT program. The discrete nature of the MAUT model is demonstrated in Scenarios 11 and 12, and 19 and 20. By changing only 1 year of flow data between these pairs of scenarios, the potential probability of the DO standard violation increases from 0.13 to 0.33 and from 0.37 to 0.64, respectively. The results for the MARK model, on the other hand, do not change as quickly as those for the MAUT model.

The potential probability of the DO standard violation for the 10-year low flow scenarios has a mean of 0.17 and a standard deviation of 0.08 for the MARK model, and a mean of 0.23 and a standard deviation of 0.20 for the MAUT model. The larger standard deviation for the MAUT model indicates a higher sensitivity of the potential probability of the DO standard violation for the MAUT model to short low flow records.

Table 6.3 indicates that the MARK model generally designs higher waste removal levels than the MAUT model. Although differences between the waste removal levels in the two management models for some scenarios are small, the potential probability that these removal levels will violate the acceptable water quality goal may be significantly different. For example, Scenario 9 shows that the winter and summer removal levels for the solution of the MAUT program are only 0.02 and 0.01, respectively, less than those for the MARK model. However, estimations of the probability of DO standard violations, when the MARK and MAUT solutions are compared with the actual distribution of removal levels from the Markov chain approach, indicate a 0.15 increase in the potential probability of the DO standard violation for the MAUT solution over that of the MARK solution.

The results of the sensitivity analysis indicate that the potential probability of the DO standard violation for the 10 year low flow scenarios for the MAUT model has a larger mean and variance than that for the MARK model. Also, for a given 10-year flow scenario, the solutions of the MARK model are more conservative than the results from the MAUT model. These observations suggest that solutions of the MAUT model may be more sensitive to short flow records and less likely to maintain an acceptable water quality level than those of the MARK model.

To analyze the robustness of the MARK and MAUT models for cases where the data are sparse, the models are evaluated for flow scenarios that represent 30 years of record in which 10 years of data are missing. Each scenario of the low flow data is developed by randomly selecting 20 years of low flow data from the actual 30 years of historical low flow data for the five flow stations. One hundred low flow scenarios are simulated and treatment levels for each scenario are evaluated. In all cases, the same stream temperatures, background DO and BOD conditions, and relationships between stream velocity, reaeration coefficients, reaction rates, streamflow and temperature are used.

To facilitate comparison and discussion of the results obtained, the following definitions are used in the rest of this section. Solution A is referred to as dominating Solution B if and only if the waste removal levels for A in both seasons are equal to or greater than the corresponding removal levels for B. Thus, if Solution A dominates Solution B, then the Solution A is either as conservative or more conservative than Solution B. In such cases, management decisions based on Solution A will have the same or higher probability of meeting the acceptable DO standard than those based on Solution B. Also, Solution C is inferior to Solution D if Solution D dominates Solution C and the two solutions are not equal. Generally, an inferior solution will be more vulnerable to failure to meet a specified water quality standard.

The results based on the 30 years of historical low flow record (i.e., the Actual Flow Scenario) and those based on the 100 simulations of 20 years of sparse record (with 10 years of missing data) indicate that the MARK model generally designs higher winter and summer waste removal levels than the MAUT model. Out of the 100 simulations, 83 solutions of waste removal levels produced by the MARK model dominate the corre-

sponding solutions from the MAUT model, while none of the MAUT solutions dominate the corresponding MARK solutions. Thus, results from the MARK model are expected to perform better than those from the MAUT model with respect to maintaining acceptable water quality. Despite the dominance of most of the MARK model solutions over the corresponding solutions of the MAUT model, the differences between these solutions are not very large. Therefore, the results of MARK model do not indicate large over-design of waste removal levels when compared with those of the MAUT model.

For this analysis, Table 6.4 shows the probability that a given set of winter and summer removal levels are dominated by the simulated solutions for the MARK and MAUT models, respectively. For example, the third row of Table 1 indicates that when winter and summer removal levels are 0.63 and 0.83, respectively, the dominance probability for the MARK model is 0.75 and that for the MAUT model is 0.54. This implies that, if the correct winter and summer removal levels are 0.63 and 0.83, respectively, then 75% of the MARK solutions will produce acceptable stream DO standards while only 54% of MAUT solutions will result in acceptable standards. The solutions of the Actual Flow Scenario of the MARK and MAUT models are marked with $\star$ and $\ddagger$, respectively.

It can be seen from Table 6.4 that the dominance probability for a given pair of winter and summer waste removal levels is always higher for the MARK than for the MAUT model. This table indicates that 48% and 42% of the simulated MARK and MAUT solutions, respectively, dominate the corresponding solutions for the Actual Flow Scenario. Although these two percentages for the MARK and the MAUT models are similar, by contrast all the remaining 58% of the simulation results for the MAUT model are inferior to the Actual MAUT solution (based on the Actual Flow Scenario) while only 36% of the simulation results for the MARK model are inferior to the Actual

MARK solution. Thus, for the MARK model, 16% of the simulations produce removal levels that do not dominate and are not dominated by the Actual MARK solution. For this 16% of the simulations, although the winter removal levels of the MARK model are lower than those of the Actual MARK solution, the summer removal levels are higher. The under-design in winter may be compensated for by higher removal levels in summer. Therefore, the ability of solutions from a MARK model to meet the acceptable DO standard is less sensitive to sparse low flow data than those from the MAUT model, for the example presented. Consequently, the MARK model may design more reliable water quality management decisions for sparse low flow data than the MAUT model.

In general, the solutions of the MARK model show less sensitivity to short and sparse low flow records than those of the MAUT model. This may be due to the method used to generate the relationship between probability of DO standard violation and waste removal level for each season, especially at the high end of the removal levels. Only a few critical historical flow conditions are utilized to obtain this relationship at high removal levels for the MAUT model. For the MARK model, the relationship is based on low flow conditions predicted from regression analysis for adjacent station flows and the probability distribution for low flows at the most upstream station. Thus, the probability distribution of waste removal level at the critical tail of the distribution (i.e., probability of the DO standard violation less than 0.1) for each season, depends on the statistical properties of the low flow data at all flow stations along the modelled section of the river. Therefore, errors or uncertainty in a few flow data points may affect the solutions of the MARK model less than they would for those of the MAUT model.

Table 6.1: Season Length Combinations

Summer Season Length (months)	Winter Months	Summer Months
1	September-July	August
2	September-June	July - August
3	October-June	July - September
4	October-May	June - September
6	November-April	May - October
10	December-January	February - November

Table 6.2: Uniform Fraction Removal Levels for the MARK and MAUT Models

D.O. Standard (mg/l)	Summer Season Length ($months$)	Uniform Fraction Removal Level			
		MARK		MAUT	
		Winter	Summer	Winter	Summer
6.00	1	0.45	0.44	0.48	0.45
	2	0.37	0.51	0.39	0.48
	3	0.35	0.49	0.35	0.48
	4	0.35	0.49	0.35	0.49
	6	0.35	0.50	0.35	0.50
	10	0.35	0.50	0.35	0.50
7.00	1	0.69	0.69	0.69	0.70
	2	0.62	0.72	0.61	0.72
	3	0.56	0.73	0.57	0.72
	4	0.47	0.74	0.48	0.72
	6	0.35	0.73	0.35	0.72
	10	0.35	0.73	0.35	0.72
7.50	1	0.81	0.82	0.81	0.82
	2	0.74	0.85	0.74	0.85
	3	0.70	0.86	0.70	0.84
	4	0.64	0.85	0.63	0.84
	6	0.40	0.85	0.35	0.84
	10	0.35	0.85	0.35	0.84

Table 6.3: Results for the MARK and MAUT Models Based on 10 Years of Continuous Low Flow Data

Scenario	MARK			MAUT		
	Fraction Removal Level		Probability of DO Standard	Fraction Removal Level		Probability o DO Standar
	Winter	Summer	Violation[1]	Winter	Summer	Violation[2]
Actual Flow	0.64	0.85	0.10	0.63	0.84	0.03
1	0.64	0.84	0.21	0.63	0.84	0.03
2	0.66	0.84	0.18	0.63	0.84	0.03
3	0.66	0.84	0.18	0.63	0.84	0.03
4	0.65	0.84	0.19	0.63	0.84	0.03
5	0.65	0.84	0.19	0.63	0.84	0.03
6	0.65	0.85	0.08	0.63	0.84	0.03
7	0.66	0.85	0.06	0.63	0.84	0.03
8	0.65	0.85	0.08	0.63	0.84	0.03
9	0.63	0.85	0.12	0.62	0.84	0.13
10	0.65	0.85	0.08	0.62	0.84	0.13
11	0.65	0.86	0.04	0.62	0.84	0.13
12	0.60	0.86	0.16	0.56	0.84	0.33
13	0.55	0.86	0.29	0.53	0.84	0.37
14	0.62	0.86	0.11	0.53	0.84	0.37
15	0.61	0.86	0.13	0.53	0.84	0.37
16	0.59	0.86	0.19	0.53	0.84	0.37
17	0.59	0.86	0.19	0.53	0.84	0.37
18	0.58	0.86	0.21	0.53	0.84	0.37
19	0.59	0.86	0.19	0.53	0.84	0.37
20	0.57	0.85	0.27	0.53	0.82	0.64
21	0.55	0.85	0.32	0.53	0.82	0.64

[1] Based on the 30-year historical flow record and the MARK model

[2] Based on the 30-year historical flow record and the MAUT model

Table 6.4: Dominance probability distribution for management solutions based on 20 years of sparse low flow records

	Waste removal level		Dominance probability level	
	Winter	Summer	MARK	MAUT
	0.63	0.83	0.75	0.54
	0.60	0.84	0.99	0.85
	0.61	0.84	0.94	0.84
	0.62	0.84	0.86	0.64
‡	0.63	0.84	0.75	0.42
	0.64	0.84	0.48	0.00
	0.60	0.85	0.99	0.00
	0.62	0.85	0.85	0.00
⋆	0.64	0.85	0.48	0.00
	0.66	0.85	0.07	0.00
	0.60	0.86	0.16	0.00
	0.61	0.86	0.11	0.00
	0.62	0.86	0.07	0.00
	0.63	0.86	0.04	0.00
	0.64	0.86	0.01	0.00

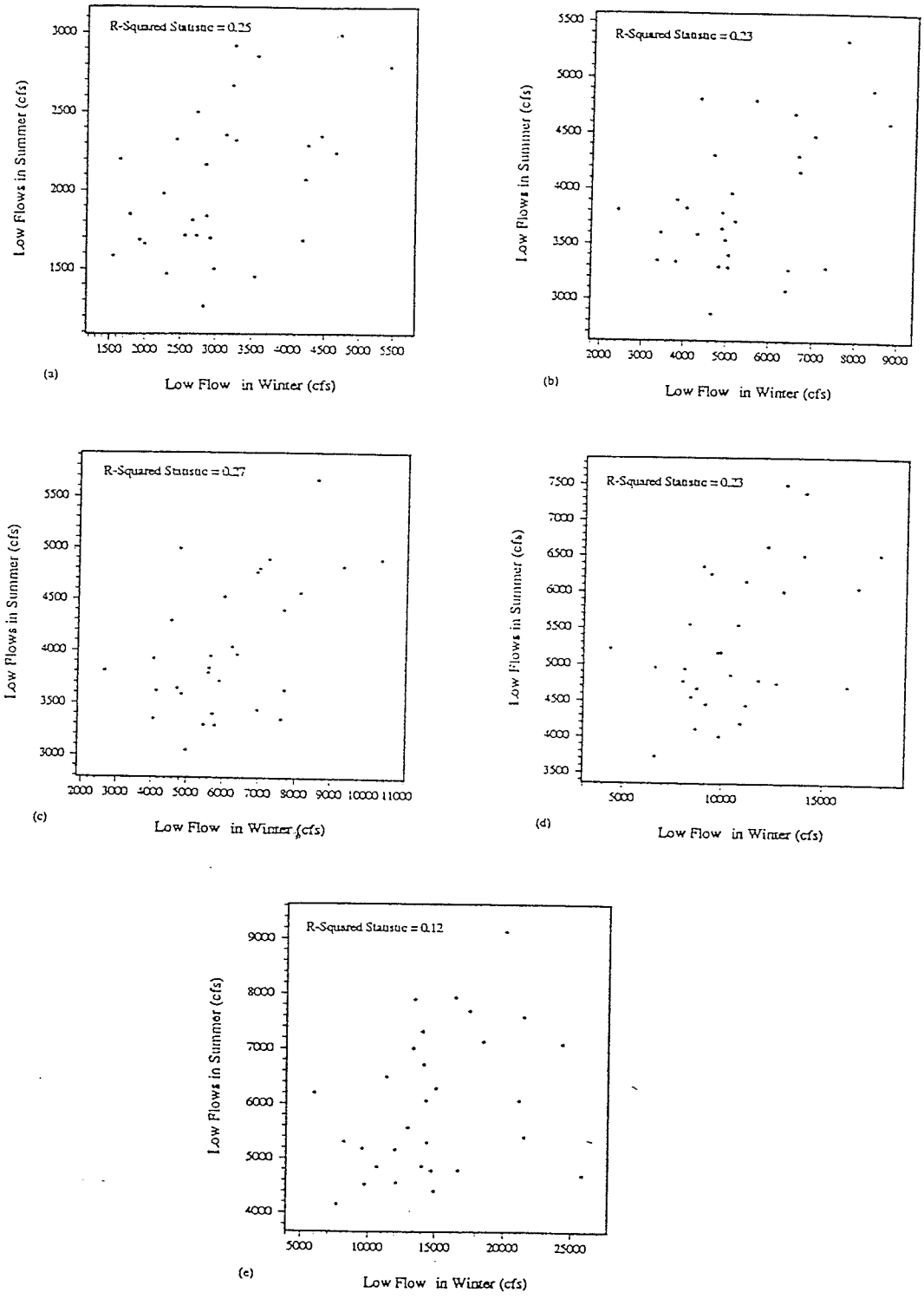


Figure 6.1: Linear Plots of Summer and Winter Flows for the Six Month Summer Season Scenario: (a) Difference between Low Flows at Stations 1 and 2; (b) Low Flows at Station 2; (c) Low Flows at Station 3; (d) Low Flows at Station 4; and (e) Low Flows at Station 5

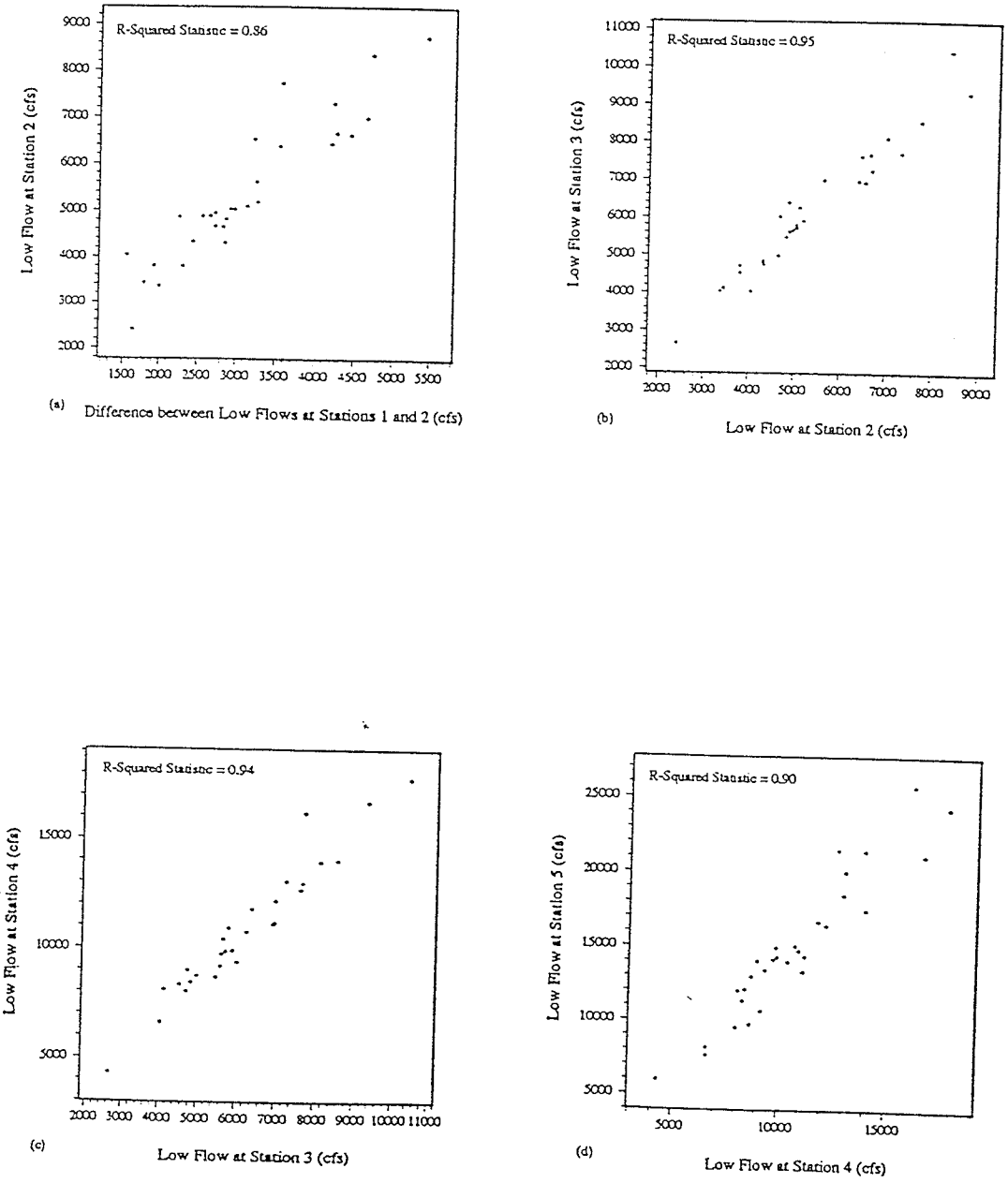


Figure 6.2: Linear Plots of Winter Low Flows at Adjacent Flow Stations for the Six Month Summer Season Scenario: (a) Tributary 1 and Station 2; (b) Stations 2 and 3; (c) Staions 3 and 4; (d) Stations 4 and 5

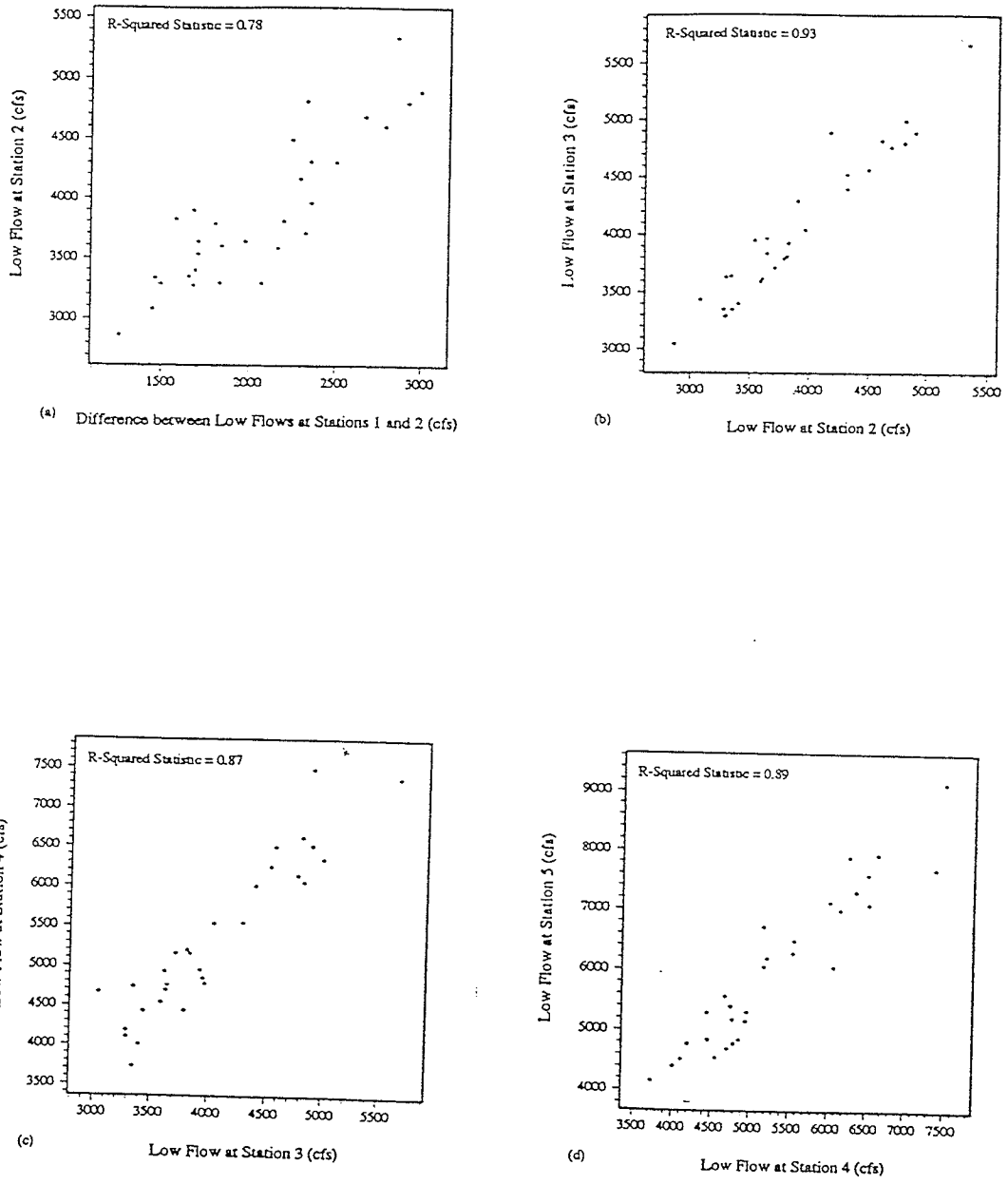


Figure 6.3: Linear Plots of Summer Low Flows at Adjacent Flow Stations for the Six Month Summer Season Scenario: (a) Tributary 1 and Station 2; (b) Stations 2 and 3; (c) Stations 3 and 4; (d) Stations 4 and 5

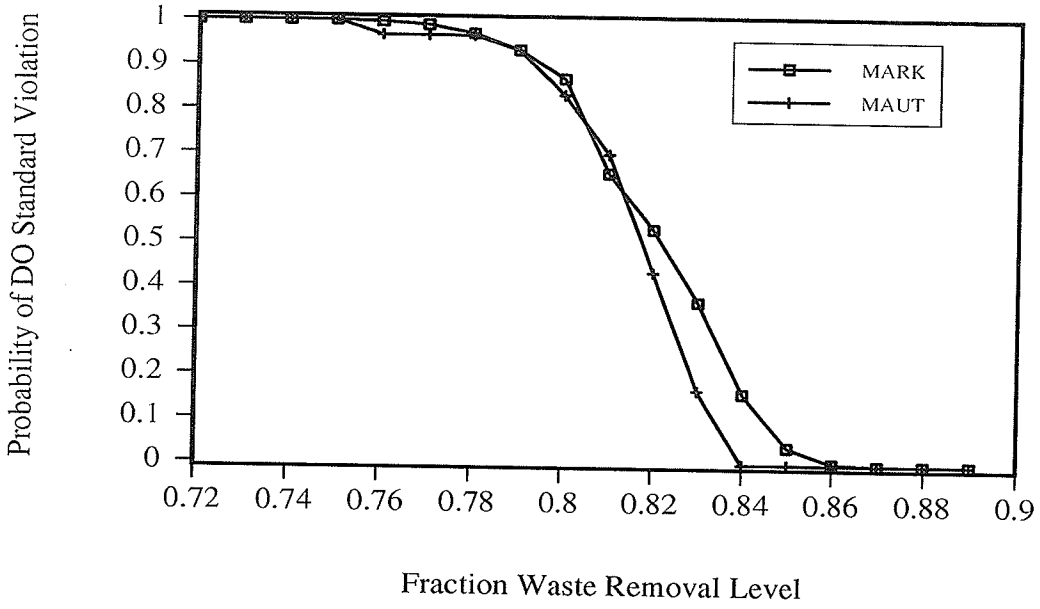


Figure 6.4: Probability of DO Standard Violation for Waste Removal Levels for the 7.50 mg/l DO Standard and the Six Month Summer Season

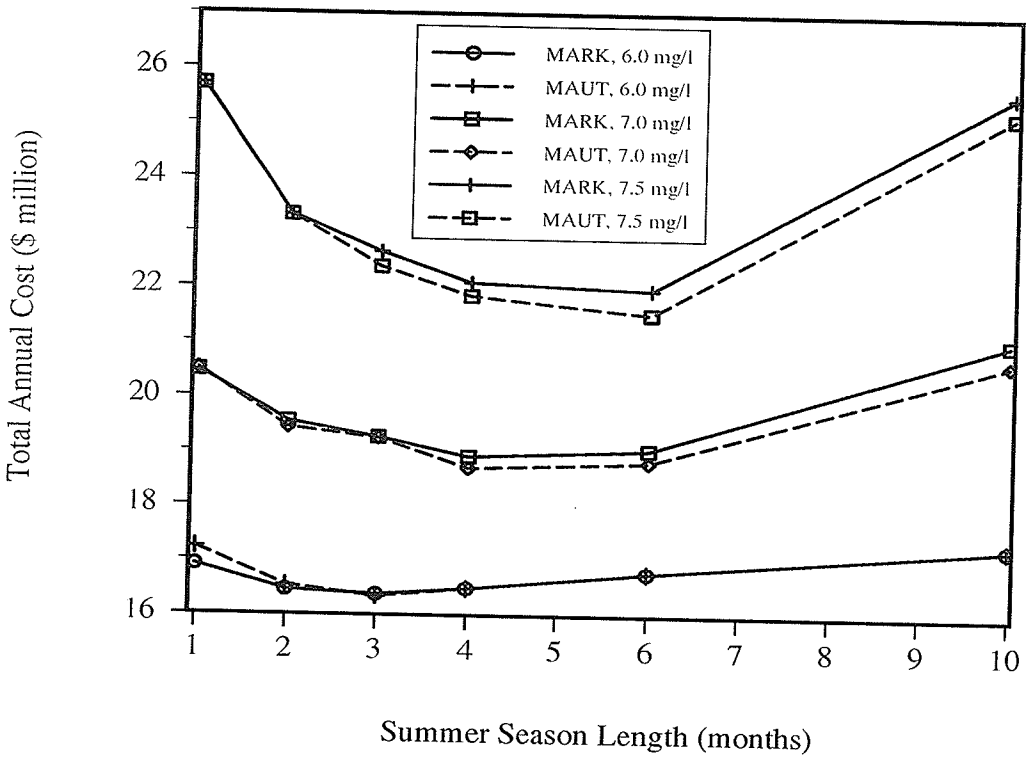


Figure 6.5: Total Annual Cost for MARK and MAUT Models

Chapter 7

Summary and Conclusions

This research incorporates Type II uncertainty in seasonal and non-seasonal wasteload allocation programs. Two different approaches are used for these programs due to different levels of complexity of the models used to represent these programs and the different levels of importance that may be assigned to the sources of Type II uncertainty under seasonal and non-seasonal conditions. The research issues that have been addressed are:

1. Obtaining the non-inferior solution set that defines the 'true' trade-off relationships between waste treatment cost and reliability using a 'new' SIOP model, the MR model;

2. Reducing the computational burden of the solution approach for the MR model to make the model more attractive to water quality management modelers; and

3. Developing an approach for designing robust seasonal wasteload allocation decisions that are somewhat insensitive to short and sparse historical flow data used

in the management model.

The above three contributions are demonstrated for the control of point sources of BOD discharges in a hypothetical river basin based on the Willamette River. The approaches developed may also be applicable to other environmental problems such as groundwater remediation design and air quality management.

Previous methods for incorporating Type II uncertainty in non-seasonal water quality management models often use CCP or SIOP. The CCP approach is often complicated and may require simplifying assumptions about the physical configuration of the water quality system, the sources and probability distribution of input information, and the objectives of the management model. By utilizes Monte Carlo simulation to generate several scenarios of stream conditions, the traditional SIOP approach has a better capability than the CCP method for representing Type II uncertainty in complex stochastic water quality systems. However, most of the solutions from this approach may be inefficient for several types of management programs that have more than one decision variable (e.g., the LC program for multi-discharger systems).

An improvement over the traditional SIOP model, the MR model, includes a large number of realizations of possible stream conditions in a single optimization model. Existing algorithms may solve this optimization model several times at several stages, dropping the most critical realization at each stage, to obtain efficient solutions that constitute the the non-inferior set for cost, or a cost surrogate, and reliability. The realizations are generated using Monte Carlo simulation or some appropriate technique to reflect the stochastic nature of the water quality management system.

Because large sized optimization models result from the MR model, and each of these models is solved several times to generate the 'true' cost-reliability trade-off relationship, the computational resources required may be very large. This large computational burden may either deter modelers from incorporating an adequate number of realizations in the MR model or even discourage them from using the model. To reduce the computational burden, several algorithms have been developed in Chapters 4 and 5 to select and include a small number of important realizations that potentially control the efficient solution at each stage of the solution process. All of these algorithms have been demonstrated to produce efficient solutions while drastically reducing the CPU time required to solve the MR model in conjunction with Morgan et al.'s heuristic algorithm. Information for selecting the important realizations are based on (1) the traditional SIOP solutions for the realizations included in the MR model, (2) the solution of the MR model for a case in which a relatively smaller number of realizations than that considered adequate are included in the MR model, or (3) a combination of (1) and (2). One of the algorithms based on (3) uses a neural network approach and appears to be the most computationally efficient if a large number of realizations (of the order of several thousands) is required to adequately represent the stochastic water quality system.

The idea of including several scenarios of stream conditions simultaneously in a single water quality management model to produce robust solutions is not new. However, most of the studies have not developed reliability-cost (or cost surrogate) trade-off relationships that are often required by decision makers, in addition to other information about unmodelled issues, to identify and assess alternative management decisions. Furthermore, although the solutions produced by these studies are often shown to be

robust when compared with solutions for the individual single scenarios of stream conditions, these robust solutions may still be inferior to the 'true' reliability-cost trade-off relationship. To obtain efficient solutions (i.e., the non-inferior solution set), particular sets of realizations of stream conditions, whose precise identification is nontrivial, are to be included simultaneously in a water quality management model.

The MR model and the algorithms developed in this research may be used to select input information for a management model that utilizes more sophisticated pollutant transport process, such as QUAL2E and WASP5. Such an approach will apply the MR model and the solution algorithms to a management model which is linked to a simpler pollutant transport model, e.g., Streeter-Phelps equation for BOD-DO interaction in a stream. The set of important realizations identified at each reliability level of interest may then be incorporated in the management model that utilizes the sophisticated pollutant transport process.

There is the potential for several improvements in the solution algorithms developed for the MR model. These may include approaches that could identify and remove some of the unimportant constraints incorporated in the set of potentially critical realizations at each stage that the MR model is solved. Future work should investigate an objective approach for determining the value of κ in the NNIM Algorithm. This could reduce the CPU time required by the NNIM Algorithm without sacrificing the quality of solution obtained. Research should also be conducted to obtain other ANN architectures and input pattern representations that may improve the estimation of the level of criticalness of realizations included in the MR model, especially for cases in which small sizes of multiple training example sets are utilized. Such improved estimation performance may decrease the appropriate value of κ and the size of the optimization models solved at

each stage of the NNIM Algorithm.

There are few studies that explicitly include Type II uncertainty in seasonal water quality management models. Often, however, there are concerns with limited historical low flow and temperature which may be considered as the most important stochastic inputs for seasonal water quality management models. Since temperature data can often be estimated reasonably well for many locations, an approach that produces robust solutions with respect to short and sparse recorded low flow data is developed. This approach uses a one-stage non-stationary Markov model for the low flows at adjacent gauging stations. The seasonal UT management decisions from this Markovian approach are shown to be less sensitive to short and sparse recorded low flow data than those shown by the decisions from an existing approach. Thus, the Markovian approach may be more attractive for designing UT decisions for many water quality systems, which often have limited recorded low flow data.

The main drawback of the Markovian approach is dimensionality. Small flow intervals are required for accurate results. The smaller the intervals used, the higher the accuracy of the output, but the larger the dimension of the problem. The interval required to produce an acceptable level of accuracy is not addressed in this research. However, the dimensionality of the model could be greatly reduced at little or no cost to accuracy by considering small intervals for flow levels that are expected to produce the design fraction waste removal level or for the flow intervals that represent critical low flow conditions, and by considering relatively large intervals for the other flow levels. An initial application of the model can be performed with large flow intervals. Results of this initial application would suggest the magnitude of flows that would be associated with the design waste removal levels. These flows could therefore be divided into

smaller intervals for a second and more refined model application.

In addition to reliability, future improvement of the MR and Markov models may incorporate other indices of pollution control performance, such as resiliency, vulnerability, and robustness. The incorporation of these indices will answer questions related to other risk descriptors and failure modes associated with a water quality management system.

References

[1] Ambrose, R. B., Jr., T. A. Fitzpatrick, J. P. Connoly, and R. W. Schanz, WASP5.x A hydrodynamic and water quality model theory, user manual, and programmer's guide. Environ. Res. Laboratory, Office of Research Development, U. S. Environ. Protection Agency, Athens, Ga, 1991.

[2] Anderson, M. W., and H. J. Day, Regional management of water quality - A systems approach, *Res. Jour. Water Poll. Control Fed., 40*, 1679-1687, 1968.

[3] Arbabi, M., and J. Elzinga, A general linear approach to stream water quality modeling, *Water Resour. Res., 11(2)*, 191-196, 1975.

[4] Ash, T., Dynamic node creation in backpropagation networks, *ICS Rep. 8901*, Inst. of Cognitive Sci., Univ. of Calif., San Diego, La Jolla, 1989.

[5] Bayer, M. B., Nonlinear programming in river basin modelling, *Water Resour. Bull., 10(2)*, 311-317, 1974.

[6] Beavis, B., and M. Walker, Interactive pollutants and joint abatement costs: achieving water quality standards with effluent charges, *Jour. Environ. Econ. Manage., 6*, 275-286, 1979.

[7] Biswas, A. K., Model of the Saint John river, Canada, in *Models for Water Quality Management*, edited by A. K. Biswas, pp. 68-90, McGraw-Hill Inc., New York, 1981.

[8] Bogardi, J. J., W.-C. Huang, and R. Harboe, Simplified assessment of typhoon-borne floods for reservoir management, *IAHS Publ. No. 213*, pp. 373-379, edited by Z. W. Kundzewicz, D. Rosbjerg, S. P. Simonovic, and K. Takeuchi, IAHS Press, Wallingford, UK, 1993.

[9] Boner, M. C., and L. P. Furland, Seasonal treatment and variable effluent quality based on assimilative capacity, *Jour. Water Pollution Control Fed., 54(10)*, 1408-1416, 1982.

[10] Brill, E. D., Jr., J. C. Liebman, and C. S. ReVelle, Equity measures for exploring water quality management alternatives, *Water Resour. Res., 12(5)*, 845-851, 1976.

[11] Brill, E. D., Jr., J. W. Eheart, S. R. Kshirsagar, and B. J. Lence, Water quality impacts of biochemical oxygen demand under transferable discharge permit programs, *Water Resour. Res., 20(4)*, 445-455, 1984.

[12] Brown, L. C., and T. O. Barnwell, Jr., The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and User Manual, *Report EPA/600/3-87/007*, U.S. Environ. Protection Agency, Athens, Ga., 1987.

[13] Burn, D. H., Water-quality management through combined simulation-optimization approach, *Jour. Environ. Eng., ASCE, 115*, 1011-1024, 1989.

[14] Burn, D. H., and B. J. Lence, Comparison of optimization formulations for waste load allocations, *Jour. Environ. Engrg. Div., ASCE, 118(4)*, 597-612, 1992.

[15] Burn, D. H., and E. A. McBean, Optimization modeling of water quality in an uncertain environment, *Water Resour. Res., 21(7)*, 934-940, 1985.

[16] Burn, D. H., and E. A. McBean, Linear stochastic optimization applied to biochemical oxygen demand - dissolved oxygen modelling, *Canadian Jour. Civ. Eng., 13*, 249-254, 1986.

[17] Burn, D. H., and E. A. McBean, Nonlinear optimization modeling of coliform bacteria, *Water, Air, and Soil Poll., 32*, 183-200, 1987.

[18] Camp, T. R., *Water and its impurities*, Rheinhold Publishing, New York, N.Y., 1963.

[19] Cardwell, H., and H. Ellis, Stochastic dynamic programming models for water quality management, *Water Resour. Res., 29(4)*, 803-813, 1993.

[20] Dobbins, W. E., BOD and oxygen relationships in streams, *Jour. Sanitary Eng., Div., ASCE, 90(3)*, 53-78, 1964.

[21] Dorfman, R., Formal models in the design of water resources systems, *Water Resour. Res., 1(3)*, 329-336, 1965.

[22] Dysart, B. C., III. Water quality planning in the presence of interacting pollutants, *Res. Jour. Water Poll. Control Fed., 42*, 1515-1529, 1970.

[23] Dysart, B. C., III. and W. W. Hines, Control of water quality in a complex natural system, *IEEE Trans. Syst. Sci. Cybern., SSC-6*, 322-329, 1970.

[24] Eheart, J. W., E. F. Joeres, and M. H. David, Distribution methods for transferable discharge permits, *Water Resour. Res., 16(5)*, 833-843, 1980.

[25] Eheart, J. W., E.D. Brill, Jr., B. J. Lence, J. D. Kilgore, and J. G. Uber, Cost efficiency of time-varying discharge permit programs for water quality management, *Water Resour. Res., 23(2)*, 245-251, 1987.

[26] Ellis, J. H., E. A. McBean, and G. J. Farquhar, Chance-constrained/stochastic linear programming model for acid rain abatement- I. Complete colinearity and noncolinearity, *Atmospheric Environment, 19(6)*, 925-937, 1985.

[27] Ellis, J. H., E. A. McBean, and G. J. Farquhar, Chance-constrained/stochastic linear programming model for acid rain abatement- II. Limited colinearity, *Atmospheric Environment, 20(3)*, 501-511, 1986.

[28] Ellis, J. H., Stochastic water quality optimization using imbedded chance constraints, *Water Resour. Res., 23(12)*, 2227-2238, 1987.

[29] Fahlman, S. E., and C. Lebiere, The cascade-correlation learning architecture, *Rep. CMU-CS-90-100*, Sch. of Comput. Sci., Carnegie Mellon Univ., Pittsburgh, Pa., 1990.

[30] Ferrara, R. A., and M. A. Dimino, A case study analysis for seasonal nitrification: economic efficiency and water quality preservation, *Jour. of Water Pollution Control Federation, 57(7),* 763-769, 1985.

[31] Fiering, M. B., A screening model to quantify resilience, *Water Resour. Res., 18(1)*, 27-32, 1982a.

[32] Fiering, M. B., Alternative indices of resilience, *Water Resour. Res., 18(1)*, 23-39, 1982b.

[33] Fiering, M. B., Estimates of resilience indices by simulation, *Water Resour. Res., 18(1)*, 41-50, 1982c.

[34] Fiering, M. B., Estimating resilience by canonical analysis, *Water Resour. Res., 18(1)*, 51-57, 1982d.

[35] Frean, M., The upstart algorithm: A method for constructing and training feed-forward neural networks, *Neural Computation 2*, 198-209, 1990.

[36] Fuessle, R. W., Air quality planning: A general chance-constraint model, *Jour. Environ. Eng., ASCE, 113(1)*, 106-123,1987.

[37] Fujiwara, O., S. K. Gnanendran, and S. Ohgaki, River quality management under stochastic streamflow, *Jour. Environ. Eng., ASCE, 112(2)*, 185-198, 1986.

[38] Fujiwara, O., S. K. Gnanendran, and S. Ohgaki, Chance constrained model for river water quality management, *Jour. Environ. Eng., ASCE, 113(5)*, 1018-1031, 1987.

[39] Glantz, M. H., Consequences and responsibilities in drought forecasting: The case of Yakima, 1977, *Water Resour. Res., 18(1)*, 3-13, 1982.

[40] Graves, G., G. B. Hatfield, and A. B. Winston, Mathematical programming for regional water quality management, *Water Resour. Res., 8*, 273-290, 1972.

[41] Hall, G. C., and L. Ortolano, Modeling water distribution in Santa Clara county, Calif., *Jour. American Water Works Assoc., 71(5)*, 242-247, 1979.

[42] Hann, H. H., and R. G. Cembrowicz, Model of the Neckar river, Federal Republic of Germany, in *Models for Water Quality Management*, edited by A. K. Biswas, pp. 158-221, McGraw-Hill Inc., New York, 1981.

[43] Hargett, D., and J. A. Seagraves, *Benefits and costs of seasonal effluent permits in North Carolina.* North Carolina Water Resource Research Institute, Raleigh, N.C. 1979.

[44] Hashimoto, T., J. R. Stedinger, and D. P. Loucks, Reliability resiliency, and vulnerability criteria for water resource system performance evaluation, *Water Resour. Res., 18(1)*, 15-20, 1982a.

[45] Hashimoto, T., D. P. Loucks, and J. R. Stedinger, Robustness of water resource systems, *Water Resour. Res., 18(1)*, 21-26, 1982b.

[46] Hirose, Y., K. Yamashita, and S. Hijiya, Back-propagation algorithm which varies the number of hidden units, *Neural Networks, 4(1)*, 61-66,1991.

[47] Hornik, K., M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks 2(5)*, 359-366, 1989.

[48] Hwang, C. L., J. L. Williams, R. Shojalashkari, and L. T. Fan, Regional water quality management by the generalized reduced gradient method, *Water Resour. Bull., 9(6)*, 1159-1181, 1973.

[49] Johansson, E. M., F. U. Dowla, and D. M. Goodman, Backpropagation learning for multi-layer feed-forward neural networks using the conjugate gradient method, *Int. Jour. Neural Syst., 2(4)*, 291-301, 1992.

[50] Johnson, E. L., A study in the economics of water quality management, *Water Resour.Res., 3(2)*, 291-305, 1967.

[51] Karnin, E. D., A simple procedure for prunning back propagation trained neural networks, *IEEE Trans. Neural Networks, 1(2)*, 239-242, 1990.

[52] Karunanithi, N., W. J. Grenney, D. Whitley, and K. Bovee, Neural networks for river flow prediction, *Jour. of Comput. in Civ. Eng., 8(2)*, 201-220, 1994.

[53] Kerri, K. D., An economic approach to water quality control, *Res. Jour. Water Poll. Control Fed., 38*, 1883-1897, 1966.

[54] Kilgore, J. D., Seasonal static transferable discharge permits for the control of biochemical oxygen demand in the Willamette River, *Working Paper No. 2, NSF Award PRA 81-21692,* Dept. of Civ. Eng., University of Illinois, Urbana, III, 1985.

[55] Lamb, J. C., III, and D. B. Hull, Current status in use of flexible effluent standards, *Jour. Water Pollution Control Fed., 57(10)*, 993-998, 1985.

[56] Lence, B. J., Dynamic transferable discharge permits for the control of biochemical oxygen demand, *Working Paper No. 1,* Dept. of Civ. Eng., University of Illinois, Urbana, 1985.

[57] Lence, B. J., J. W. Eheart, and E. D. Brill, Jr., Risk equivalent seasonal discharge programs for multidischarger streams, *Proc. Am. Soc. Civ. Eng. Jour. Water Resour. Planning and Manage., ASCE, 116(2),* 170-186, 1990.

[58] Lence, B. J., and A. K. Takyi, Data requirements for seasonal discharge programs: An application of a regionalized sensitivity analysis, *Water Resour. Res., 28(7)*, 1781-1789, 1992.

[59] Liebman, J. C., *The optimal allocation of stream dissolved oxygen resources*, Ph.D. thesis, Cornell Univ., Ithaca, N. Y., 1965.

[60] Liebman, J. C., and W. R. Lynn, Optimal allocation of stream dissolved oxygen resources, *Water Resour. Res., 2(3)*, 581-591, 1966.

[61] Logan, J. A., W. D. Hatfield, G. Russell, and W. R. Lynn, An analysis of the economics of wastewater treatment, *Jour. Sanit. Eng. Div. Am. Soc. Civ. Eng., 34(9)*(S60), 1962.

[62] Lohani, B. N., and K. B. Hee, A CCDP model for water quality management in the Hsintien River in Taiwan, *Int. Jour. Water Resour. Dev., 1*, 91-114, 1983.

[63] Lohani, B. N., and A. R. Saleemi, Recent developments on stochastic programming model for water quality management, *Water Supply Manage., 6*, 511-520, 1982.

[64] Lohani, B. N., and N. C. Thanh, Stochastic programming model for water quality management in a river, Journal of the *Water Pollution Control Fed., 50(9)*, 2175-2182, 1978.

[65] Lohani, B. N., and N. C. Thanh, Probabilistic water quality control policies, *Jour. Environ. Div., ASCE, 105(EE4)*, 713-725, 1979.

[66] Loucks, D. P., J. R. Stedinger, and D. A. Haith, *Water Resource Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1981.

[67] Loucks, D. P., and R. Lynn, Probabilistic Models for predicting stream quality, *Water Resour. Res., 2 (3)*, 593-605, 1966.

[68] Loucks, D. P., C. S. ReVelle, and W. R. Lynn, Linear programming for water pollution control, *Management Science, 14(4)*, B-166-B-181, 1967.

[69] Michels, C. M, *Incorporating reserve capacity in the waste load allocation process*, M.Sc. thesis presented to the University of Illinois at Urbana-Champaign, Urbana, Illinois, 1987.

[70] Montgomery, W. D., Markets in licenses and efficient pollution control programs, *Jour. Econ. Theory, 5*, 395-418, 1972.

[71] Morgan, D. R., *Decision making under uncertainty using a new chance constrained programming technique: A groundwater reclamation application*, Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1990.

[72] Morgan, D. R., J. W. Eheart, and A. J. Valocchi, Aquifer remediation design under uncertainty using a new chance constrained programming technique, *Water Resour. Res., 29(3)*, 551-561, 1993.

[73] Mozer, M. C., and P. Smolensky, Skeletonization: A technique for trimming the fat from a network vis relevance assessment, *Rep. CU-CS-421-89*, Dep. of Comput. Sci., Univ. of Colorado, Boulder, 1989.

[74] Neter, J., W. Wasserman, and M. H. Kutner, *Applied linear regression models (2nd ed.)*, Richard D. Irvin Inc., Homewood, IL, pp. 62-87 and 145-151, 1989.

[75] Newsome, D. H., The Trent River Model - An aid to management, in *Modeling of Water Resource Systems*, edited by A. K. Biswas, pp. 490-509, Environment Canada, Policy Planning and Research Service, Ottawa, Ont., Canada, 1972.

[76] Nicholson, G. S., E. E. Pyatt, and D. H. Moreau, A methodology for selecting among water quality alternatives, *Water Resour. Bull., 6*, 23-33, 1970.

[77] O'Connor, D. J., and W. E. Dobbins, Mechanism of reaeration in natural streams, *Trans. Am. Soc. Civ. Eng., 123*, 641-684, 1958.

[78] O'Connor, D. J., and D. M. DiToro, Photosynthesis and oxygen balance in streams, *Jour. Sanit. Eng. Div. Am. Soc. Civ. Eng., 96*, 547-571, 1970.

[79] O'Neil, W. B., *Pollution permits and markets for water quality*, Ph.D Thesis, University of Wisconsin, Madison, Wisconsin, 1980.

[80] Orlob, G. T., *Mathematical Modelling of Water Quality*, Wiley/IIASA, Chichester, 1982.

[81] Park, H., and J. C. Liebman, Redundancy-constrained minimum-cost design of water-distribution nets, *Jour. Water Resour. Planning and Manage., ASCE, 119(1)*, 83-98, 1993.

[82] Ponnambalam, K., E. A. McBean, and T. E. Unny, Impacts of meteorological variations on acid rain abatement decisions, *Jour. of Environ. Eng., ASCE, 116(6)*, 1063-1075, 1990.

[83] Pratishthananda, S., and A. B. Bishop, A nonlinear multilevel model for regional water resources planning, *Water Resour. Bull., 13(3)*, 611-625, 1977.

[84] Ranjithan, S., J. W. Eheart, and J. H. Garrett, Jr., Neural network-based screening for groundwater reclamation under uncertainty, *Water Resour. Res., 29(3)*, 563-574, 1993.

[85] Reheis, H. F., J. C. Dozier, D. M. Word, J. R. Holland, Treatment cost savings through monthly variable effluent limits, *Jour. Water Pollution Control Fed., 54(8)*, 1224-1230, 1982.

[86] ReVelle, C. S., D.P. Loucks, and W.R. Lynn, Linear programming applied to water quality management, *Water Resour. Res., 4, (1)*, 1-9, 1968.

[87] Ritzel, B. J., J. W. Eheart, and S. Ranjithan, Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resour. Res., 30(5)*, 1589-1603, 1994.

[88] Rizzo, D. M., and D. E. Dougherty, Characterization of aquifer properties using artificial neural networks: Neural kriging, *Water Resour. Res., 30(2)*, 483-497, 1994.

[89] Roberts, M. J., *Environmental protection: the complexities of real policy choice, in Managing the water environment*, edited by N. A. Swainson, University of British Columbia Press, Vancouver, 157-235, 1976.

[90] Rogers, L. L., and F. U. Dowla, Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling, *Water Resour. Res., 30(2)*, 457-481, 1994.

[91] Rose-Ackerman, S., Market models for water pollution control: their strengths and weaknesses, *Public Policy, 25*, 383-406, 1977.

[92] Rossman, L. A., Risk equivalent seasonal waste load allocation, *Water Resour. Res., 25(10),* 2083-2091, 1989.

[93] Saad, M., A. Turgeon, P. Bigras, and R. Duquette, Learning disaggregation technique for the operation of long-term hydroelectric power systems, *Water Resour. Res., 30(11),* 3195-3202, 1994.

[94] Shih, C. S., System optimization for river basin water quality management, *Res. Jour. Water Poll. Control Fed., 42,* 1762-1804, 1970.

[95] Smith, E. T., and A. R. Morris, Systems analysis for optimal water quality management, *Jour. Water Pollution Control Fed., 41(9),* 1635-1646, 1969.

[96] Sobel, M. J., Water quality improvement programming problems, *Water Resour. Res., 1(4),* 477-487, 1965.

[97] Sobel, M. J., Chebyshev optimal waste discharges, *Oper. Res., 19,* 308-322, 1969.

[98] Somlyódy, L., M. Kularathna, and I. Masliev, Water quality management of the Nitra River Basin (Slovakia): Evaluation of various control strategies, *Working Paper, WP-93-63,* International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 1993.

[99] Streeter, H. W., and E. B. Phelps, A study of the pollution and natural purification of the Ohio River, *Bull. 146,* U.S. Public Health Serv., Washington, D. C., 1925.

[100] Rossman, L. A., Risk equivalent seasonal waste load allocation, *Water Resour. Res., 25 (10),* 2083-2091, 1989.

[101] Takyi, A. K., *Uncertainty analysis for environmental management models: Application of the generalized sensitivity analysis*, M.Sc. thesis submitted to the Faculty of Graduate Studies, University of Manitoba, Winnipeg, 1991.

[102] Tang, Z., and P. A. Fishwick, Feedforward neural nets as models for time series forecasting, *Jour. of Comput. Oper. Res. Soc. Am., 5(4)*, 374-385, 1993.

[103] Thomann, R. V., Mathematical model for dissolved oxygen, *Jour. Sanit. Eng. Div. Am. Soc. Civ. Eng., 89*(SA5), 1-30, 1963.

[104] Thomann, R. V., *Systems Analysis and Water Quality Management*, McGraw-Hill, New York, 1974.

[105] Thomann, R. V., and M. J. Sobel, Estuarine water quality management and forecasting, *Jour. Sanit. Eng. Div. Am. Soc. Civ. Eng., 90*(SA5), 9-35, 1964.

[106] Wagner, B. J., and S. M. Gorelick, Optimal groundwater quality management under parameter uncertainty, *Water Resour. Res., 23(7)*, 1162-1174, 1987.

[107] Wagner, B. J., and S. M. Gorelick, Reliable aquifer remediation in the presence of spatially variable hydraulic conductivity: From data to design, *Water Resour. Res., 25(10)*, 2211-2226, 1989.

[108] Warn, A. E., The Trent mathematical model, in *Mathematical Models in Water Pollution Control*, edited by A. James, chap. 17, pp 357-375, John Wiley, New York, 1978.

[109] Warren, J., and J. K. Bewtra, A model to study the effects of time-variable pollutant loads on stream quality, *Water Res., 8*, 1057-1061, 1974.

[110] Worley, J. L., *A system analysis method for water quality management by flow augmentation*, M.S. thesis, Oreg. State Univ., Corvallis, 1963.

[111] Yapo, P., S. Sorooshian, and V. Gupta, A Markov chain flow model for flood forecasting, *Water Resour.Res., 29 (7)*, 2427-2436, 1993.

Appendix A

The backpropagation (BP) learning algorithm

The BP algorithm finds the values of all the connection weights of a feed-forward ANN such that the quadratic error function (*Equation* 5.6) is minimized using the gradient descent method. For a given iteration of the training process, the error is used to adjust each connection weight by an amount given by *Equation* A.1 which performs a gradient on $\mathcal{E}$.

$$\triangle \mathcal{W}_{mn} = -\tau \frac{\partial \mathcal{E}}{\partial \mathcal{W}_{mn}} \tag{A.1}$$

where

τ = a constant of proportionality (learning rate)

The gradient of the error surface, $\frac{\partial \mathcal{E}}{\partial \mathcal{W}_{mn}}$, can be written as:

$$\frac{\partial \mathcal{E}}{\partial \mathcal{W}_{mn}} = \frac{\partial \mathcal{E}}{\partial \iota_m} \frac{\partial \iota_m}{\partial \mathcal{W}_{mn}} \tag{A.2}$$

From *Equation* 5.1, $\frac{\partial \iota_m}{\partial \mathcal{W}_{mn}}$ can be written as:

$$\frac{\partial \iota_m}{\partial \mathcal{W}_{mn}} = a_n \tag{A.3}$$

The first term on the right hand side of *Equation* A.2 can be written as:

$$\frac{\partial \mathcal{E}}{\partial \iota_m} = \frac{\partial \mathcal{E}}{\partial a_m}\frac{\partial a_m}{\partial \iota_m} \tag{A.4}$$

From *Equation* 5.3,

$$\frac{\partial a_m}{\partial \iota_m} = \mathcal{F}'(\iota_m) \tag{A.5}$$

To compute $\frac{\partial \mathcal{E}}{\partial a_m}$, two cases which assume unit m as (1) an output unit and (2) a hidden unit, are considered.

For case (1),

$$\frac{\partial \mathcal{E}}{\partial a_m} = -(t_m - a_m) \tag{A.6}$$

Here, $a_m = O_m$, therefore *Equation* A.6 becomes:

$$\frac{\partial \mathcal{E}}{\partial a_m} = -(t_m - O_m) \tag{A.7}$$

For case (2), the error measure can be written as a function of the net inputs from hidden to output layer, i.e., $\mathcal{E} = \mathcal{E}(\iota_1, \iota_2, \ldots, \iota_m, \ldots)$, and using the chain rule $\frac{\partial \mathcal{E}}{\partial a_m}$ can be written as:

$$\frac{\partial \mathcal{E}}{\partial a_m} = \sum_{q=1}^{Q_0}\frac{\partial \mathcal{E}}{\partial \iota_q}\frac{\partial \iota_q}{\partial a_m} = \sum_{q=1}^{Q_0}\frac{\partial \mathcal{E}}{\partial \iota_q}\frac{\partial}{\partial a_m}\sum_{r=1}^{R_q}\mathcal{W}_{qr}a_r = \sum_{q=1}^{Q_0}\frac{\partial \mathcal{E}}{\partial \iota_q}\mathcal{W}_{qm} = -\sum_{q=1}^{Q_0}(t_q - O_q)\mathcal{F}'(\iota_q)\mathcal{W}_{qm} \tag{A.8}$$

Combining *Equations* A.1, A.3, A.4, A.5, and A.7, the change in weight during a given iteration for cases in which the target unit is an output unit for the network is given as:

$$\triangle \mathcal{W}_{mn} = \tau \mathcal{F}'(\iota_m)a_n(t_m - O_m) \tag{A.9}$$

For cases in which the target unit is a hidden unit of the network, the weight adjustment is obtained from *Equations* A.1, A.3, A.4, A.5, and A.8 as:

$$\triangle \mathcal{W}_{mn} = \tau \mathcal{F}'(\iota_m) a_n \sum_{q=1}^{Q_0} (t_q - O_q) \mathcal{F}'(\iota_q) \mathcal{W}_{qm} \tag{A.10}$$

Generally, the weight adjustments in *Equations* A.9 and A.10 are computed recursively for all units in the network.

For the sigmoid activation function,

$$\mathcal{F}'(\iota_m) = \frac{\partial}{\partial \iota_m} \left(\frac{1}{1 + \exp^{-\iota_m}} \right) = a_m (1 - a_m) \tag{A.11}$$

therefore, *Equations* A.9 and A.10 become:

$$\triangle \mathcal{W}_{mn} = \tau a_m (1 - a_m) a_n (t_m - O_m) \tag{A.12}$$

and

$$\triangle \mathcal{W}_{mn} = \tau a_m (1 - a_m) a_n \sum_{q=1}^{Q_0} (t_q - O_q) a_q (1 - a_q) \mathcal{W}_{qm} \tag{A.13}$$

respectively.

A.1 Conjugate gradient modification of the BP learning algorithm

During a given iteration, the steepest gradient used by the traditional BP algorithm searches for the minimum error point in the direction of the gradient of $\mathcal{E}$, $\bigtriangledown \mathcal{E}$ (see, *Equation* A.1). A more efficient approach is the conjugate gradient method which uses a search direction that is a compromise between the direction of the gradient $\bigtriangledown \mathcal{E}$, and the previous search direction. Thus a new search direction d^{new} is given as:

$$d^{new} = \bigtriangledown \mathcal{E}^{new} + \delta d^{old} \tag{A.14}$$

where δ is obtained using the Polak-Ribiere rule (*Equation* A.14), so that the new search direction deteriorates as little as possible the minimization achieved by the previous search.

$$\delta = \frac{(\bigtriangledown \mathcal{E}^{new} - \bigtriangledown \mathcal{E}^{old}). \bigtriangledown \mathcal{E}^{new}}{(\bigtriangledown \mathcal{E}^{old})^2} \tag{A.15}$$