STOCHASTIC RESERVOIR OPERATION FOR WATER QUANTITY AND QUALITY USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES

ＡＩ手法を導入し水量・水質を対象とした
確率論的貯水池操作

Paulo Chaves

Doctoral Thesis

Department of Civil Engineering Systems
Graduate School of Engineering, Kyoto University

December 2004

For my mother, Ana

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Professor Toshiharu Kojiri, for all his teachings and support during the development of my research. I am grateful for the research topics he has proposed to me and his guidance in solving the proposed problems. During the past six years working under his supervision, in addition to his professional advice, I could also learn a lot about Japan, its culture and Japanese society. I greatly appreciate his concern, not only about my research activities, but many aspects of my life in Japan. More than a supervisor, he is a friend.

I am also grateful to Professor Shuichi Ikebuchi and Associate Professor Kunio Tomosugi of the Disaster Prevention Research Institute for their support throughout my stay at Kyoto University.

A special thanks to Professor Tomoharu Hori of the Department of Global Environment Engineering, Kyoto University, for his constructive and essential comments on the first drafts of this dissertation.

I am also grateful to Professor Tsuneo Tsukatani of the International Institute of Economics of Kyoto University for his encouragement and support for my participation in the international conference organized by IMACS in 2004. Also, to Professor Wen-Cheng Huang of the Department of Harbor and River Engineering of National Taiwan Ocean University for his much appreciated advice and encouragement during his three months as a visiting professor in my laboratory, regarding the first drafts of this dissertation. Furthermore, thanks to Professor Tetsuo Sawaragi of the Department of Precision Engineering of Kyoto University, for his valuable comments and suggested regarding the proposed CFNN.

I am thankful also to Dr. Yosuke Yamashiki and Dr. Vicente Santiago, my supervisors during my six-month period internship at the UNEP-IETC, Shiga office. During my master’s course, they played an important role in deciding the case study, putting me in contact with people and data from the region and also helping me throughout the two years. This case study has also been used in this dissertation. They also assisted with the physical model and concepts related to water quality modeling.

I would also like to thank Professor Jose Tundisi, Dr. Carlos Saggio and the International Institute of Ecology, Sao Carlos, Brazil, for their much appreciated support during my 3-day technical visit to the site of my case study, the Barra Bonita reservoir.

Thank you to the Ministry of Education of Japan (Monbusho) for the grant I received to pursue my research in the masters and doctoral courses at Kyoto University.

A sincere thanks to the Brazilian organizations CETESB, AES Tietê S.A. and INMET, which contributed useful data and information for my research.

I am thankful to the Water Resources Research Center (WRRC) of Kyoto University; the International Lake Environmental Committee and the International Institute of Economics of Kyoto University, for the financial support for my attendance and presentation at conferences in Singapore (2002), Argentina (2001) and Kazakhstan (2003), respectively.

I would like to offer a sincere thanks to all my professors at the Federal University of Rio de Janeiro, especially Professor Paulo Renato Barbosa and Professor Tarcisio de Castro, for the excellent education I received during my time as an undergraduate student in their care.

Thanks to my friends Sandra Chaves, Marc-Alain Widdowson and James Orton for their general comments and English proof-reading on some of the manuscripts for this dissertation.

Thanks to all my friends and colleagues at Kyoto University for constantly supporting and motivating me during this research.

Thank you to all my dear friends for their support and for making life in the country of the Rising Sun much more enjoyable, including Gaelle Pery, Galit Aviman, Jean Cluzel, Karen Harada, Mariko Takei, Michiro Ishikawa, Natsuko Tochiyama, Sigrid Vivo and Yuuka Nojima.

Finally, and most important, thanks to my wonderful family for their unconditional support and love.

ABSTRACT

In the last few decades most of the water storage systems, such as lakes and reservoirs, have suffered from poor water quality, mainly caused by human activities within their watersheds and inefficient operation of these systems. Even though there are a variety of techniques to improve water quality in storage, most are either very costly or not sustainable. Therefore, efforts to find new methods to achieve better water quality are extremely important. By taking advantage of the close relationship between the quality (limnological) and quantity (hydrological and climatological characteristics) of water, a better way to manage stored volumes is proposed to yield optimal benefits for water utilization while at the same time improving its quality.

The aim of this research is to integrate the management of water resources, that is to say storage reservoirs, considering quantity and quality issues. The goal is to generate a formulation for the integration of quantity and quality within multipurpose storage reservoirs, which can serve as a basis for practical improvements of environmental and water quality and future developments in the field by using state-of-the-art techniques of artificial intelligence (AI), such as artificial neural networks (ANN), fuzzy theory and genetic algorithm (GA). Long-term planning operation is addressed considering multiple objectives, such as flow stabilization, power generation and improvement of water quality. The analysis focuses on the planning operation of the reservoir for monthly time-step considering optimization and simulation models. The relationship between storage volume, release from a single outlet, and quality in the reservoir are considered as the main driving forces of the system.

Many works using dynamic programming (DP) in the optimization of water resources systems can be found in the literature, but they rarely address the water quality issues. The few works that attempted to operate reservoirs by stochastic DP (SDP) considering water quality, rarely integrate simultaneously water quality simulation and optimization models; instead they consider quality through transition matrix developed based on *a priori* simulated quality results. Moreover, in the water quality optimization models, usually only a few, one or two, quality parameters are considered as the increased number of quality parameters or state variables can result in a problem computationally intractable by DP. Besides the problem related to the curse of dimensionality, SDP models face other drawbacks, such as uncertainty due to the discretization of variables and the calculation direction, being the last responsible for the difficulties on integrating simulation and optimization models simultaneously. With the use of artificial intelligence techniques, such as ANN, the fuzzy sets theory and genetic algorithm (GA), a methodology to overcome the problems of DP models related to the operation of storage reservoir optimization, when water quality is also of concern, is proposed.

Artificial neural networks (ANN) have been applied successfully in various fields. However, ANN models depend on large sets of historical data and their use is limited when only vague and uncertain information is available, due to the low reliability of their results. Therefore, a conceptual fuzzy neural network (CFNN) is proposed, and applied as a water quality model to simulate the Barra Bonita reservoir system, Southeast region of Brazil. The CFNN model consists of a rationally-defined architecture based on accumulated expert knowledge about variables and processes included in the model. A genetic algorithm (GA) is used as the training method for finding the parameters of fuzzy inference as well as the connection weights. Fuzzy inference system handles the uncertainties related to the system itself, model parameterization, the complexity of the concepts involved and the scarcity and inaccuracy of data. CFNN showed greater robustness and reliability, when dealing with systems for which the data are considered to be vague, uncertain or incomplete.

A new approach for system optimization, named stochastic fuzzy neural network (SFNN), is proposed and it can be defined as a neuro-fuzzy system stochastically trained to yield a *quasi* optimal solution. The method intends to overcome some of the problems related to stochastic dynamic programming (SDP) models, such as the required backward calculation scheme. For validation of the proposed method, a real case of storage reservoir optimization and operation is considered with inflow discharge information as a stochastic variable. After comparing results with other dynamic programming (DP) approaches for the same optimization problem, the SFNN model showed improvements in the optimization results. Moreover, to deal with the uncertainties related to discretization of inflows in finding the stochastic inflows, we propose the use of conditional probability of fuzzy events. The multiple purpose characteristic of the system is handled with fuzzy theory to be able to easily compare different objectives.

After training and validation of both ANN based models, optimization and simulation, they are integrated for the optimization considering simultaneously water quantity and quality under different optimization schemes. The operating rules are presented as neural networks, which has been trained for the improvement of water quality, having initial storage and previous inflow as the net input information. The results clearly showed improvements in the water quality. The results are presented in such a way that they can be easily understood by operators.

要旨

最近の数十年で、湖や貯水池といった、貯水システムの大半は、流域における人間の活動や、非効率な貯水池システムの操作によって、水質の劣化に苦しんでいる。貯水池の水質を改善する技術は多く存在するが、大半は高価であるか、もしくは維持が困難であるのが実状である。そのため、より良い水質を達成するための新しい方法を見いだすことは、非常に重要である。そこで、水の質（陸水学）と量（水文学、気候学的特徴）の密接な関係を利用し、水利用にとって最適な利益を成し遂げると同時に水質を改善するために貯水システムを効率的に管理する方法を提案する。

この研究の目的は水資源の管理を総合化する、すなわち、質と量の問題を考慮した、貯水池の管理を統合化することである。最終目標は、多目的貯水池での水量と水質の総合のための定式を生みだすことである。これはニューラルネットワーク（ＡＮＮ）、ファジー理論、遺伝的アルゴリズム（ＧＡ）などの最新の人工知能の技術を用いることによって、この分野での将来の発展と環境、水質の実際的な改善の基礎として有用なものとなる。長期的計画操作は、水量安定、発電、水質などの、多目的に関して述べられている。解析は、最適化シミュレーションモデルに関して月単位の貯水の計画操作に焦点を当てる。貯水量、ひとつの放流口からの放水、貯水池の質の関係がこのシステムの主要素として考えられている。

水資源システムの最適化に関して動的計画法（ＤＰ）を使用したものは多くあるが、水質の問題に言及するものはほとんどない。また水質に関して確率動的計画法ＳＤＰを用いて貯水池を操作しようと試みたものはあるが、水質シミュレーションと最適化モデルを同時に統合したものはほとんどない。代わりに、前もってシミュレートされた水質結果に基いて発展された推移マトリックスを通して水質を考察している。さらに、水質最適化モデルでは、水質パラメーターや状態変数が増えると、ＤＰによって計算上追跡できない問題になるために、１つか２つの水質パラメータのみが考察されている。この次元性問題に加えて、ＳＤＰモデルは、変数の離散化や計算方向による不確定性といった他の障害に直面し、シミュレーションと最適化モデルを同時に統合することを困難にしている。ＡＮＮ、ファジー理論、ＧＡなどの人工知能技術を用い

ることによって、水量とともに水質も考慮する際の貯水最適化操作に関するＤＰモデルの問題を克服する方法が提案される。

ＡＮＮは様々な分野で応用されてきた。しかし、ＡＮＮモデルは過去のデータセットに依存しており、漠然とした不確定な情報だけが利用可能なときには、結果の信頼性が低いために、その使用は制限される。よって、ＣＦＮＮが提案され、ブラジルの南東地域のバラ・ボニタ貯水システムをシミュレートするための水質モデルとして応用される。概念的ニューラルネットワーク（ＣＦＮＮ）モデルは、モデルに含まれる変数と過程に関する蓄積された専門的知識に基く合理的に定義された構造からなる。ＧＡは結合荷重とファジー推論のパラメーターを見つけるための学習方法として使われている。ファジー推論システムはシステムそのもの、モデルのパラメーター化、概念に関連する複雑性、データの欠乏と不正確性に関する不確実性を扱う。ＣＦＮＮは、データが漠然であったり、不確実、不完全であると考えられるとき、システムを扱う上で、頑健さと信頼性を示した。

確率的ニューラルネットワーク（ＳＦＮＮ）と名付けられたシステムの最適化のため新しい方法が提案され、それは推計学的に準最適解を生みだすために訓練されたニューロ・ファジーシステムとして定義される。この方法は、必要とされる後方計算スキームといった、ＳＤＰモデルに関するいくつかの問題を克服するためのものである。この提案された方法の検証のために、実際の貯水池最適化と操作のケースが確率的変数として流入量の情報で考慮されている。同じ最適化問題に関して、他のＤＰアプローチの結果と比較すると、ＳＦＮＮは最適化の結果に改善が見られた。さらに、確率的流入量を見つける上での流入量の離散化に関する不確実性を扱うために、ファジー事象の条件付確率を使用することを提案している。この貯水池操作の多目的最適化は、容易に他の目的と比較できるようにファジー論理が扱われている。

ＡＮＮに基いたモデルである、最適化とシミュレーションを有効化と学習をさせた後、それらは、異なる最適化スキームの下で水質と水量を同時に考察する最適化のために統合される。操作規則は、初期貯水と前の流入量を入力情報として、水質改善のために訓練されたニューラルネットワークとして提示される。結果は明らかな水質の完全を示した。結果は技師に容易に理解されるように提示されている。

CONTENTS

ACKNOWLEDGEMENTS v
ABSTRACT vii
要旨 ix
LIST OF FIGURES xiii
LIST OF TABLES xv
SYMBOLS AND ABBREVIATIONS xvi
1 Introduction 1
1.1 Background 1
1.2 Research Objectives 2
1.3 Proposed Methodology 4
2 Water Quality in Storage 6
2.1 Water Quality and Storage 7
2.2 Water Quality Evaluation 9
2.3 Water Quality Simulation 11
2.4 Water Quality and its Uncertainty 12
2.5 Management of Storage Water Quality 14
2.6 Conclusions 15
3 Storage Reservoir Operation 17
3.1 Operation for Water Quality 17
3.2 Multi-objective Analysis 19
3.3 Operation with Fuzzy Objectives 20
3.4 Multistage Fuzzy Control and Optimization 21
3.4.1 Aggregation Operator 22
3.4.2 Fuzzy Type of Decision 22
3.5 Reservoir Optimization through Dynamic Programming 24
3.5.1 Deterministic, Probabilistic and Stochastic DP 26
3.5.2 Inverted and Noninverted form of the Mass (Continuity) Equation 27
3.5.3 Backward and Forward Calculation 28
3.6 Reservoir Operation through AI Techniques 30
3.6.1 Basic Characteristics of the Developed GA Model 32
3.7 Conclusions 34
4 Study Area 36
4.1 General information 36
4.1.1 Hydrological Situation in Brazil 36
4.1.2 Platina Basin, or Plata River Basin. 37
4.1.3 State of São Paulo 37
4.1.4 Tietê River Basin 38
4.2 Barra Bonita Reservoir 39
4.2.1 Basic Characteristics 39
4.2.2 In-site Inspection to the Reservoir 41
4.3 Data Collection 41
4.3.1 Hydrological Data 41
4.3.2 Water Quality Data 44
5 Water Quality Simulation Under Uncertainty 45
5.1 Introduction 45
5.2 Conceptual Fuzzy Neural Network 47
5.3 Architecture of a CFNN 47
5.3.1 Advantages & Applicability of CFNN 49
5.3.2 Disadvantages of CFNN 50
5.4 Mathematical Formulation 50

5.4.1 Fuzzification ... 50
5.4.2 Connection Type ... 51
5.4.3 Transfer Function ... 53
5.4.4 Defuzzification of $\mu_{o,k}$ ... 53
5.5 Performance Function ... 54
5.5.1 Performance Function (PF_{mse}) of MSE of Linguistic variables ... 57
5.5.2 Performance Function (PF_{vague}) of Output Vagueness ... 58
5.5.3 Aggregation of the Performance Functions ... 58
5.5.4 Training Algorithm ... 59
5.6 CFNN Water Quality Model ... 59
5.6.1 Basic Concepts of Reservoir Water Quality ... 60
5.6.2 Water Quality CFNN Model Architecture ... 63
5.7 Results ... 67
5.8 Conclusions ... 71
6 Operation for Quantity Objectives ... 73
6.1 Introduction ... 73
6.2 Methodology ... 74
6.3 Conditional Probability of a Fuzzy Event ... 75
6.4 Stochastic Fuzzy Neural Network (SFNN) ... 79
6.4.1 Fuzzy Neural Network Model ... 82
6.4.2 Operation Objectives ... 85
6.5 Validation of the SFNN Model by Comparison with DP ... 86
6.6 Results ... 88
6.7 Conclusions ... 94
7 Operation for Quality Objectives ... 96
7.1 Introduction ... 96
7.2 Simultaneous Combination of Simulation and Optimization Models ... 99
7.3 Water Quality Objectives ... 100
7.4 Considering Quality Objectives in the Reservoir Optimization ... 103
7.4.1 Case 1: Quantity and Quality Objectives are Considered Equally Important ... 104
7.4.2 Case 2: Quantity Objectives are Given Priority over Quality Ones ... 105
7.5 Results ... 105
7.5.1 Results for Case 1 ... 105
7.5.2 Results for Case 2 ... 107
7.6 Conclusions ... 109
8 Conclusions ... 111
REFERENCES ... 114
APPENDIX I: Total inflow assumed into Barra Bonita Reservoir (from AES) ... 121
APPENDIX II: Transition Matrix of conditional probabilities (Discrete Inflow Event) ... 122
APPENDIX III: Transition Matrix of conditional probabilities (Fuzzy Inflow Event) ... 124
APPENDIX IV: Attribute Maps of Tiete river Sub-basins (from CETESB) ... 126
APPENDIX V: Photos from Inspection to the Reservoir site (March, 2001) ... 130
APPENDIX VI: Other publications by the author ... 132

LIST OF FIGURES

Figure 1. Basic concept of integrated water resource management 1
Figure 2. Flowchart for the proposed water quality operation framework 5
Figure 3. Top and bottom release 17
Figure 4: System of reservoirs 18
Figure 5: Storage and water quality 18
Figure 6. Recursive equation or fuzzy type decision and its degree of conservatism........... 23
Figure 7. Backward DP (calculation: right to left; optimum sequence: left to right). 29
Figure 8. Forward DP (calculation: left to right; optimum sequence: right to left) 29
Figure 9. Triangular (cumulative) probability distribuition 33
Figure 10. Flowchart of training process using GA 34
Figure 11. Main Brazilian watersheds... 36
Figure 12. Sub-basins within the state of São Paulo (source SIGRH - Sao Paulo State) 38
Figure 13. Schematic profile of the Barra Bonita dam site with volumes and elevations 39
Figure 14. Location of dams in the Tietê river... 40
Figure 15. Storage variation for Barra Bonita reservoir (1970 – 2000) 42
Figure 16. Location of observation points of the Barra Bonita reservoir system 42
Figure 17. Consolidation of observed data considering the water balance 43
Figure 18. Average storage and discharges of Barra Bonita reservoir................................ 43
Figure 19. Components of the layer structure of CFNN between two consective layers 49
Figure 20. Membership function for the fuzzification of input x_i of a layer structure 51
Figure 21. Rule inference for each connection type.. 52
Figure 22. Graphic representation of the resulted fuzzy number $\mu_{y,k}$ 53
Figure 23. Performance function PF_{mse} as a function of MSE. 57
Figure 24. Basic processes and relations considered for the CFNN quality model. 61
Figure 25. Architecture of the reservoir water quality CFNN based model. 63
Figure 26. Climatological and hydrological input data. 64
Figure 27. Transfer Matrix for the CFNN quality model: from input to hidden layers 66
Figure 28. Transfer Matrix for the CFNN quality model: from hidden to output layers 66
Figure 29. Evolution of the GA final performance function 67
Figure 30. Results of DO for training and validation.. 67
Figure 31. Results of BOD for training and validation 67
Figure 32. Results of TN for training and validation 68
Figure 33. Results of TP for training and validation.. 68
Figure 34. Results of CHA for training and validation 68
Figure 35. Results of hidden unit “BOD input” and the observed calculated values 69
Figure 36. Results of hidden unit “TN input” and the observed calculated values.............. 69
Figure 37. Results of hidden unit “TP input”and the observed calculated values 69
Figure 38. BOD concentration (observed and calculated values) x BOD load (observed) .. 70
Figure 39. DO x BOD concentrations within the reservoir.. 70
Figure 40. Results of the hidden unit “DO consumption”. 70
Figure 41. Results of the hidden unit “growth”. 71
Figure 42: Correlogram for the inflow discharge... 75
Figure 43. Discretization of discharge: discrete (upper) and fuzzy (lower) intervals........... 76
Figure 44. Flowchart of the calculation procedures for the SFNN 79
Figure 45. Example for the calculation of the expected probability P_{exp} 80
Figure 46. Architecture of the proposed neuro-fuzzy inference system 83
Figure 47. Membership function for the fuzzification of input variables x_i......................... 84

Figure 48. Fuzzy objective membership functions for the reservoir operation 86
Figure 49. Relation between storage discretization and DP optimization results 87
Figure 50. Cumulative CP assuming discrete intervals... 89
Figure 51. Cumulative CP assuming fuzzy linguistic variables... 89
Figure 52. Results after five optmization schemes (weighted-average aggregation) 89
Figure 53. Resuts of optimal useful storage after three different optimization schemes 90
Figure 54. Results of operation using the SFNN model 91
Figure 55. Results of objective function versus GA generations 91
Figure 56. Results from five different optmization schemes (product aggregation) 91
Figure 57. Objective function results after six SFNN formulations - training...................... 92
Figure 58. Objective function results after six SFNN formulations - validation 92
Figure 59. Results of stabilization (left) and hydropower (right) evaluation functions........ 93
Figure 60. Minimum value of objective function for six SFNN formulations 93
Figure 61. Conceptual integration of ANN based optimization and simulation models 99
Figure 62: Flowchart for the SFNN model considering water quality objectives 100
Figure 63: Fuzzy membership functions for reservoir water quality evaluation 101
Figure 64. Evaluation of fuzzy results 103
Figure 65: Quantity (left) and Quality (right) objectives after Training for Case1............. 106
Figure 66: Quantity (left) and Quality (right) objectives after Validation for Case1 106
Figure 67: BOD concentration (left) and BOD evaluation function (right) for Case 1 107
Figure 68: CHA concentration (left) and CHA evaluation function (right) for Case 1 107
Figure 69: Optimal storage sequence for training (left) and validation (right) for Case 1 . 107
Figure 70: Quantity (left) and Quality (right) objectives after Training for Case2............. 108
Figure 71: Quantity (left) and Quality (right) objectives after Validation for Case2 108
Figure 72: Quantity (left) and Quality (right) objectives against time after Validation for Case 2 108
Figure 73: Optimal storage sequence for training (left) and validation (right) for Case 2 . 109

LIST OF TABLES

Table 1. Basic characteristic of Tietê river sub-basins...38
Table 2: Water quality with the central part of the reservoir near to the dam 44
Table 3. Characteristics of the CFNN compared to the traditional ANN 50
Table 4. Three methods for defining the center values of fuzzy membership of linguistic variables *k*...54
Table 5. Center values used for the definition of representative numbers 56
Table 6. Center values assumed for the membership functions of input variables 66
Table 7. Cumulative (unconditioned) probabilities of monthly inflows 76
Table 8. Water quality optimization considering the hydraulic components and methods ..97
Table 9. Attributes for quality parameters in the evaluation of trophic state 102
Table 10. Attributes for quality parameters in the evaluation of aquatic life preservation.102
Table 11. Comparison between Quantity and Quality objectives 104

SYMBOLS AND ABBREVIATIONS

Some of the symbols used in this dissertation are listed below. It is difficult to avoid using the same symbol to represent more than one concept, therefore only symbols which have only one meaning are listed. The other symbols are defined when ever used, so no confusion should result.

Abbreviations:

AI	artificial intelligence
ANN	artificial neural networks
BOD	biochemical oxygen demand
CFNN	conceptual fuzzy neural network
CHA	chlorophyll-a
CP	conditional probability
DO	dissolved oxygen
DP	dynamic programming
EUTRO	index for CHA, TP and TN
GA	genetic algorithm
OXYG	index for BOD and DO
POW	fuzzy objective for hydropower generation
SDP	stochastic DP
SFNN	stochastic fuzzy neural network
STAB	fuzzy objective for flow stabilization
TN	total nitrogen
TP	total phosphorous
UNEP	United Nations Environment Programme

Symbols:

E	evaluation function
F	recursive function
I	inflow
L, M, H	linguistic variables representing low, medium and high, respectively
P	probability
PF	performance function
R	release
S	storage
W	weights
X, V	linguistic variables representing extreme and very, respectively
μ	membership function
r	representative value
α and β	weights

1 Introduction

1.1 Background

Socio-economic and environmental issues related to water quantity and quality have highly complex aspects associated with them. Such issues include increasing water demand, multiple-use, water pollution due to rapid urbanization, eutrophication of water bodies, and increasing costs related to water treatment. Assessment of seasonal, social and ecological changes for practical reservoir management has become more relevant during recent decades. Use of statistics, optimization and Artificial Intelligence (AI) techniques, such as fuzzy set theory, genetic algorithm (GA) and artificial neural network (ANN), in the development of decision support systems, has proved to be a suitable method for this analysis.

The history of reservoir management has been developed basically focusing on single-objective and single-purpose problems. This approach has been used in many countries worldwide, based mostly on economic cost-benefit analysis. As a consequence, many water related problems have arisen, creating the need for a better and more comprehensive evaluation process. The consideration of various usage purposes, such as recreation, environment and navigation, has to be addressed from an integrated water resource management point-of-view. The basic relation between the water sector and other sectors is shown in Figure 1.

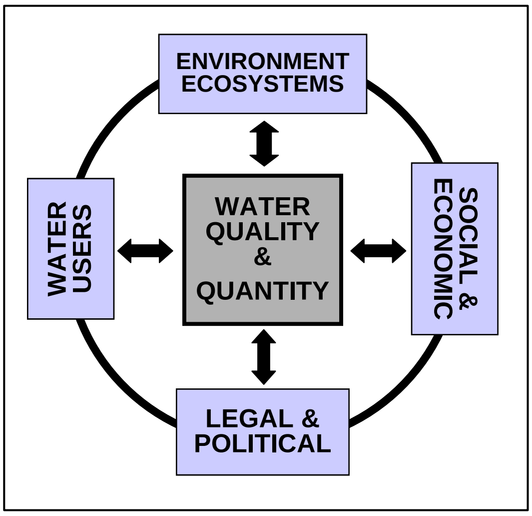


Figure 1. Basic concept of integrated water resource management

The integrated water resources management should include different sectors keeping a holistic view towards the development and problems related to the water sector. In this way, the reservoir plays an important role as one of the most important sources of freshwater for the environment and human socio-economic activities. The management and operation of reservoirs should focus on multiple purpose and multiple stakeholders, including water quality condition as well, which can enormously influence other sectors, such as the environment, human health and the aesthetic characteristics of water bodies.

But how may water quality be included in the reservoir operation framework? How can we simulate and evaluate water quality? How do we deal with the variability and uncertainties relating to the analysis of water quality data and model results? And in a more concrete way, to what extent is water quality influenced by reservoir operation and vice-versa? Can water quality in storage and release be improved by the more efficient operation of reservoirs? These are just some of the questions we intend to discuss and propose appropriate solutions to hereafter.

1.2 Research Objectives

The aim of this research is to integrate the management of water resources, more specifically storage reservoirs, considering aspects of quantity and quality. A formulation for the integration of water quantity and quality within a multipurpose storage reservoir is proposed, which can serve as a basis for practical improvements in environmental and water quality and assist further future developments in the field. Long-term planning operation is addressed considering multiple objectives, such as flow stabilization, power generation and water quality characteristics. By taking advantage of the close relationship between the quality (limnological) and quantity (hydrological and climatological characteristics) of water, proper management of stored volumes is carried out to yield optimal benefits for water utilization while at the same time improving its quality.

Barra Bonita reservoir is chosen as the study area for the application of the proposed methodology. The reservoir is located in the upper part of the Tiete river basin which suffers from severe water quality-related problems. Some other attempts to improve water quality condition through the more efficient management of water resources in this region has been carried out, such as by Azevedo *et al.* (2000), who attempted to integrate the evaluated results on water quality and quantity for the whole river basin. They presented water quantity optimization, followed by quality simulation for different scenarios applied for the hydrologic net of the Tiete river basin. However, the optimization and simulation models were not simultaneously integrated and water quality within the reservoirs was not considered.

One of the most applied techniques for storage reservoir optimization and operation is definitely dynamic programming (DP). Many works using DP in the optimization of water resources systems can be found in the literature, such as in Yakowitz (1982), Esogbue (1989), Kelman and Stendinger (1990), Kojiri *et al.* (1993), Simonovic (1991) and Nandalal and Bogardi (1995). However, water quality is rarely taken into account. Nevertheless, some attempts to use DP for the optimization of water quality can also be found, such as in Lohani *et al.* (1983) and Cardwell *et al.* (1993). However, simultaneous integration of water quality simulation and optimization models is seldom considered. Moreover, in the water quality optimization models, usually only few, one or two, quality parameters are considered as the increase in the number of quality parameters or state variables correspond to a increase in problem size which may soon become computationally intractable by DP models (Cardwell *et al.*, 1993).

Chaves *et al.* (2003a, b) and Chaves *et al.* (2004) successfully combined DP optimization models with simultaneous water quality simulation. However, to be able to do so, some assumptions had to be made due to some limitations related to the stochastic DP, such as the curse of dimensionality and the backward-moving calculation in which SDP has to be realized (Chaves *et al.*, 2003a, b). Further explanation about the limitations of DP models is given in Chapter 3.

With the use of artificial intelligence techniques, such as ANN, fuzzy set theory and genetic algorithm (GA) – evolving neuro-fuzzy system, we propose a methodology to overcome the problems of DP models relating to the operation of water quality in storage reservoir optimization. In Chapter 2, we discuss the main points relating to the quality of water in storage, such as the relation between quantity and quality, water quality modeling, uncertainty analysis and aspects of the decision-making process in water quality management. This is followed by Chapter 3, which presents some of the main topics related to the operation and optimization of reservoirs, including some aspects related to DP and multistage fuzzy optimization. Chapter 4 introduces the case study area and includes a general description about the Tiete river basin and data situation. In Chapter 5, a new water quality simulation model based on a neuro-fuzzy system is proposed. Then, Chapter 6 introduces a new stochastic optimization method, based on ANN, used in an attempt to overcome the problems that arise when combining DP and simulation models. Finally, in Chapter 7, both proposed models for simulation and optimization, are combined for the simultaneous optimization of storage considering water quantity and quality. Final conclusions and potential future research are presented in Chapter 8. The next section provides further explanation about the proposed methodology and the techniques developed.

1.3 Proposed Methodology

Techniques such as the dynamic programming (DP), time series analysis, Markov chain process, genetic algorithm (GA), artificial neural networks (ANN) and the fuzzy sets theory are used throughout this research. Each of these techniques is applied with different degrees of complexity or combined with one another. In this section, a general explanation about the two new developed techniques is provided. First, the conceptual fuzzy neural network (CFNN) is proposed for the simulation of water quality within the reservoir. Then, the stochastic fuzzy neural network (SFNN) is introduced briefly. This section concludes with a flowchart representing the connections among the different components, such as simulation and optimization models, of the system, and the overall proposed water quality optimal operation framework.

Artificial neural networks (ANN) have been applied successfully in various fields. However, ANN models depend on large sets of historical data and are of limited use when only vague and uncertain information is available, which leads to difficulties in defining the model architecture and low reliability of its results. Therefore, in Chapter 5, a conceptual fuzzy neural network (CFNN) is proposed, and applied for the simulation of water quality of the Barra Bonita reservoir system, located in the Southeast region of Brazil. The CFNN model consists of a rationally-defined architecture based on accumulated expert knowledge about variables and processes included in the model. A genetic algorithm (GA) is used as the training method for finding the parameters of fuzzy inference and the connection weights. The fuzzy inference system handles the uncertainties related to the system itself, model parameterization, complexity of the concepts involved and the scarcity and inaccuracy of data. CFNN tends to present more robustness and reliability than traditional ANN models, when dealing with systems for which data are considered to be vague, uncertain or incomplete. The CFNN model structure is easier to understand and to define when compared to other ANN based models. Moreover, it can help to understand the basic behavior of the system as a whole, being a successful example of cooperation between human and machine.

In Chapter 6, we introduce a new approach for system optimization, named stochastic fuzzy neural network (SFNN), which can be defined as a neuro-fuzzy system which is stochastically trained to yield a *quasi* optimal solution. The method intends to overcome some of the problems relating to stochastic dynamic programming (SDP) models, such as the rough discretization of state variables, the increase in the number of state variables and limitations that may arise due to the backward calculation scheme required by SDP models. For validation of the proposed method, an actual case of storage reservoir optimization and operation, with stochastic inflow discharge information, is considered. After comparing the results to other traditional dynamic programming (DP) approaches, for the same optimization problem, the SFNN model showed improvements in the optimization results. Then, it is combined with the water quality simulation model for the reservoir optimal operation, when the improvement of water quality is of concern. Moreover, to deal with the uncertainties relating to the discretization of inflows in finding the conditional probability of inflows, we propose the use of conditional probability of a fuzzy event. The multiple purpose characteristic of the system is handled by fuzzy theory to be able to compare different objectives easily. The overall flowchart for the proposed reservoir water quality operation framework is presented in Figure 2.

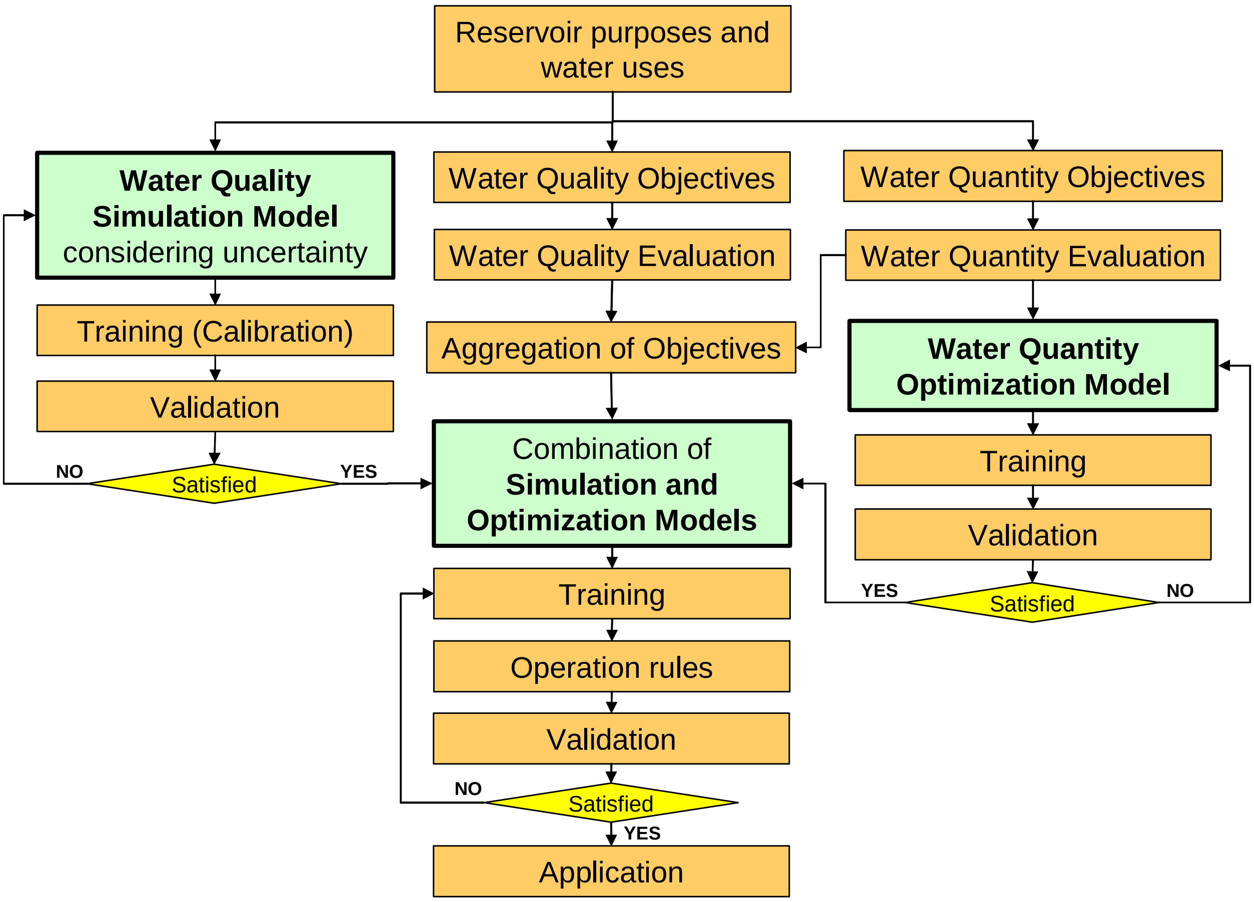


Figure 2. Flowchart for the proposed water quality operation framework

2 Water Quality in Storage

Water is the most precious resource found in nature. It is not only indispensable for the survival of all living beings, but it is also an important contribution for many socio-economic activities. It is not a coincidence that most of the first human settlements were located near areas rich in freshwater resources. However, due to population growth, increase of standard of living, urbanization and development of extensive areas for agriculture, the available water resources may become insufficient to fulfill all the needs of human living, particularly during dry periods which are caused by the natural randomness of the hydrological cycle.

Therefore, it becomes necessary to find and to develop new sources of freshwater in order to secure the constant and appropriate supply of water. Certainly, there are many methods to maintain the stable supply of water, such as groundwater exploitation and water deviation from wet to dry areas. But undoubtedly, the most popular way worldwide to stabilize water supply is through the operation of stored water, particularly in lakes and manmade reservoirs. Reservoir construction has a long history, with the first dam believed to be built in ancient Egypt around 2800 B.C. (Biswas, 1970).

In recent decades, most of these water storages have been suffering from poor water quality conditions, mainly caused by human activities within the watersheds and inefficient operation of these reservoirs. Furthermore, the quality of water in storage has been severely jeopardized. Even though there are a variety of techniques to improve the water quality, most of them are very costly and are often not sustainable. Therefore, finding new alternatives and methods that could achieve better water quality are extremely important. Taking advantage of the close relationship between the quality and quantity of water and knowing that quality can directly affect most of water uses, the objective of this research is to propose an efficient way to better operate the stored water to yield the optimal benefits for water utilization while at the same time trying to improve its quality.

In the next sections, the main issues regarding the quality of water in storage are presented. First, the important relation between quantity and quality of water, and the potential improvement of water quality through the proper operation of storage volumes are highlighted. Then, the main issues related to evaluation of water quality are explained, showing the importance of proper water quality for different uses. Next, a general introduction to water quality simulation models is given followed by a discussion about the uncertainties involved in the water quality analysis.

Finally, all these topics are combined within the storage management and operation framework when water quality is considered, emphasizing the recent techniques for the multiobjective optimization of multipurpose reservoirs.

2.1 Water Quality and Storage

This section intends to give a brief explanation of the relation between the quantity and quality of water, emphasizing the potential use of better operation of water quantity to achieve potential improvements in water quality.

Manmade reservoirs and lakes used for water storage, like other water bodies, are vulnerable to escalating and more frequent water quality related problems. Some of the most common problems concerning the quality of water in storage are organic pollution, anoxia, eutrophication, toxicity, turbidity, siltation and water-born diseases. Similarly, there is a variety of possible causes regarding such problems, such as industrialization, an increase in population, the over-exploitation of agriculture soil, bad agricultural practices, poverty, a lack of proper sanitation systems, runoff from cities, deforestation, a lack of proper enforcement of regulations for pollution control and an overuse of water resources with a decrease in water volumes and levels.

Probably, the most effective method to control pollution and improve the quality of water in storage is through a comprehensive watershed management, which includes reduction and control of non-point and point sources of pollution, such as nutrient and organic matter loads. On the other hand, there are other techniques that may be applied directly to the water bodies themselves such as artificial mixing, aeration, sediment removing, chemical manipulation, algaecides and bio-manipulation of species like fishes and aquatic plants. However, most of these techniques are costly and difficult to apply; therefore the proposal of new methods becomes extremely important.

Basically, there are three main components of a reservoir system: inflow, storage and release. Inflows may be defined as natural events (e.g. when upstream does not exist), whereas storage and release may be considered as under human control. Nevertheless, both natural variability and human decision can greatly influence the water quality of a reservoir. Reservoirs tend to present water of different quality during wet and dry hydrologic periods. With the increased runoff during periods of rain, higher nutrient loads would be expected in the reservoir, and these nutrients stimulate algae blooms and therefore may lower the water quality. Moreover, higher levels of inflow may also increase the input of organic matter resulting in DO depletion, which is defined as the decrease of dissolved oxygen concentrations, vital for sustaining aquatic life. On the other hand, higher inflows increase flushing rates and enhance mixing characteristics, decreasing stratification and potentially improving water quality. Hydrological factors, however, are not the only influences on water quality. Climatologic characteristics also play an important role, since during warm periods there is a higher probability of reservoir thermal stratification, due to a difference in temperature between water layers, leading to a difference in water density. The phenomenon of stratification, defined as the difference in density through the water column which is commonly caused by variations of temperature, restricts the mixing of water between different layers. The isolated lower layer, called hypolimnion, can suffer from DO depletion by anaerobic processes and an increase in nutrients and other materials from bottom sediments.

Water quality is closely tied to storage volumes. Clearly an increase of water volume results in increased dilution of pollutants but in addition, other physical, chemical and biological processes that influence the quality of water may be closely related to controllable hydrological factors such as the variation of release discharge, storage volumes and water levels. For example, an increased retention time (defined as the ratio between reservoir volume and discharge rate) can influence water quality in many ways, such as decreasing mixing proprieties, increasing phosphorous retention, increasing the difference between surface-bottom temperatures (stratification) and decreasing turbidity that can enhance chances of algae bloom. Moreover, fluctuation of water levels may also enormously affect water quality, such as increased erosion from the shores that may result in higher turbidity, increasing washout of margins and decreasing reproduction of some fish species (which lay eggs on costal vegetation) consequently changing the phytoplankton (algae) population of the aquatic system. The quality of stored water may also be affected by the release of water of poor quality from different levels, such as bottom-spillway discharge.

Zalewski *et al.* (1997) emphasizes the importance of ecohydrology to use ecosystem properties as management tools for sustainability of water resources, such as through adequate management of quantities to improve water quality. Moreover, his paper presents an example of the relationship between storage levels and the eutrophication process for the Sulejow Reservoir. Straskraba and Tunidisi (1999) discuss in a general way, many aspects relating to the management of reservoir water quality, such as eco-technologies, the relation between storage and water quality, the possible impact of different volumes and the quality of discharged water from reservoirs.

Therefore, the proper operation of release discharges and storage volumes considering single or multiple outlets and offtakes can also be considered an efficient method to improve the quality of stored water, by discharging water of undesirable quality or by increasing the storage volume with additional water, hence diluting water of poor quality (Welch and Patmont, 1980 and Chaves *et al.*, 2003a, b). Even though, some general idea of the relationships between hydrologic and climatologic variables and water quality condition could be described, it is not possible to make a definitive statement about such relations. Moreover, specific studies for each reservoir system are needed to truly understand its limnology characteristics. Such studies need to include data collection and analysis, application of mathematical models, biological and hydrologic investigations and lastly, an understanding of the socio-economic characteristics of the reservoir basin.

It becomes necessary to have a flexible model able to represent the quality condition based on the available information, such as inflow, retention time and air temperature. Therefore, the quality model proposed in Chapter 5 is intended to fulfill these needs.

2.2 Water Quality Evaluation

Water quality directly affects virtually all water uses, such as a) fish survival, diversity and growth; b) recreational activities; c) municipal, industrial and domestic water supplies; d) agricultural uses like irrigation and livestock; e) waste disposal; f) corrosion of turbines of hydropower plants and g) general aesthetics. All of these uses are affected by the physical, chemical, biological and microbiological conditions that exist in water bodies. Some water uses may affect water quality which can influence other water uses, such as navigation that may increase bank erosion increasing turbidity. This is of considerable interest in regards to water clarity and light penetration which is important for phytoplankton growth. Water bodies used for waste water assimilation may suffer from DO depletion and an increase in total dissolved solids, which are of main interest in terms of sustaining aquatic life and water supply, respectively.

The minimum acceptable quality of water depends very much on the water use itself. Water for irrigation, for example, should be low in dissolved salts, but water intended for livestock should be low in bacteria. The water used in industrial processes should usually be of a much higher quality than the water used for industrial cooling. As for municipal supply, water must not only be safe to drink, but ideally should contain low concentrations of materials such as calcium, iron or similar materials, as they may cause costly infrastructure damage or add unpleasant characteristics to the water even after treatment (Kiely, 1997).

The best efforts of professionals to define the quality characteristics of water that are necessary or desirable for various specific uses are summarized periodically in the form of “water quality index or criteria”. The water body in question with the appropriate criteria provides a basis for judging if the resource would be satisfactory for the beneficial uses that are under consideration (Lamb, 1985). A water quality index may be an efficient way to objectively understand the water quality situation under the vast number of existing water quality parameters, helping decisions to be made in a more efficient and less subjective way. Many of the water quality indices tend to be developed based on local characteristics and local expertise (Thanh and Biswas, 1990). Water quality indices may be also developed to express the conformity or net benefits of a water body regarding the particular type of water use. Indices can be based on either a single parameter or a combination of many quality parameters, such as the well-known Water Quality Index developed by the American Environment Protection Agency (EPA). Therefore, their utilization in the management and operation framework has to be closely studied and discussed among all stakeholders involved in the decision making process. Due to the vagueness involved in defining these indices, some new attempts to evaluate water quality are carried out based on fuzzy theory, refer to Norwick and Turksen (1984), Dombi (1990) and Chang, Chew and Ning (2001).

For a case study, it is decided to use the knowledge of local authorities regarding the evaluation of each individual parameter or the index developed by them. For the sake of simplicity among the large variety of quality parameters, only five are chosen: dissolved oxygen (DO), biochemical oxygen demand (BOD), total nitrogen (TN), total phosphorous (TP) and chlorophyll-a (CHA). These indices are used worldwide for the evaluation of water quality. The first two are selected due to their importance regarding the aquatic life preservation and the last three for their relevance on the evaluation of water trophic state. The limits used for the elaboration of each evaluation function are also based on real values used for the public authorities responsible. More detail on the development of the water indices used in this research can be found in Chapter 7.

2.3 Water Quality Simulation

Particular mathematical models can play an important role in the decision making process, as these tools may support the decision makers in making the "best" decisions, as the outcome of different decisions may be simulated and explored virtually with relative low costs and in a short period of time. Chaves *et al.*, 2003a presents an operation model combined with a physical model for the optimization of water quality within a reservoir.

There are a variety of mathematical models which are able to assess the water quality situation in reservoirs and lakes. These models may present different targets and may be developed for particular goals, such as mathematical calculation of a) stratification condition based on the heat balance of the water body that can help in the estimation of the thermocline and the temperature profiles, which can greatly influence water quality; b) trophic condition that can consider quality parameters such as phosphorous, nitrogen and chlorophyll which can be used for the estimation of algae blooms; c) dissolved oxygen that is crucial for the survival of most aquatic species, in which factors such as organic matter loads, suspended sediments and aeration characteristics of the water body play an import role; and d) dispersion models that, for example, can help to understand the movement of pollutants after a chemical accident leaking into a reservoir.

These models vary enormously; hence they can be classified in different groups, relating to their complexity, structure and applicability. The most common type of quality model is deterministic, which provides a single value for each situation and unit of time. Stochastic models, which calculate a range of values that can be expected for each situation and unit of time, are also available but are usually much more complex. On the other hand, they are usually more efficient in their simulation due to the high degree of uncertainty encountered in the quality simulation and natural processes.

Today we can count on quality models that are rather sophisticated, known as physical models, which are intended to describe theoretically many of the existing natural processes through a structure based on physical, chemical and biological concepts and relations. Therefore, many experts believe that the use of physical models may be extrapolated for future unknown situations or areas presenting a lack of data. For a more detailed explanation about the physical, biological and chemical processes behind the mathematical models, see Jorgensen (1983), Orlob (1983) and Thomann and Mueller (1990).

However, the determination of physical parameters for the calibration of such models may be a complex task. On the other hand, input/output or data-driven models, such as regression models and artificial neural network (ANN) models (Maier and Dandy, 1996 and Schleiter *et al.*, 1999), can be used in situations where data is fairly available. Once calibrated (trained), due to their low computational requirements, these models may be easily and efficiently combined simultaneously with the optimization model (Chaves *et al.*, 2004). Still, in recent years, some works have intended to overcome the problems of these two types of models by, for example, combining the characteristics of both models and introducing physical concepts into ANN models, such as in works by Towell and Shavlik (1994).

It is important to clarify that any model is just an attempt to represent the real world through approximation and abbreviation which intends to describe perfectly the unsolvable complexity of the ecological systems. Therefore, any simulation model will have a certain degree of imprecision and uncertainty involved in its results. To effectively incorporate the model's results into the storage management and operation framework, it is crucial to consider the involved uncertainties. Beck (1987) presents an excellent review about the uncertainties related to the water quality modeling. He also suggests that an efficient way to reduce such uncertainties could be through a model capable of using linguistic variables; maybe what he had in mind is exactly the one proposed in Chapter 5. In the next section, we shall discuss further the topics relating to the uncertainty of analysis of water quality before introducing the actual issue of storage operation.

2.4 Water Quality and its Uncertainty

The objective of this section is to discuss the main issues related to uncertainty in water quality modeling and management and how this can affect the decision making process of storage operation. But, where does uncertainty come from? Basically, uncertainty is inherent in many processes such as measurement, natural randomness and human perception. In a more concrete way, some sources of uncertainty relating to water quality can be listed as follows:

- Structure and parameters of simulation models,
- Uncertainty related to mis-measurement,
- Temporal and spatial incompleteness of data sets,
- Construction problems such as related to treatment facilities,
- Future and changing conditions (e.g. new socio-economic activities in the watershed and global warming)

- Vagueness of objectives (e.g. “improvement” of water quality)
- Situations of accidents and disaster (e.g. traffic accidents or chemical leakage from factories)
- Natural randomness of climate and hydrological variability (e.g. extreme events such as floods or long drought periods)
- Approximations, discretization and interpolations of variables depending on spatial and temporal scales.

To illustrate the importance of this topic, some of the various ways of how uncertainty analysis can improve water quality management and modeling are presented as follows:
- Uncertainty analysis can improve decision making when an organized and structured decision framework is considered, where decision makers and environment modelers work together. Some of the advantages of considering uncertainty analysis are an increase of net benefits, a reduction of treatment costs and more public acceptance of a certain decision.
- Uncertainty prediction should play an important role in the model selection process. In the decision making process based on model results and prediction, a decision taken without the consideration of uncertainty analysis is incomplete and may even be irresponsible (Reckhow, 1994a, b)
- Proper identification of the system uncertainties can support a more flexible and effective management framework, whereas components identified with greater uncertainty are considered for further investigation, the ones with less may become more important decision tools as the consequences can be more accurately predicted.

Results after uncertainty analysis can be given in different ways, such as probability density function and confidence intervals, giving a range of possible outcomes avoiding the common “unexpected” results. There are a variety of methods to handle uncertainty analysis. Some of these methods are first-order variance estimation, point estimate (PE) techniques, analytical techniques of integral transformations, Monte Carlo simulation, Bayesian technique and Fuzzy set theory. Yen (2002) presents some explanation and a list of reference for the first four methods. For Bayesian theory and Fuzzy theory, refer to Bernardo (1994) and Kaufmann (1975), respectively.

In this research uncertainty is handled by the use of Fuzzy theory in the development of the water quality model presented in Chapter 5, in an attempt to overcome the uncertainties related to data incompleteness and possible mis-measurements, and in the identification of conditional probabilities of inflow which are used in the stochastic optimization model of Chapter 6 and 7.

Moreover, quantity and quality objectives are also structured as fuzzy evaluation functions which are able to represent more concretely the degree of satisfaction in fulfilling a certain objective.

2.5 Management of Storage Water Quality

Quantity volumes have been considered since the beginning of storage reservoir construction, but quality has not. In the same way as for quantity operation, water can be stored and released for the improvement of water quality and the net benefits. As described in the previous sections, the assessment, evaluation and modeling of water quality are very complex tasks for water resources managers. Some of the reasons why water quality is still being left out of the decision-making framework process are:

- multiple stakeholders and multiple objectives are presented in most cases,
- there are a large number of physical, chemical and biological water quality parameters to be considered,
- quality conditions present great spatial and temporal variability,
- it is difficult to evaluate water and environment quality economically,
- quality may be highly influenced by human activities and natural conditions,
- complete and representative spatial and temporal water quality data are seldom available, and
- water quality simulation is a complex task due to the high levels of uncertainty involved in this process.

Moreover, information about costs is frequently needed to improve the decision prior to the water quality impact assessment. Such costs may be related to water treatment, penalties for excessive pollution, socioeconomic losses such as the ones related to tourism activities and decrease of benefits of certain objectives when multi-purpose operation is considered.

The operation of single and multiple outlet reservoir systems can present different objectives such as the improvement of water quality in stored volumes, released discharges and intake water. Moreover, the management horizon can be specified to deal with different problems such as those related to long-term planning, real-time operation and emergency situations. Besides the improvement of water quality condition itself, there are many advantages to considering water quality in the operation process, such as the wider acceptance of reservoirs among the general public, a decrease in maintenance costs, an increase in net benefits, preservation of aquatic life and the achievement of sustainable policies.

Management of water quality can be considered together with the various types of reservoirs which may present different forms of water usage, such as reservoirs for water supply, irrigation, hydropower generation, flood control, recreation and, of course, multipurpose systems. As explained previously, for each type of reservoir, a different evaluation method may be necessary.

Straskraba and Tundisi (1999) present some models that combine quality simulation with management tools. Research in this field is in constant development and some of the new models are imbedded with complex components that can facilitate the water quality analysis, such as remote sensing information, geographic information system (GIS), management and optimization techniques, artificial intelligence (AI) and decision support systems (DSS).

Even though there are a great number of water quality management models, few of them consider the optimization problem itself. This is due to many reasons such as computational efforts when the simulation and optimization models are combined and the lack of proper methodologies to incorporate the uncertainty analysis and multiobjectives characteristics of the system. For the practical optimization of storage reservoir the most common used technique is dynamic programming, as it is very efficient in dealing with stage-wise and non-linear systems. Some application of dynamic programming to water quality optimization can be found in only a few works, such as Fontane *et al.* (1981) for temperature with selective withdrawn, Ikebuji and Kojiri (1992) for downstream turbidity, Dandy *et al.* (1992) and Nandalal *et al.* (1995) considering downstream and reservoir salinity, Chang *et al.* (1996) considering inflow quality and Chaves *et al.* (2003a, b) for multiple water quality parameters for storage optimization.

2.6 Conclusions

It seems that there is still much to be accomplished and developed in the field of water quality in storage. The focus should be directly applied to integrated quality management, which must consider quantity and quality along with the three main components of a storage reservoir system: inflows, storages and releases. Moreover, other topics that should be considered are: i) simultaneous combination of stochastic optimization and water quality simulation models ii) uncertainty analysis and iii) multiple water quality parameters in a multiobjective optimization framework and multipurpose reservoirs considering the different stakeholders and water uses. The potential success in solving the water quality related problems will depend on the ability of different professionals including researchers, hydrologists, limnologists and decision-makers working together in an interdisciplinary and holistic way.

Some of the above topics discussed in this chapter are handled throughout this research. As shown in the following chapters, the quality simulation model is simultaneously combined with a stochastic optimization model under the same operation framework, considering the quantity and quality objectives and focusing on the improvement of water quality in storage. For the proposed monthly operation, storage and release quality may be considered with similar water quality; hence, improvements in storage quality are expected to represent direct improvements in release quality. Nevertheless, this assumption may not be at all realistic for the case of real-time operation, as the daily variation of quality and prediction components should be also taken into consideration. Even though the quality indices considered are composed by five quality parameters (DO, BOD, TN, TP and CHA), the addition of other parameters and indices could be easily implemented as shown in Chapter 7.

Although the improvement of water quality just through appropriate hydraulic control may be limited to physical constraints, such as quality of inflow, the combination of proper operation and other techniques may synergistically yield efficient results. Further improvements desired in the water may require other approaches or methods. Still all quality parameters (and consequently their indices), which depend on hydrological and climatological conditions, may have concrete chances for improvement by the appropriate operation of reservoir water volumes.

3 Storage Reservoir Operation

In this chapter, the main topics related to the operation and optimization of storage reservoirs are presented as background information for the following chapters where the new techniques for reservoir quality simulation and stochastic optimization are proposed. Firstly, we discuss the operation problem when the quality of water is of concern, followed by fuzzy control formulation, which is used for the optimization of the reservoir in chapters 6 and 7. Secondly, some of the important characteristics of dynamic programming are reviewed presenting the limitations regarding its simultaneous combination with simulation models. Finally, in the last section of this chapter, the genetic algorithm (GA) technique is presented showing its potential used and limitations for the reservoir optimization problem and the role of this technique in this research, followed by a description of the basic components of the developed GA model application for the training of the proposed simulation (Chapter 5) and optimization models (Chapters 6 and 7).

3.1 Operation for Water Quality

The Quantity and quality of water are often related. Some examples of how the management of quantity may interact with quality conditions are presented below in Figure 3 to Figure 5. The first case considers a reservoir with two outlets, as illustrated in Figure 3. As water in a reservoir may present different quality characteristics in the vertical water column, the effective operation of outlets could improve water quality conditions either in the reservoir itself, the released water, or both, depending on the operation objectives. Definitely in many cases, a tradeoff between the improvement of reservoir and release quality should be expected.

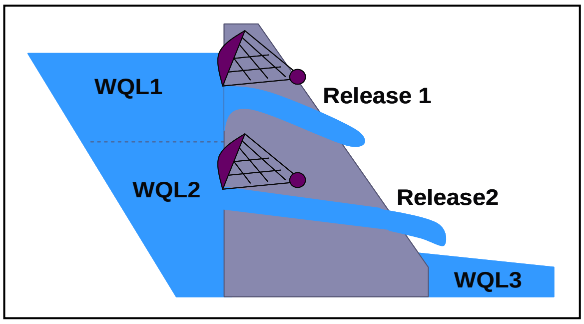


Figure 3. Top and bottom release

Considering that reservoir operation has a great influence on its volume and its release quality, it is easy to imagine this in a system of reservoirs, such as in Figure 4. The volumes and releases of each reservoir have important effects on each other when quality is considered as operation objectives. For example, in an attempt to improve quality condition of a downstream reservoir by proper operation of the upper one, the water quality of the latter may be deteriorated. This is also another example of the tradeoff between quality objectives for different components of the same system. Both examples indicate the necessity of an integrated analysis, presenting a high degree of complexity.

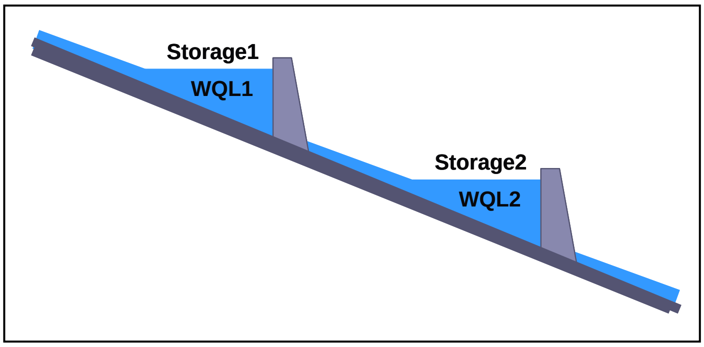


Figure 4: System of reservoirs

Each situation outlined may be found in our case study, which is described in more detail in Chapter 4. For simplicities sake and based on the proposed methodology and some of the actual characteristics of the case study in question, some assumptions can be made, fairly decreasing its complexity. Therefore, this research focuses on a simpler case, shown in Figure 5, which considers the relationship between storage volume, release from a single outlet (used for hydropower generation) and quality within the reservoir, which is basically the real situation encounter in the reservoir system of the proposed case study.

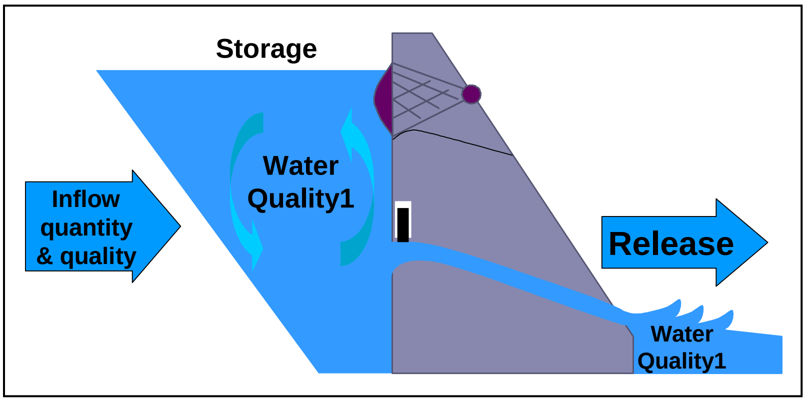


Figure 5: Storage and water quality

In the proposed case study, the spillway is basically only used during very high water periods. As the analysis focuses on the monthly time-step, it is possible to assume that the quality of water in release is similar to the one in the reservoir water body, which does not present long-term stratification due to its shallow depths and the climatological conditions of the region.

Therefore, the reservoir water body can be considered to be a completed mixed system, which may then avoid differences between the quality of upper and lower layers and the released water. Hence, the following assumption may be made: improvements on the storage water quality are expected to result in better released quality. This can greatly reduce the complexity of the system analysis. However, for short-term and real time operation these assumptions may no longer be valid and appropriate considerations should then be made.

3.2 Multi-objective Analysis

In the operation of multi-purpose reservoirs, quantitative and qualitative objectives must be considered. Hence, the calculation becomes complicated and difficult to execute (Simonovic, 1991; Yakowitz, 1998). The need to consider reservoir as multi-purpose systems has become even more important in recent years, due to the increased emphasis on environmental protection, especially in water bodies suffering from pollution from nearby urban areas. Attempting to evaluate objectives only economically, through the cost/benefits analysis, may fail to represent the satisfaction among all stakeholders. Moreover, it is almost impossible to only evaluate some objectives economically such as environmental quality, recreation and health problems.

Various approaches for solving the multi-objective management of natural resources have been performed, such as by El-Swaify and Yakowitz (1998). Traditional multi-objective optimization techniques applied to reservoir operation includes: the weighting method, the ε-constraint method, the surrogate worth trade-off method, goal programming, and compromise programming (Changchit and Terrell 1989; Simonovic 1991).

The application of the fuzzy sets theory can provide a viable way to handle situations when problems with objectives are difficult to be defined due to vagueness and imprecision. The application of fuzzy sets theory to water resources analysis can be found in Hipel (1982), Kojiri (1992), Russel and Campbell (1996), Shrestha (1996) and Fontane *et al*. (1997). Fuzzy set theory is applied in this research as it provides the ability to work with objectives regarding the degree of satisfaction to achieve them by the use of fuzzy membership functions. In the case of reservoir operation, fuzzy membership functions may be described in terms of water level, release and concentration of water quality parameters.

The main objective of this work is not to detail the formulation process of the membership functions. However, it is important to mention, the more realistic the formulation of these functions, the higher the chance of operation success. The functions have to represent the desire of all groups related to the water resource system in question and have to be constantly evaluated and updated to properly represent operation objectives such as described in section 2.2, in the case of quality objectives.

3.3 Operation with Fuzzy Objectives

Objectives of reservoir operation are commonly represented as a function of quantifiable characteristics such as release, storage and power generation. However, other objectives such as environment quality may not be so simple to evaluate, due to difficulties in the reduction to comparable commensurable units. One attempt to solve this problem is through the comparison of actual values with expected objective targets. These comparisons can generate functions that represent the degree of satisfaction between operation decisions and desirable values. Most of these satisfaction functions are based on the relationship between targets and expected values – for example, goal programming. Nevertheless, some objectives cannot be easily perceived and understood by water stakeholders.

With the intention to make the interpretation of those degrees of satisfaction easier (especially because of their complexity and vagueness) linguistic descriptors such as “good” water quality or “adequate” water supply may be used. The degree of satisfaction will depend on stakeholders’ perceptions and experience, the system’s characteristics considering the constraints encountered, knowledge about the state of the system, risk perception and impact related to water quantity and quality problems.

The basic method of formulating the membership functions is provided by survey with water manager and users who have knowledge about the subjectivity of the system. Elaboration and analysis of these surveys (Norwich and Turksen (1984), and Rea and Parker (1992)), as well as the fitting of curves on the acquired data (Dombi (1990) and Klir and Yuan (1995)) can be located in various references.

This method of categorizing reservoir objectives through the use of linguistic values can be handled with fuzzy sets, allowing a gradual transition from a situation that completely fulfills a concept to a situation that does not. A more detailed explanation on the theory of fuzzy sets is presented in Kaufmann (1975) and Kacprzyk (1997). More detail on the fuzzy objectives adopted for our optimization problems is presented in Chapters 6 and 7.

3.4 Multistage Fuzzy Control and Optimization

In fuzzy logic based models, variables are imprecise or vague, and the source of uncertainty is not merely due to the random quality of the natural event. The fuzzy optimization approach can address the problem of subjective and noncommensurable objectives in a way that is easily interpreted. It indicates relatively how each objective has been satisfied. Fuzzification allows decision makers to specify the goals and/or constraints in subjective and linguistic terms.

The multi-objective operation framework presented hereinafter is based on fuzzy control theory, particularly because operation objectives are presented with different commensurable units. Moreover, a high degree of uncertainty and fuzziness in the definition of such objectives is also present. Usually in the theory of fuzzy control, the fuzzy decision is obtained after the consideration of fuzzy goals $\mu_G(x)$ and fuzzy constrains $\mu_C(x)$, as presented by Kacprzyk (1997) in his work on multistage fuzzy control. However, here it is decided to use the term fuzzy objective $\mu(x)$ (performance or evaluation functions), to mean both, the fuzzy goals and the fuzzy constraints. This is done for the sake of simplicity, especially because the definition of fuzzy goals and constraints differ very little from each other, and when they present equal importance, the mathematical difference between them becomes even more irrelevant. However, just as a formality, the general formulation of the decision making problem in a fuzzy environment suggests that the “goal should be attained and the constraint satisfied.”

Kacprzyk (1997) describes various kinds of fuzzy control formulations applied to deterministic, stochastic and fuzzy systems with different optimization horizons, such as with fixed, implicit fixed and infinite termination time. He also proposes different solution methods to solve these optimization problems, such as dynamic programming (DP), stochastic DP (SDP), branch-and-bound, genetic algorithms (GA) and artificial neural networks (ANNs), which may be applied with different aggregation operators (related to the performance functions for a single stage or time-step) and different fuzzy types of decisions (related to the recursive equation for the optimization horizon).

The proposed optimization model presented in Chapter 6 is tested under different aggregation methods and fuzzy decision types, to be able to test its applicability and efficiency under various calculation schemes. However, it may be argued that the aggregation method and fuzzy type of decision to be used may depends significantly on the judgment of operators.

3.4.1 Aggregation Operator

A certain decision in stage *t* can be evaluated through the fuzzy (objective) membership function related to a fuzzy objective. The resulted evaluation from each performance function can be then combined to grasp the overall performance of the control decision. Various types of aggregation operators which can be used to combine or aggregate the multiple objectives can be formulated, such as minimization, product, weighted-sum, maximization and algebraic-sum.

The general mathematical formulation for three methods of the aggregation operator, minimization, product and weighted-sum is displayed in equations (2) to (4). Furthermore, some of these operators are tested with the optimization model proposed in Chapter 6. In general terms, the minimization and product types tend to assume the same importance for all objectives (Kacprzyk, 1997). The general representation of the performance function *E* is presented in equation (1), where *n* refers to the fuzzy objectives.

$$E = f(\mu_1(x),...,\mu_n(x),...,\mu_N(x)) \qquad (1)$$

- *Minimization*

$$E = \min_n [\mu_n(x)] \qquad (2)$$

- *Product*

$$E = \prod_{n=1}^{N} \mu_n(x) \qquad (3)$$

- *Weighted-sum*

$$E = \sum_{n=1}^{N} \alpha_n \cdot \mu_n(x) \quad \text{where} \quad \sum_{n=1}^{N} \alpha_n = 1 \qquad (4)$$

3.4.2 Fuzzy Type of Decision

As mentioned before, the aggregation operators are responsible for the combination of the results from the fuzzy objective membership functions referent to a certain time-step or stage *t*. To cope with the multistage operation problem, it is then necessary to present the operation formulations considering the total optimization horizon. This can be achieved through the recursive equation or objective function. The general formulation of the performance function *F* is shown in equation (5), and basic formulations for each type of decision are present in equations (6) to (8), where *T* is the total number of stages, and the *max* operator means the maximization of the benefit values, which are here represented by the evaluation functions *E*.

$$F = \max_{1 \le t \le T} \left[f(E_1 ... E_T) \right] \qquad (5)$$

- *Minimization Type*

$$F = \max_{1 \le t \le T} \left[\min[E_t] \right] \qquad (6)$$

- *Product Type*

$$F = \max_{1 \le t \le T} \left[\prod_{t=1}^{T} E_t \right] \tag{7}$$

- *Weighted-sum Type*

$$F = \max_{1 \le t \le T} \left[\sum_{t=1}^{T} \beta_t \cdot E_t \right] \quad \text{where} \quad \sum_{t=1}^{T} \beta_t = 1 \tag{8}$$

Where β is the weight, here we consider the weights to be the same and equal to *1/T*. Nevertheless, different weights may be incorporated to increase the importance of a particular season in time or as a discount factor. Figure 6 summarizes the influence of each fuzzy type of decision regarding the decision-making process, in a schematic scale measuring the degree of conservatism.

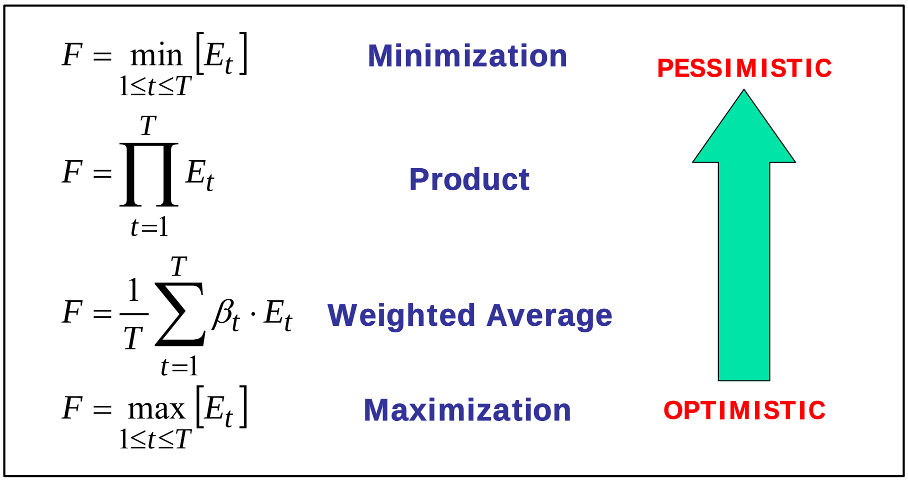


Figure 6. Recursive equation or fuzzy type decision and its degree of conservatism

No doubt the most important operator in fuzzy control is the minimization method. One of the disadvantages of this decision type is the inability to reflect any correlation among intermediate evaluation values, as it always chooses the smallest among different evaluation values, which in some operation problems may not be sufficient for proper achievement of operation objectives. For example, appropriate operation of reservoir is argued here to be capable of increasing (improving) the values of quality indices. However, due to physical constraints, such as inflow quality and storage volumes, this improvement may be limited only to some extent. Hence, during certain times of the year, some of the quality parameters or quantity objectives may be impossible to be satisfied at all, independently of any effort of operators to do so, resulting in a null (or zero) evaluation value. This would mean that any operation policy carried out for the whole operation horizon would be the same. However, we know that design considering a zero-fail system can be extremely inefficient and, therefore, any system is subjected to a risk of failure, even if minimal.

As already mentioned, the above optimization problems can be solved by a diversity of methods. However, the classical method used for the case of reservoir operation is definitely the one based on dynamic programming (DP). Even though DP presents many advantages over other methods, its application may not be feasible in certain cases, such as when the number of state variables increases too much resulting in the problem called the curse of dimensionality (Bellman, 1959) or when the stochastic DP requires the forward direction calculation scheme (Chaves, 2002 and Chaves *et al*, 2003a). In the next section, application and limitations of DP for the case of reservoir operation is presented. Finally, a brief background on the use of artificial intelligence (AI) techniques in the reservoir operation is given. In Chapter 6, a new methodology based on ANN, to solve the above mentioned optimization problems, is proposed and compared to some DP schemes.

3.5 Reservoir Optimization through Dynamic Programming

Dynamic programming (DP) was developed in the 1950's by RAND Corporation and was sponsored by the US Air Force. It was named and described by Richard Bellman (1959). Applications of dynamic programming to water resources systems can be found in many works, such as Yakowitz (1982) and Esogbue (1989). DP presents various advantages over other methods to approach water resources management aspects, and can be associated with other techniques named such as stochastic DP (SDP) and fuzzy DP (FDP). Some applications of SDP for water resources management is found in Torabi *et al.* (1973), Kelman *et al.* (1990), Fontane *et al.* (1997) and Chaves *et al.* (2003a, b), where river discharge is treated as a stochastic variable. In fuzzy DP, decision and state variables as well as constraints, can be represented as fuzzy membership functions, some application of fuzzy DP for the operation of storage reservoirs can be found in Kojiri *et al*. (1992), Kojiri *et al*. (1993), Hori *et al.* (1997) and Fontane *et al*. (1997). The basic characteristics of water resources and reservoir operation that lead to the use of DP are stage-wise structure and non-linearity of the system. DP can also be divided into two different approaches depending on the problem, continuous or discrete – with the latter the most commonly used for the sake of computational simplicity. One important characteristic of DP is the calculation direction in which the model is developed, known as forward-moving and backward-moving. The advantages and disadvantages of DP can be summarized from Labadie (1993) as follows:

Advantages:

i) It is suitable for solving sequential decision problems. Reservoir operation can be easily modeled in a stage-sequence problem, facilitating the use of DP, with storage as the state variable and release (or end-of-period storage) as the decision variable.

ii) It is quite simple to handle nonlinear modeling. This characteristic can be very important when modeling functions based on variables such as water quality and hydropower.

iii) Efficiency increases with an increasing number of constraints because possible calculation iterations decrease.

Disadvantages:

i) The biggest disadvantage is definitely the curse of dimensionality, where the required computational effort increases linearly with the number of stages and exponentially with the number of state and decision variables considered. The maximum possible number of variables able to be handled with DP is usually only four to seven.

ii) Other drawbacks may be the loss in efficiency originating from the discretization of state and decision variables and possible problems relating to the calculation direction.

Dynamic Programming converts a large, complicated optimization problem into a series of smaller interconnected ones, each containing only a few variables. The result is a series of partial optimizations requiring a reduced effort to find the optimum. The DP algorithm can be applied to find the optimum of the entire process by using the connected partial optimizations of smaller problems. In each process, the functional equation governing the process was obtained by an application of the following intuitive principle stated by Bellman:

> *"Principle of Optimality. An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."*

The principle of optimality was stated mathematically as the dynamic programming algorithm to maximize a serial process with T stages. A unit of time can be represented as a stage, like weeks or months. Every stage t has a variety of variables and functions. The return or evaluation function (E) gives the measure of profit or cost for the stage. Decision variables, for example release, are those that can be manipulated independently. State variables, like storage and water quality parameters, are inputs to a stage from an adjacent stage. Consequently, such variables cannot be manipulated independently. In the case of reservoir optimization, the transition function for the storage state variable is based on the continuity equation. The return function depends on the decision and state variables. In order to determine the optimal value for the return function for each stage, it is necessary to exhaustively list individual values of the state variables, searching for the correspondent decision variables. The objective function (*F*) is based on the maximization (or minimum) of the aggregation for the return functions for each stage. The accumulation of the objective function between each stage can guarantee the continuity of optimization. The basic deterministic DP formulation for reservoir quantity optimization is shown in equation (9), below:

$$F_t(S_t) = \max_{S_{t+1}} \left\{ E_t(\cdot) + F_{t+1}(S_{t+1}) \right\} \qquad (9)$$

Where, *F* is the recursive objective function, *E* is the evaluation function, *S* is the storage at the beginning of period or stage *t* and *R* is the release from reservoir in stage *t*.

3.5.1 Deterministic, Probabilistic and Stochastic DP

Basically, dynamic programming models can be classified as discrete or continuous, deterministic or stochastic and serial or non-serial (Bellman, 1959; Esogbue, 1989). Discrete is defined here as a process consisting of a finite number of stages. Discrete DP models tend to result in considerable saving of computational time.

In deterministic models there is one value for each situation at each unit of time. The outcome of a decision is uniquely determined by the decision. The calculations are based on physical, chemical, biological and economical formulations. Basically, DP can consider probability characteristics of variables in two different ways, as probabilistic models and as stochastic models. In both models, a range of values can be expected for each situation at each unit of time. They may be applied when the probabilities of the occurrences are known (risk). The probabilistic model does not consider conditional probability. There is no dependency of time regarding the probability characteristics of variables. On the other hand, the variables of stochastic models will have time-dependent probabilities (conditional probability). The SDP can be assumed as the most efficient method as unknown (but probabilistic) future characteristics of the system can be handled, increasing the robustness of the optimization model. As the objective of this research is to develop a monthly planning operation, if it is necessary to replace the SDP, the proposed technique should also be embedded with a component able to handle such stochastic characteristics of the system. Therefore, in Chapter 5 the stochastic optimization is compared with SDP as the latter can be considered an efficient technique for problems with few state variables. The basic stochastic DP formulation for reservoir quantity optimization is shown in equation (9), below:

$$F_t(S_t, I_{t-1}) = \max_{S_{t+1}} \left\{ \sum_{k=1}^{K} P\left(I_{t,k} / I_{t-1}\right) \cdot \left[E_t(\cdot) + F_{t+1}(S_{t+1}, I_t)\right] \right\} \qquad (10)$$

Where, F is the recursive objective function, E is the evaluation function, S is the storage at the beginning of period or stage t and R is the release from reservoir in stage t. $I_{t,k}$ represents the discrete inflow during period t, at discrete probability level k and $P(I_{t,k}/I_{t-1})$ is the probability of occurrence of $I_{t,k}$ conditioned to the previous known monthly inflow I_{t-1}, note that $\Sigma_k P = 1$.

3.5.2 Inverted and Noninverted form of the Mass (Continuity) Equation

The transition function used in the case of reservoir operation is normally the mass balance equation, or simply, the continuity equation. There are two ways the reservoir mass balance can be considered (Labadie, 1993). Ignoring other losses such as evaporation and infiltration, they can be defined for the noninverted form as in (11) and for the inverted form as in (12):

$$S_{t+1} = S_t - R_t + I_t \qquad (11)$$

$$R_t = S_t - S_{t+1} + I_t \qquad (12)$$

Where S, R and I are respectively storage, release and inflow for time t.

Normally, the noninverted form is found in most of the studies related to DP applied to water resources optimization. The noninverted form would be a more natural way to approach the optimization problem, considering that the goal is optimization of release with direct influence to storage. In the noninverted form, end-of-period storage is treated as the variable to define the grid to be searched, and release is directly optimized, resulting in optimal release policies $R_t(S_t, I_{t-1})$. A disadvantage in the use of this form is that the continuity equation is an implicit function of the end-of-period storage (S_{t+1}), which results in the need for several iterations to find an explicit solution for the end-of-period storage. Suharyanto and Goulter (2004) presented an interesting way of dealing with this form in the case of SDP using fuzzy theory.

On the other hand, in the inverted form, releases are regarded as random variables, as optimization is performed directly over the end-of-period storage S_{t+1}. Storage can then be directly calculated with no need for iterative procedures, resulting in optimal storage guidecurves $S_{t+1}(S_t, I_{t-1})$. Moreover, this form is considered to be more flexible for reservoir operation, as the storage discrete grid and optimization continuity can be elaborated on more easily, allowing direct trace-back of possible optimal storage levels at each stage.

However, with this form, there may be a risk of failure to satisfy constraints on release. Setting penalty terms for the objective function or specifying a maximum risk of failure to satisfy downstream requirements could overcome this problem. Even though it may look odd, it is important to note that release is still being the decision variable, because it is used in the evaluation function, which drives the decision in the optimization process. The inverted form is used hereafter as it may decrease the complexity of the SDP model, especially when considering the optimization model proposed in Chapters 6 and 7.

3.5.3 Backward and Forward Calculation

The backward-moving is a common calculation procedure for DP, where calculation starts from the stage in the end of the optimization horizon (right side), moving backwards until the first time stage. In other words, the calculation sequence proceeds in reverse chronological direction. Calculation continuity is guaranteed based on the beginning-of-period storage. Therefore, there is only one optimal value for each beginning-of-period storage value. Optimum sequence is defined in the chronological direction (from left to right), as can be seen in Figure 7, where $S_{t,i}$ represents discrete storage values for calculation. F and E represent objective and evaluation functions, respectively. Enumeration of stages can be done either way. Hereafter, it is decided to enumerate stages considering chronological time sequence for both forward- and backward-moving, so that the stages are enumerate starting from 1 to total number of stage or time-steps T. Equation (13) and (14) show the deterministic and stochastic DP formulation, respectively, for the backward-moving calculation.

Backward Deterministic DP:

$$F_t(S_t) = \max_{S_{t+1}} \left\{ E_t(\cdot) + F_{t+1}(S_{t+1}) \right\} \tag{13}$$

Backward Stochastic DP:

$$F_t(S_t, I_{t-1}) = \max_{S_{t+1}} \left\{ \sum_{k=1}^{K} P\left(I_{t,k} / I_{t-1}\right) \cdot \left[E_t(\cdot) + F_{t+1}(S_{t+1}, I_t) \right] \right\} \tag{14}$$

Where S_t represents the initial storage for stage t and S_{t+1} is the end-of-period storage for time-step t. Therefore, S_{t+1} is also the initial storage for the next stage.

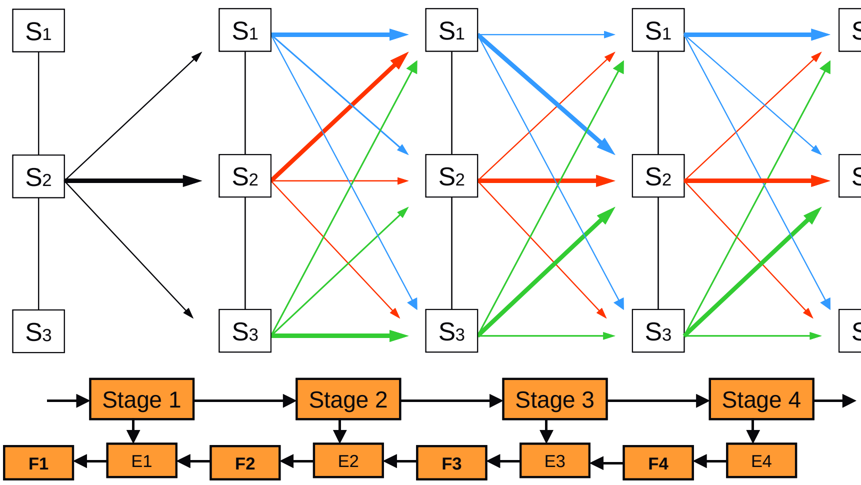


Figure 7. Backward DP (calculation: right to left; optimum sequence: left to right).

The other way to execute calculation is called forward-moving, as shown in Figure 8. Basically, it works in the opposite way to the backward-moving. Optimization continuity is ensured based on the end-of-period storage. Hence, resulting in stage optimal sequences associated to end-of-period storage.

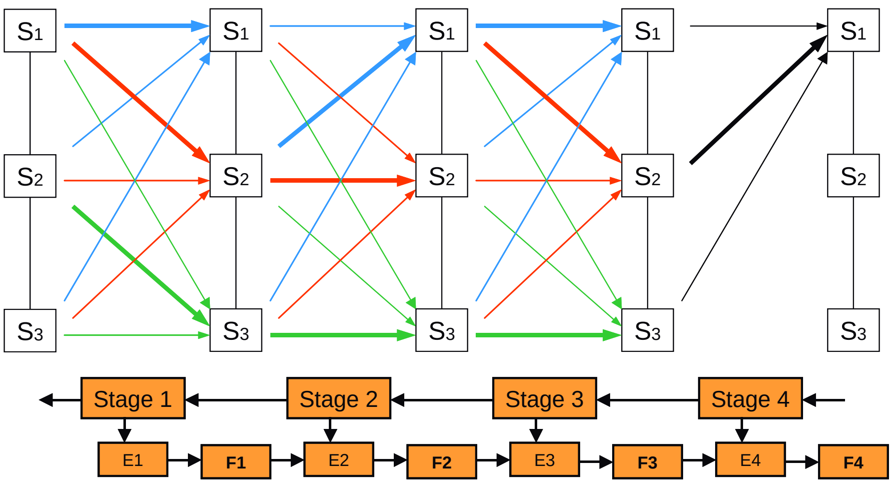


Figure 8. Forward DP (calculation: left to right; optimum sequence: right to left)

However, in the stochastic case, it can be a problem as optimal guidecurves can only be defined for a restricted range of initial storage and it may present multiple "optimal" values for an initial storage, which is not desired. Moreover, also for the SDP, the continuity of the recursive equation is compromised as the next stage to be calculated is dependent on two-previous time-steps regarding the stochastic variable, as can be seen in equation (16) in the term $I_{t-2,k}$. Equation (15) and (16) show the deterministic and stochastic DP formulation, respectively, for the forward-moving calculation.

Forward Deterministic DP:

$$F_t(S_t) = \max_{S_{t-1}} \left\{ E_t(\cdot) + F_{t-1}(S_{t-1}) \right\} \tag{15}$$

Forward Stochastic DP:

$$F_t(S_t, I_{t-1}) = \max_{S_{t-1}} \left\{ \sum_{k=1}^{K} P\left(I_{t,k} / I_{t-1}\right) \cdot \left[E_t(\cdot) + F_{t-1}(S_{t-1}, I_{t-2,k}) \right] \right\} \tag{16}$$

Although it may seem that the choice of method (backward or forward) makes no difference in terms of results, the way the calculation is developed can be extremely important in some applications. This is particularly true when dealing with stochastic variables, or considering the combination of DP with other models, which may require a forward-moving calculation.

As explained in Chaves *et al.* (2003a, b), stochastic dynamic programming (SDP) is not suitable for the simultaneous combination with water quality simulation models. Basically, it is due to the different directions in which the two models must be calculated and the increase in the number of state variables (curse of dimensionality) when water quality variables are considered. Mathematically, the combination may be possible but the number of assumptions taken would make the problem too complicated and probably computational infeasible, especially when considering the water quality optimization. Therefore, in Chapter 6 and 7, a stochastic model based on neuro-fuzzy systems is proposed in an attempt to overcome these limitations of SDP models presented above.

3.6 Reservoir Operation through AI Techniques

The application of artificial intelligence (AI) techniques for the optimal operation of reservoirs, also called heuristic models or soft-computing, has become more popular in the last decades. Different from traditional optimization algorithms, these models cannot guarantee the optimal solution. Nevertheless, they are usually capable of achieving global optimal solutions to problems where traditional models would fail to converge due to dimensionality and local optima related problems (Labadie, 2004).

The most popular AI optimization technique is surely the genetic algorithm (GA), which is categorized as an evolutionary programming, in that their optimization process is analogous to the mechanism of natural selection and genetics of biological sciences. Genetic algorithm (GA) models have also been developed for water resources system optimization, such as in works by

Goldberg (1987) for pipeline optimization and Wardlaw and Sharif (1999) for reservoir operation. Some of the advantages of GA models are: (a) it is a powerful tool to deal with nonlinear system global optimization, (b) it can also be directly linked with hydrologic and water quality models without much simplification, (c) robust ability to solve highly nonlinear, nonconvex problems, specially when using floating number or real number coding, instead of traditional binary coded (Wardlaw and Sharif, 1999) and (d) it may become a very efficient technique for tackling the dimensionality problem when combined with DP in the optimization of multi-reservoir system, such as shown in Huang *et al.* (2002).

GA presents some disadvantages when compared to other techniques such as DP. Some of these disadvantages are: (a) increase in the number of variables proportionally to the size of the time series data set, as the variables of each time step becomes a state or decision variable, aggravating the convergence problem, (b) GA may be too problem-specific requiring modification for each new application making its practical use rather difficult, (c) difficulty of explicitly accounting for constraints (particularly, inequalities constraints) and maintaining feasible solutions in the population (Labadie, 2004) and (d) expensive computational efforts when combined with explicitly stochastic optimization models when applied to multi-reservoir systems. The latter disadvantage can be overcome if operating rules can be parameterized in some way, for example by combining GA with other techniques such as artificial neural networks (ANN). Chang and Chang (2001) combine genetic algorithm (GA) and an adaptive neural fuzzy inference system (ANFIS), such in a way that GA is applied to search the optimal reservoir operating histogram, which is used as the training pattern of the ANFIS model. The later is built to estimate the optimal water release based on the current storage level and inflow conditions, for a deterministic problem.

The difference of the proposed method is that, reservoir operating rules, which are represented by an neuro-fuzzy system, are stochastic trained (optimized) by a GA model, as proposed in Chapter 6 for the operation considering quantity objectives and as in Chapter 7 when quality also becomes a concern of operators. ANN models trained by GA are called evolving networks, different from other training techniques such as the traditional back-propagation. This evolving process is also applied for the training of the conceptual fuzzy neural network (CFNN), proposed in Chapter 5. Some basic characteristic of the GA model developed for the training of the neuro-fuzzy system presented later are described as follows.

3.6.1 Basic Characteristics of the Developed GA Model

Yao (1999) presents a full review of the evolution of ANN, as for its weights, structures, learning rules and input features. According to Yao (1999), the evolutionary training approach is attractive because it can handle global optimization in much more complex problems. Moreover, the same evolutionary algorithm (EA) can be used to train different types of networks, saving a lot of human effort in the development of different training algorithms for different types of ANN's. Also he presents an extensive list of references where evolutionary approaches have been successfully used to train recurrent, high-order and fuzzy ANN's.

Therefore, a genetic algorithm (GA) based model is developed here to deal with the optimization of the parameters of the fuzzy inference variables and connection weights of the proposed nero-fuzzy systems. The evolutionary approach for training the network has two major points to be considered: decision on how to represent the connection weights, i.e., binary or real numbers representation; and, definition of evolution operators such as crossover and mutation. Nevertheless, as the purpose of this research is not to deeply investigate the evolving ANN's itself, in this section just a brief explanation of the developed GA model is given. For further information on evolving ANN readers are recommended to other references such as Fogel (1995), Yao (1999) and Renner and Ekart (2003).

The GA optimization model is constructed with five operators - elitism, crossover BLX-α, multi-points crossover, simple mutation and creep mutation, applied for real number coding (Galvao, 1999) - instead of the more common binary formulation. Yao (1999) explains some of the advantages of evolving algorithms based on the real number formulation.

Calculation is carried out for the optimization of the total number of network parameters and connection weights to be trained, accounting each for the same number of existing connections. The total population is assumed to be around three times the number of genes (variables to be optimized). For the elitism operator, the two best chromosomes are saved out of the total chromosomes size population, generating the new offspring of the new population in the next generation.

In crossover BLX-α, a chromosome son $S1$ is generated as: $S1 = F1 + \beta (F1 - F2)$, where β is a random number between 0.5 and 1.5; $F1$ and $F2$ stand for the genes from two different father chromosomes from the total population through a triangular probability distribution function. If the generated chromosome does not fit in the specified range for the related parameter, the calculation is repeated using a new random value for β, until an acceptable value is found. This operator generates around one-quarter of the new chromosomes.

The triangular probability distribution function is applied for the selection of father chromosomes not only for the crossover BLX-α, but also for the other operators. The function tends to prioritize the chromosomes with higher aptitude (higher values for the performance function); hence, improving the efficiency of the evolutionary algorithm. A random value between 0 and 1 is generated and the chromosome with the closest cumulative probability is chosen, as shown in Figure 9, which shows a plot of the triangular probability for each chromosome and the cumulative probability distribution for a 20-chromosome population. The triangular probability distribution function is defined by equation (17):

$$P_i = P_{i-1} + \frac{2(m+1-i)}{m(m+1)} \qquad (17)$$

where $P_0 = 0$ and i correspond to the ith chromosome of an arranged population from best to worst aptitude, $i = 1,\ldots, m$, and m is the total number of chromosomes in the population.

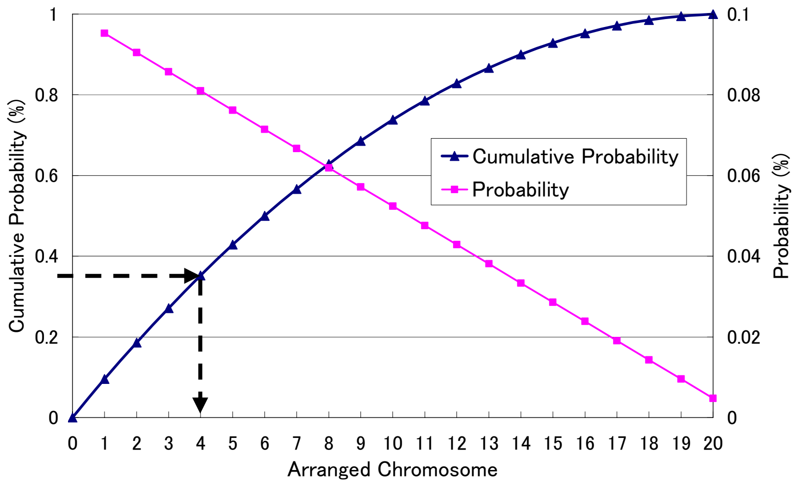


Figure 9. Triangular (cumulative) probability distribuition

In the multi-points crossover, a chromosome son *S1* is generated from two different chromosomes chosen from the whole population through the triangular probability distribution function (17). This operator also contributes to the convergence of the calibration model. This operator generates also around one-quarter of the new chromosomes of the new population. This is found to be the most relevant operator in the convergence of the optimization process.

The creep mutation operator generates a son *S1* from a chromosome father *F1* chosen by the triangular probability distribution function, as: $S1 = F1 \cdot k$, where k is a random number from 0.8

to 1.2. The Creep mutation intends to refine the results towards the optimum solution. This operator generates one-quarter of the new population.

Finally, simple mutation is applied to avoid problems related to premature convergence. This operator generates around a quarter of the new population. The search space is limited to the minimum and maximum values of the parameters and weights to be optimized. The objective function used for optimization is based on the performance or objective functions. Figure 10 shows the flowchart for the development of an ANN model with GA as the training algorithm.

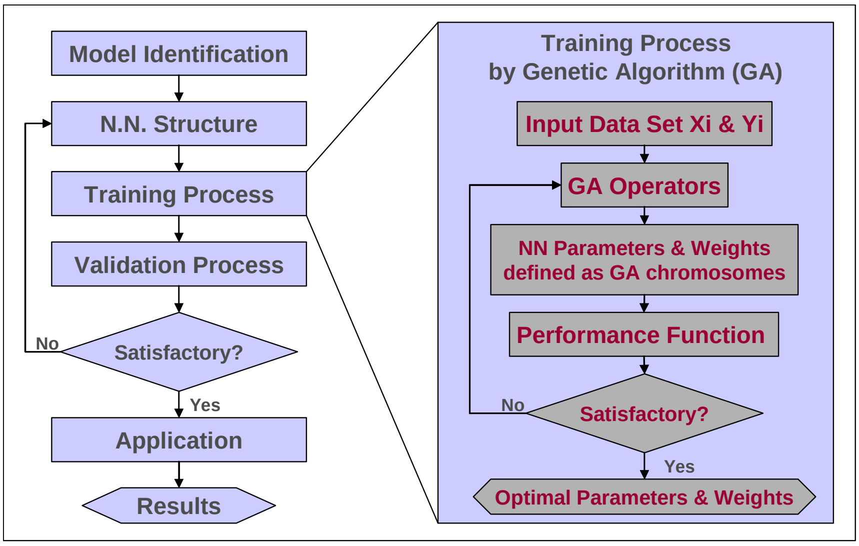


Figure 10. Flowchart of training process using GA

3.7 Conclusions

This chapter showed an overall idea of the present stage of basic reservoir operation and optimization techniques, together with their advantages and disadvantages. It also explained some of the background and concepts used hereafter for the development of this research.

In his recently extend review about the state-of-the-art techniques in the optimal operation of multi-reservoir system, Labadie (2004) concluded indicating some potential future developments in this field. The present work addresses some of his observations, which are described as follows:

(a) "Reservoir system operators may increasingly rely on sophisticated computer modeling tools to better respond to new environmental and ecological constraints",… "by the improvement of the linkage of optimization and simulation models, which they are more ready to accept."

(b) "Past difficulties in inferring operating policies from implicitly stochastic optimization models may be alleviate through applications of the fuzzy rule-base systems and neural networks. The computational challenges of explicitly stochastic optimization may also be overcome through judicious application of these heuristic techniques."

Labadie's remarks reinforce the importance, originality and state-of-the-art status of the research proposed here, in the field of optimal reservoir operation, where through the development of a new approach combining simultaneously an original stochastic neuro-fuzzy model, chapter 6, with a water quality simulation model, chapter 5, the multiobjective optimal reservoir operation was successfully achieved, after the optimal operation policies could be successfully derived .

4 Study Area

4.1 General information

4.1.1 Hydrological Situation in Brazil

Brazil has a vast and dense fluvial net. Many of its rivers are notable for their great width and depth, due to the nature of the relief with prevailing areas of plateau. Besides other characteristics, there is a great potential for hydropower generation. Among the great national rivers, only the Amazon and Paraguay rivers are predominantly and broadly used for navigation. The rivers São Francisco and Paraná are the main rivers located within the plateau area. Generally, the rivers do not have their origins in areas of very high elevation, with the exceptions being the Amazon and some of its tributaries that originate in the Andean mountain range.

In general, the Brazilian fluvial net can be divided into seven main basins, as illustrated in Figure 11: Amazon River Basin; Tocantins-Araguaia River Basin; South Atlantic Basin - north and northeast regions; São Francisco River Basin; South Atlantic Basin - east region; Platina Basin, composed of the sub-basins of the rivers Paraná and Uruguay; and Atlantic South Basin - southeast and south regions.

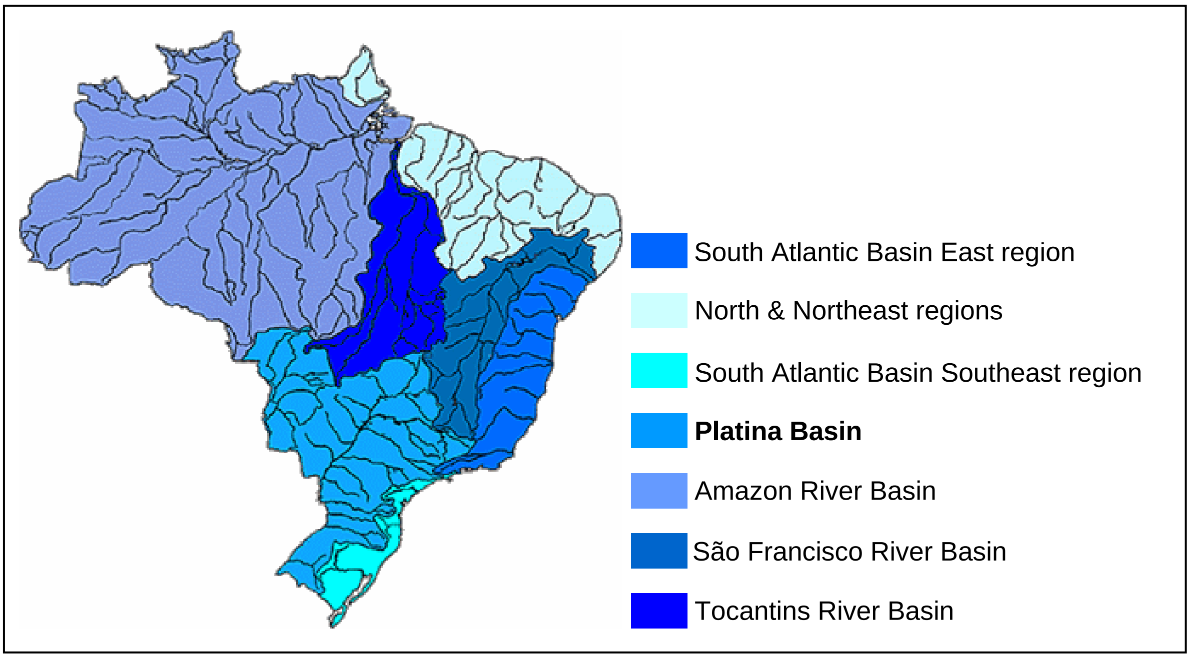


Figure 11. Main Brazilian watersheds

4.1.2 Platina Basin, or Plata River Basin.

The Platina basin, or Plata river basin, consists of the sub-basins of the rivers Paraná, Paraguay and Uruguay, draining water from areas of Brazil, Bolivia, Paraguay, Argentina and Uruguay. The Paraná River is approximately 4,900 km in length, being the second longest in South America. Parana River is formed by the junction of Grande and Paranaíba rivers. Its main tributaries are the rivers Paraguay, **Tietê**, Paranapanema and Iguaçu.

4.1.3 State of São Paulo

The Tietê river basin is located in the Southeast region of Brazil running entirely within the state of São Paulo. The state of São Paulo has a total area of 248,809 km² with a population of 35 million. The State has the greatest GDP as compared with other Brazilian states. São Paulo is the Brazilian state with the largest population, the largest industrial area and is by far the largest consumer center. Approximately 30% of all private investment in the country is centered in the State. It generates approximately 40% of national industrial production, while the metropolitan area of São Paulo itself generates roughly 20% of the total national production. It has the most extensive and best-structured highway network in the country, which has been privatized in recent years, as have many other formerly public sectors, such as telecommunication and energy production.

The State is the Brazilian leader in agribusiness, representing one third of the national agro-industrial GDP. Sugar cane crops and the agribusiness have a particular economic relevance for the region. It is the largest fruit producing state, with one third of the total. São Paulo grows 80% of the industrialized orange crop, 70% of the sugar cane crop, and a quarter of Brazil’s vegetables. It is also the third largest coffee producer in Brazil. Brazil’s first major commercial inland waterway, the Tietê-Paraná Waterway, and its area of influence correspond to more than 80% of the economic and social characteristics of the State.

It was only after the 1950’s that the city of São Paulo started to show great indices of economic growth. Its population increased at a surprising rate, basically due to the migration from other parts of Brazil, mainly from the rural areas. In order to achieve these extraordinary rates of growth, São Paulo mismanaged some of its social and natural resources. This has resulted in many problems, such as deforestation, air and water pollution and social inequalities. The origins of the basic problems relating to the river’s water are the untreated domestic, industrial and agricultural runoffs, which in most cases flow directly into the rivers, especially the Tietê River. As a result, Tietê is considered to be one of the most polluted rivers in the country.

4.1.4 Tietê River Basin

São Paulo is one of the few states in Brazil, which has a formal division for its hydrographic basins. The sub-basins of Tietê River is identified by the numbers 5, 6, 10, 13, 16, 18 and 19 as shown in Figure 12.

Sub-basins of Tiete River:

5 Piracicaba/Capivari/Jundiaí (PCJ)

6 Alto Tietê (AT)

10 Sorocaba e Médio Tietê (SMT)

13 Tietê/Jacaré (TJ)

16 Tietê/Batalha (TB)

18 São José dos Dourados (SJD)

19 Baixo Tietê (BT)

Figure 12. Sub-basins within the state of São Paulo (source SIGRH - Sao Paulo State)

The basic water uses in the Tietê River basin are water supply for domestic and industrial use, waste water effluent dilution, also for domestic and industrial waste, irrigation, navigation, hydropower generation, tourism and, within the reservoirs, fishery production. Some basic characteristics of the sub-basins of the Tietê River are listed in Table 1. Illustrative maps of the sub-basins 5, 6, 10 and 13, with their main rivers, land use and observation stations are presented in Appendix IV.

Table 1. Basic characteristic of Tietê river sub-basins

Characteristic	Unit	5	6	10	13	16	19	Total
Area	km^2	15205	5650	12099	11537	13394	15347	**73232**
Population								
Urban	1000 hab	2914	15772	1119	1031	301	525	**21662**
Rural	1000 hab	190	347	182	94	98	73	**984**
Total	1000 hab	3104	16119	1301	1125	399	598	**22646**
Water Demand								
Urban	m^3/s	13.3	62.3	3.7	3.1	0.9	1.6	**84.9**
Industry	m^3/s	12.5	4.6	6.8	8	1.4	1.8	**35.1**
Irrigation	m^3/s	7.6	4.2	10.9	3.3	6.5	9.8	**42.3**
Total	m^3/s	33.4	71.1	21.4	14.4	8.8	13.2	**162.3**
Groundwater	m^3/s	5 - 60	5 - 40	5 - 200	5 - 200	5 - 50	8 - 50	

Source: SIGRH - Sao Paulo State

Some hazards related to flooding are found within the basin, especially around São Paulo City. Basically, it is caused by mismanaged urbanization resulting in the increase of runoff and, consequently, the increase of flood peak discharge. The irregular occupation of the river bed by low income communities during drought periods, trying to get a space near to the formal urbanized areas, increases the problems related to flooding during intense rainy periods, and has further degraded the river water quality. Moreover, other issues, such as water scarcity, pose serious problems for local residences. This scarcity could be caused not only by the lack of quantity of water, due to population increase or overexploitation of aquifers, but especially by the lack of minimum water quality conditions, which is a result of the excessive pollution and the lack of regard for environmental protection.

4.2 Barra Bonita Reservoir

4.2.1 Basic Characteristics

Barra Bonita reservoir is located in the middle Tietê River basin, São Paulo, Brazil, (22°29' S and 48°34' W) with maximum surface water at an altitude of 453 m. The reservoir has a water surface area of approximately 340 km^2, total volume of 3.6 km^3 and length of 50 km. Maximum and average depth are around 25 and 10 meters, respectively. Average water fluctuation in the reservoir seems to stay around 5 meters. A schematic profile of the dam with volumes and elevations is presented in Figure 13.

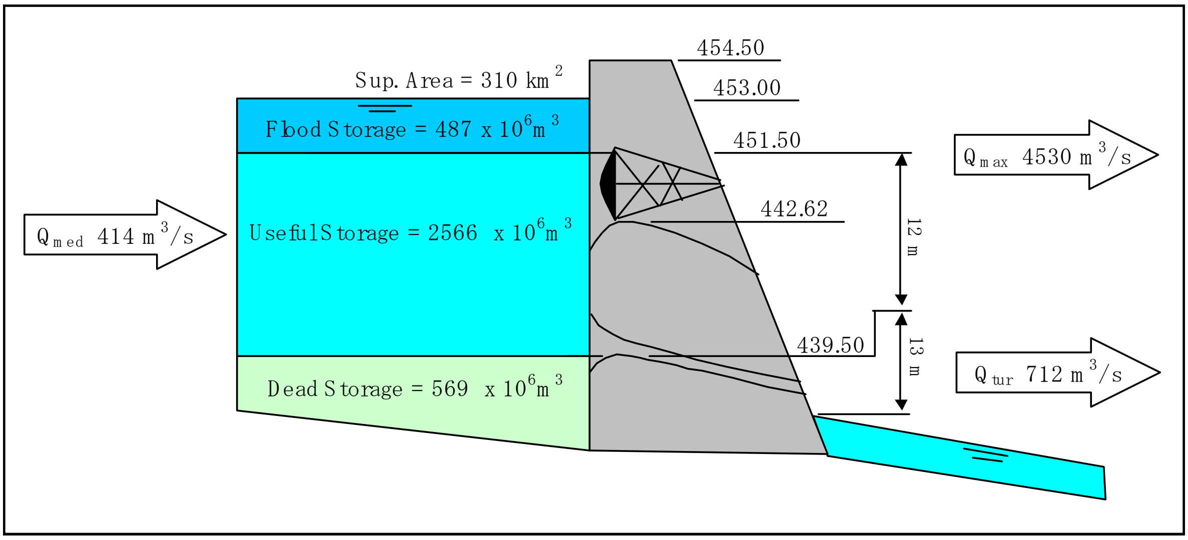


Figure 13. Schematic profile of the Barra Bonita dam site with volumes and elevations

Hydropower energy is the primary water use of the reservoir, having an installed potential for generation of 140 MW. Other uses include navigation, recreation, water supply and fishery production. The location of Barra Bonita and other dams of the Tietê River are schematically shown in Figure 14.

Figure 14. Location of dams in the Tietê river

The Reservoir is the first of a series of six reservoirs, around 300 km downstream from Brazil's biggest city, São Paulo. It can be classified as a subtropical/tropical reservoir with intermediate retention time of around one month. Air temperature normally varies within a range of only 15^{o}C between winter and summer. The wet season occurs between September and March. Annual cumulative precipitation is around 1400 mm, with maximum wind velocity of 5 to 7 m/s during winter. Rainfall, wind and flushing rates are the major driving forces on water behavior in this reservoir.

Water level and volume are related to both climatologic conditions and water use. The average annual flushing rate is 414 m^3/s. Changes in release discharge rates are an important driving force in this system, rapidly modifying ecological conditions within the reservoir, as well as downstream. Stratification does not seem to be very strong in the reservoir due to the medium residence time and its shallowness characteristics. Nevertheless, stratification is significant enough to create undesirable water quality problems. Two main rivers, Tietê and Piracicaba, flow into the reservoir. Due to the considerable difference of the water quantity and quality of the two rivers, the reservoir presents spatial heterogeneity concerning the quality conditions.

The general source of impacts to the Reservoir are nitrogen and phosphorous input from non-point and point sources; input of suspended material from agricultural activities and runoff during precipitation; navigation, tourism and recreation related activities; deforestation in the watershed and the introduction of exotic fish species. Some consequences of these impacts include: eutrophication; siltation; blooms of Cyanophyta (Microcystis), especially in summer and Anabaema, in tropical winter – this impact may be minimized by the controlled operation of the spill water (Tundisi, 1990).

4.2.2 In-site Inspection to the Reservoir

In March 2001, a technical visit was carried out to the site of the Barra Bonita reservoir for the in-site inspection of its quality. This inspection was carried out together with the cooperation of the International Institute of Ecology (IIE), Brazil and the International Lake Environment Committee Foundation (ILEC), Japan and the UNET-IETC, Japan. All of them cooperated with staff and technical support for the success of this technical visit. In Appendix V some of the photos taken during this inspection are shown. Note that green-color tones presented in the water surface indicate the high degree of eutrophication of the Reservoir.

4.3 Data Collection

With the progress of information technology, it is becoming easier to find data on the Internet and in digital format. Hence, it may increase confidence in acquired data. Much of the information used in this work was obtained from different Brazilian organizations, either from direct contact or through their homepages. However, it is still very difficult to have access to more detailed data, especially due to the non-existence of such data.

Most of the hydrologic observation stations in Brazil have a very short period of observed data. Usually, this data do not show good spatial distribution and time continuity. It is due not only to the vast area of the Brazilian territory, but also to political instability and lack of proper investments in the area.

4.3.1 Hydrological Data

For this study, discharge data and reservoir water elevation data from the last thirty-two years (from January, 1970 to December, 2002) are used. Through the use of the elevation versus storage curve, it was possible to convert water elevation into storage volumes. Figure 15 shows the observed storage variation for the years from 1970 to 2000, and the average value.

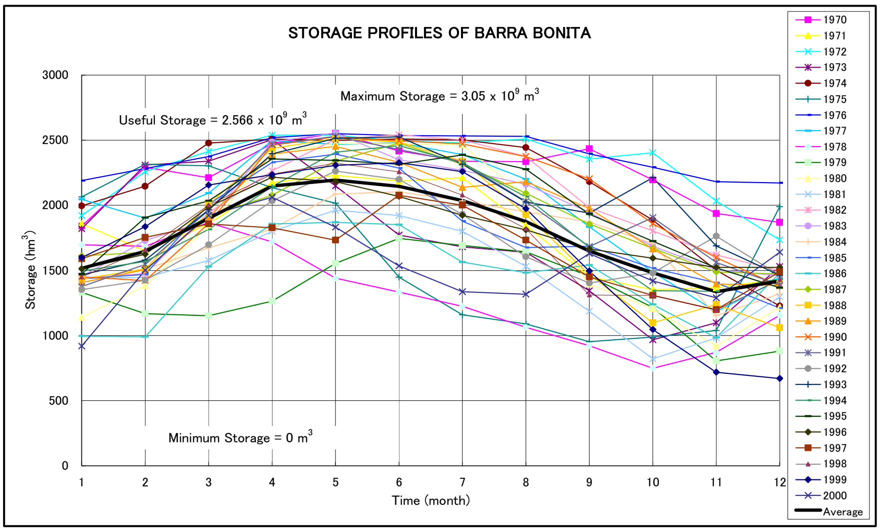


Figure 15. Storage variation for Barra Bonita reservoir (1970 – 2000)

As for the inflow data into the reservoir, two observation points were considered and compared with expected values. Both of Laranjal Paulista and Piracicaba stations are located upstream from the reservoir in the Tietê and Piracicaba rivers, respectively. Release from reservoir was also considered in the calculation of the expected total inflow. The hydrological data was kindly provided by the AES Tiete S.A., which is responsible for the operation of most of the dams of the Tiete river system, including the Barra Bonita reservoir. Figure 16 shows the Reservoir and the location of the observation points.

Figure 16. Location of observation points of the Barra Bonita reservoir system

The expected total inflow into the reservoir was calculated through the use of the mass conservation, or continuity equation, which considered inflow to be equal to the variation of storage plus release. After plotting expected inflow and the total discharge of Tietê and Piracicaba rivers, a coefficient of 1.35 was assumed to be able to achieve an acceptable consolidation between the two sets of values, as shown in Figure 17. Missing data were also completed considering the water balance of the reservoir.

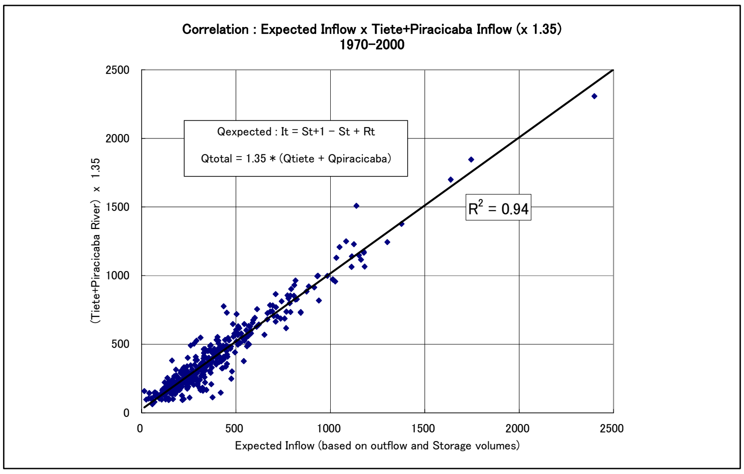


Figure 17. Consolidation of observed data considering the water balance

Figure 18 shows the average of storages and discharges of Barra Bonita reservoir, before and after the consolidation of discharge values. Note that evaporation or other losses were not considered for the development of this work. These losses are also compensated by the 1.35 coefficient used in the continuity equation.

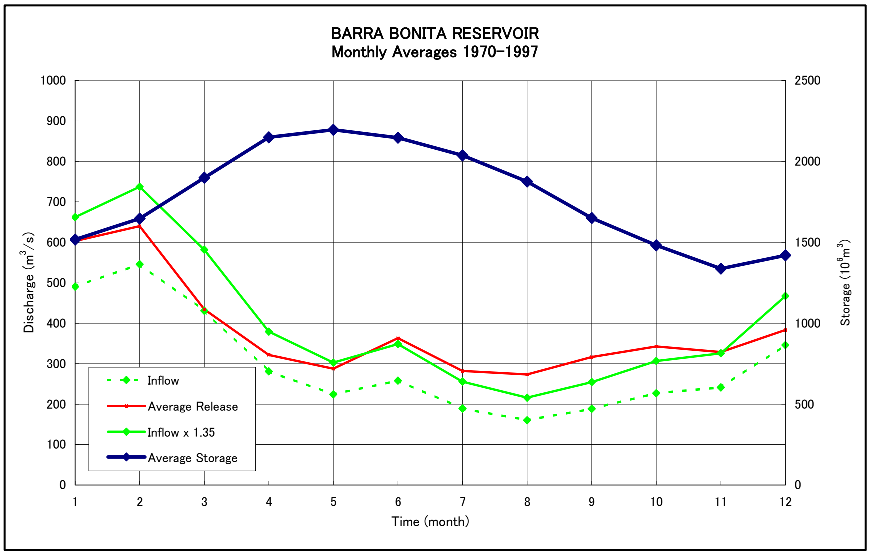


Figure 18. Average storage and discharges of Barra Bonita reservoir

4.3.2 Water Quality Data

Quality data were obtained from the CETESB, Brazil, from different locations within the basin. The location of some of the observation stations can be found in Figure 16 or with more detail in Appendix IV. The TIET02450 observation station provides the inflow quality data for the Tietê River. However, for chlorophyll, it was necessary to use information from stations TIET02350 and TIET02300, due to lack of enough data. As for the Piracicaba River inflow quality, data from the PCB02500 station is used. As for the data used in the calibration of the water quality model only the central part of the reservoir after the junction of the two main tributaries, Tietê and Piracicaba, is considered. Therefore, data from stations TIBB02900 and TIBB02700 were used. Table 2 shows the average values of some water quality parameters for the central part of the reservoir.

Table 2: Water quality with the central part of the reservoir near to the dam

Month	Twater	pH	DO	$BOD_{5,20}$	COD	TN	TP	Turbidity	Coliforme	CHA*	WQI
	ºC		mg/l	mg/l	mg/l	mg/l	mg/l	UNT	NMP/100ml	mg/l	
JAN	28.2	8.0	5.9	5.0	21.7	3.7	0.1	23.0	378.7	0.0258	65.8
FEB	27.1	7.6	5.9	2.3	19.3	2.8	0.1	16.5	3702.5	0.0308	77.5
MAR	27.8	7.3	5.9	2.4	21.7	1.6	0.1	13.2	22.7	0.0050	77.2
APR	25.5	6.9	6.4	1.5	11.0	1.7	0.1	20.8	395.0	**N/A**	87.5
MAY	22.5	7.1	6.1	2.5	16.3	1.7	0.0	11.9	33.0	0.0022	75.3
JUN	19.0	6.9	7.0	1.0	9.8	1.8	0.1	5.3	94.0	**N/A**	89.0
JUL	19.0	7.1	7.7	3.5	25.0	2.5	0.0	6.0	26.8	0.0265	75.5
AUG	20.0	7.6	9.6	6.3	17.8	3.7	0.1	5.6	236.0	**N/A**	63.5
SEP	24.7	7.7	9.0	6.3	26.3	5.8	0.1	8.4	17.2	0.0437	72.8
OCT	22.5	6.7	5.0	4.3	14.3	3.4	0.0	2.4	365.0	**N/A**	64.5
NOV	26.3	9.0	9.5	7.7	47.2	3.1	0.1	15.2	41.0	0.0535	58.5
DEC	25.8	7.6	7.1	1.8	16.3	3.6	0.0	3.6	10.5	**N/A**	86.5

CETESB (Monthly averages from 1998 to 2000)

N/A – data not available

5 Water Quality Simulation Under Uncertainty

5.1 Introduction

Artificial neural network (ANN) models have been used widely in various fields, such as aerospace, finance, robotics, environmental assessment and hydrology, for different purposes such as classification, pattern identification, simulation and prediction (Hagan *et al.*, 1995). However, ANN models depend on large sets of historical data, and are of limited use when only vague and uncertain information is available, which leads to difficulties in defining the model architecture and low reliability of results. Therefore, new proposals are needed towards the improvement of ANN. Towell and Shavlik (1994) present some other disadvantages and weaknesses of traditional ANN.

ANN models are usually characterized as “black-boxes”, since such models are basically data-driven engines. Natural concepts and characteristics of the system are usually not taken into account in a traditional ANN model structure. Only input and output represent some of the system characteristics. After training, it is expected that the ANN model is able to generalize the behavior of the system, when new data of the same type is given as input.

Not having to understand the processes within an ANN model makes application of ANN models straightforward, particularly when compared to physical models. However, if the training data sets do not adequately represent the system behavior, this may result in unrealistic or imprecise outputs. Problems associated with data may be a consequence of data sets that are too small, inappropriate methodology in data collection or errors in the observed values themselves. Hence, the data used in the training process is critical to obtain an efficient ANN model. However, in many practical cases, the problem of unrepresentative training data sets are commonly faced by ANN users, especially in environmental modeling, where data acquisition is usually time-consuming and very expensive.

Due to the absence of physical concepts in traditional ANN models, when input variables assume extreme values or new patterns out of the scope of the training data set, the output may be inaccurate or even completely out of the realistic physical range. In an attempt to overcome such problems, the combination of ANNs with other techniques, such as fuzzy theory, has already been proposed by some authors. The use of fuzzy theory with the ANN aims to combine the ability of fuzzy sets to represent knowledge that is understandable to human beings with the learning capability of ANNs. Therefore, a neuro-fuzzy system is a fuzzy system that can learn from data and as well as experts' experience. This can be very effective for dealing with a system where data and information are vague or uncertain, but some knowledge about the system is available. Application of fuzzy theory with ANNs may change the model from being a "black-box" into a "gray-box" type. As described by Jin (2003), there have been some attempts to develop neuro-fuzzy models and the best known example is the adaptive neuro-fuzzy inference system (ANFIS) proposed by Jang and Sun (1993, 1995).

In general, fuzzy theory is used in the input layer, to deal with the fuzzification of input values. Hence, it enhances the net understanding of the input values and helps prevent the model from behaving completely unrealistically when faced by unexpected input conditions. Nevertheless, in fact neuro-fuzzy models still behave mostly as data-driven engines, where model outputs are mostly dependent on the quality of the training data. Moreover, the training process may become more complex and time-consuming, as extra parameters are included in such models (Jin, 2003).

Some attempts to combine conceptual characteristics with or within the ANN have been carried out. Huttunen and Vehvilainen (1996) proposed coding a conceptual model into an ANN, where the nodes, or neurons, represent the components of the conceptual model. Their proposal is limited by the time required to find the optimal structure, weights and other parameters, a process that can take weeks or even months. Moreover, their model depends on various types of data, and vagueness of the system is not at all considered. Another attempt to combine ANNs with physical concepts was that by Cozzio-Bueler (1995) who coded differential equations directly into an ANN. Additionally, Oussar and Dreyfus (2001) have proposed a semi-physical modeling approach, where parts of a physical model are substituted by a traditional ANN. Wilby *et al.* (2003) attempted to detect the presence of a rainfall-runoff model process inside a traditional ANN model. This study concluded that some characteristics of the rainfall-runoff process, such as base flow and quick flow, have been naturally represented by the hidden neurons of particular structures of an ANN model.

Le Cun (1990) proposes a new method for the elimination of connections within an ANN, which may be considered irrelevant to the model's output. By removing such connections, Le Cun suggests that many improvements may be obtained, such as better generalization, fewer training examples required and improved speed of learning and classification. However, the proposed method does not consider the physical processes involved for the removal of connections. It is decided automatically in an interactive way as a tradeoff between network complexity and training error. As stated by Le Cun (1990), simpler structures may reduce the necessary training time of ANN based models. However, the most relevant method up until now for the insertion and extraction of symbolic knowledge considering ANNs are the knowledge-based artificial neural networks or KBANNs proposed by Towell and Shavlik (1994). Some examples using the KBANN are presented by Mitra *et al.* (2001) and Yiyao *et al.* (2001). Fu (1995) briefly describes three types of KBANN, which are based on rule inference, decision trees and semantic constraints.

5.2 Conceptual Fuzzy Neural Network

We propose a new type of fuzzy neural network model, hereafter referred to as conceptual fuzzy neural networks (CFNN) that can be applied even when data provide some information about the system; however it is either not complete or is not considered entirely trustworthy. This can be achieved by combining characteristics of ANNs with fuzzy theory and basic biological, chemical and physical concepts, rationally defining the network structure and connections through the use of accumulated expert knowledge about the system and processes involved. The necessary knowledge for the development of a CFNN model can be obtained as it is done for other models. Basically, there are three general ways of extracting knowledge about a certain system in order to define types of connections for the CFNN model:

a) Well-established physical equations, concepts and laws.
b) Accumulated knowledge: classification, descriptive and statistical analysis from past works; empirical analysis and “common-sense”.
c) Analysis of the accuracy and reliability of available data.

A method of knowledge extraction, which obtains crude knowledge from the available data, is known as rough set theory. The extracted knowledge can then be encoded into an ANN model (e.g., Banerjee *et al.*, 1998; Mitra *et al.*, 2001) or combined with it (Yasdi, 1995). Once again, however, this approach may be problematic when available data are not fully representative.

5.3 Architecture of a CFNN

Hagan *et al.* (1995) pointed out that: “a two-layer network having a sigmoid first layer and a linear second layer can be trained to approximate most functions arbitrarily well,” and “as for the number of layers, most practical neural networks have just two or three layers (four or more layers are used rarely)”. Moreover, the elimination of irrelevant connections may not affect the efficiency of the model but on the contrary it may increase the ability of generalization and avoid problems related to over fitting as well. Based on these considerations, a CFNN is structured as a three-layer network, with input, hidden and output layers. In contrast to the traditional ANN, the number of neurons in the hidden layer of a CFNN must be rationally defined, based on previous accumulated understanding of the system. The neurons of the hidden layer(s) are to be defined as fuzzy values, representing the output variables. Even though the basic proposed structure presents three layers, depending on the nature and complexity of the system to be modeled, more hidden layers may be introduced.

Moreover, extra units may also be introduced into the hidden layer(s), accounting for the fuzzy representation of intermediate variables or processes, which are not included in the input layer. The extra neurons must be correlated to the input and output variables. These neurons may be omitted in the input layer due to lack of data, even though some knowledge about them may be available. The introduction of such neurons may increase the efficiency and flexibility of the CFNN. The hidden neurons representing fuzzy outputs are also expected to influence the final value of other output variables, also in a predefined, rational way. This is different from the other fuzzy ANN models, as in the CFNN, fuzzy inference rules are defined based on the type of connections between two consecutive nodes.

Each neuron of the input layer represents important factors regarding the output variables. The input variables do not need to be normalized before being input into the model, as the CFNN is prepared to accept their actual values and units. The connections between input and hidden neurons are carried out through the fuzzification process of input values, defined by fuzzy inference with rules stipulated based on the type of connection. Then, the input information can be combined through a transfer function generating the fuzzy rule inference output.

The last layer to be introduced is the output layer, which is composed of output neurons representing the variables of interest to be modeled. The neurons of the output layer in a CFNN model may either provide results with the actual value and unit of the simulated variables or as linguistic values (in a scale from 1 to K, where K is the maximum number of linguistic variables). This information can be extremely useful, for example as shown later in the section about training the network, as the accuracy of the results are evaluated based on their fuzzy evaluated importance (or linguistic value, e.g. “high” concentration) rather than their actual units. Moreover, it becomes easier to stipulate or assume some vagueness when final results are given as linguistic values.

Figure 19 presents a schematic representation of basic components of the CFNN model architecture, identified here as the layer structure. A layer structure is defined between two consecutive layers, for example input/hidden and hidden/output. The input and output of a layer structure is defined hereinafter as *x* and *y*, respectively, where each neuron of input and output layers are represented by the indices *i* and *o*, respectively. Between two layers, there are four basic processes composing a layer structure, as explained before: 1) fuzzification, 2) connection type, 3) weighted combination of results from previous units by a transfer function and 4) defuzzification. For the fuzzification process, the center and limits values of each linguistic variable, such as *L* (low), *M* (medium) and *H* (high), are defined during the training process or by users, depending on the system characteristics and available data. The rule inference is then applied based on the connection type, defined by users based on the kind of relation or process between two variables or units. Then, the transfer function is responsible for the combination of each fuzzy rule, accounting to the weights *W*, which are also defined during the training process, resulting in the fuzzy inference variable, which after defuzzification results in a crisp value for the output node *y*. Detailed mathematical formulation for each of the components mentioned above is given later.

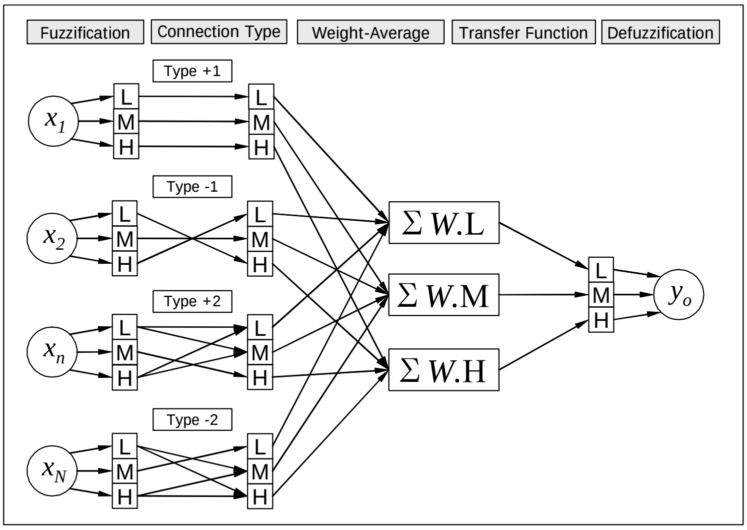


Figure 19. Components of the layer structure of CFNN between two consective layers

5.3.1 Advantages & Applicability of CFNN

- CFNN may reduce the number of connections and parameters to be defined if compared to other neuro-fuzzy systems. Hence, the efficiency of the training process may be improved.
- CFNN is ideal to deal with situations when traditional ANNs should be applied but historical data are either limited, do not represent the system and processes properly or have a high degree of uncertainty.
- The structure of a CFNN model may be easily defined, if the modeled system and the

relevant variables are properly understood.

- The CFNN may also consider the relationship between the output variables themselves, as actually occurs in many real processes when hidden neurons are identified as intermediate fuzzy variables or complex processes.
- CFNN can handle fuzzy inputs and the vagueness of training datasets, addressing uncertainty in a quantitative way.

5.3.2 Disadvantages of CFNN

The process to be modeled has to be understood by the user for defining the network architecture. However, much less understanding is necessary if compared to the KBANN for instance, or even less when compared to a physical model. Moreover, as it may occur with the traditional ANNs, the performance functions used for the training process may greatly influence the final results of a CFNN model. Table 3 summarizes some of the important characteristics of the CFNN comparing them with traditional ANNs.

Table 3. Characteristics of the CFNN compared to the traditional ANN

CHARACTERISTICS	Traditional ANN	CFNN
Handling uncertainty of data	Not Applicable	Applicable
Handling data scarcity	Not Applicable	Applicable
Handling uncertainty of processes	Not Applicable	Applicable
Handling fuzzy values	NO	YES
Definition of models architecture	Trial & Error	Logically defined
Different types of connection	NO	YES
Performance functions	Usually single	Multiple
Learn from data	YES	YES
Learn from Expert	NO	YES
Over fitting	May occur	More robust

5.4 Mathematical Formulation

In this section, more detailed mathematical description of each component of a CFNN model layer structure is presented in general terms, and can be applied for the connections between all layers, such as input-hidden and hidden-output.

5.4.1 Fuzzification

First, let's consider a set of crisp values x_i, where i refers to each input information or neuron unit, as the input information into a layer structure as the one presented in Figure 19. To assess the uncertainty of each input value x_i, the value is fuzzified, as shown in Figure 20. The parameters X_k, defined as the center value of the fuzzy membership function of each linguistic variable k, can be found during the training process or may be defined by users, depending on the characteristics of the system and on the available information. For the CFNN input layer these parameters may be more easily defined by users than the ones for the hidden layer (when it becomes the input layer for the hidden/output layer structure). Then, the membership value $\mu_{i,k}$ of each linguistic value k can be easily obtained. Note that the extreme values can be easily defined taking into account physical characteristics of the input variables. The number of linguistic variables is stipulated according to the fuzziness of the system being modeled or by trial-and-error. Systems presenting high levels of uncertainty may be more efficiently modeled with fewer linguistic variables, as the model may lose some of its capability to deal with uncertain information if the number of linguistic variables increases too much. Moreover, such systems may perform better with fix X_k parameters.

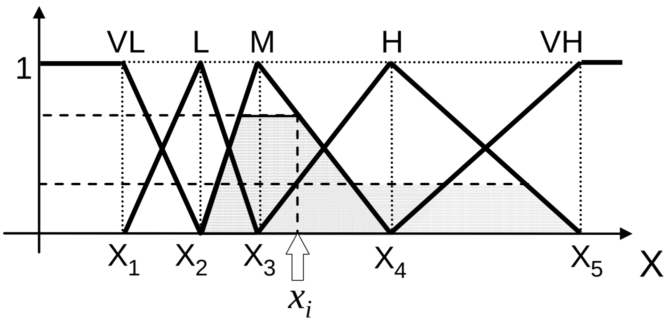


Figure 20. Membership function for the fuzzification of input x_i of a layer structure

Membership functions are defined here as triangular functions; however, other types such as the bell or Gaussian functions may also be used. The problem with such functions is that there would be a larger number of parameters to be defined and in some cases their relation with physical characteristics could be more difficult to understand.

5.4.2 Connection Type

In the CFNN, the input values are fuzzified through fuzzy membership functions. Then, the type of connection, i.e. kind of influence or relation, between two units or variables, may be defined as *increasing*, *decreasing*, *convex* or *concave*. For example, this guarantees that if there is an *increasing* relation, increasing the input x of a layer structure results in an increase of the output value y. This is valid for both, the connections between input/hidden and hidden/output layer structures.

The connections are established by the introduction of a transfer matrix, which is defined by the users and can be somehow compared to the rule inference mechanism of a traditional fuzzy inference system. The transfer matrix defines not only the type of connection between one node and another but also may annul some of the connections. Here, the transfer matrix is constituted by values of *+1* and –1, *+2* and *–2* and *0* for *increasing*, *decreasing*, *convex*, *concave* or *null* influence, respectively. The transfer matrix is defined as C_{io}, representing the connections between two consecutive layers. The transfer matrix is the most important way to provide the expert knowledge about physical, chemical and biological processes to the CFNN model. Besides the five possible types of connections described above, some particular connections may be introduced as predefined functions. Such functions may represent a well-known relationship between two nodes, including physical relationships or any other kind of constraints to be introduced into the model.

Figure 21 shows the rules of the fuzzy inference in the case of five linguistic variables, which associates the linguistic value (e.g. "high") of input information to the one of an output unit. When after fuzzification, a variable is classified as, for example, "high" and its relation to a variable is defined to be *decreasing*, the latter will present a "low" linguistic value associated to it equal to the same value of the membership function of the former. In a more concrete example, for a reservoir with low BOD concentration, high levels of dissolved oxygen (DO) can be expected, therefore represented by a connection type *-1*. Another example considering the connection type *+2*, here defined as *convex*, can be given. The growth rate of algae increases with an increase in temperature, however, it starts to decrease after the temperature is superior to an optimal algae growth temperature. This procedure may greatly simplify the rule inference process of a neuro-fuzzy system, when comparing it to other techniques such as the one used in the ANFIS model (Jang, 1993), which demands the computation of all possible combinations among the rules of all input information.

	Linguist Domain for actual Unit				
Connection Type / Previous Unit Contribution	VL	L	M	H	VH
+1	VL	L	M	H	VH
−1	VH	H	M	L	VL
+2	VL or VH	(VL+L)/2 or (VH+H)/2	L or H	(L+M)/2 or (H+M)/2	M
−2	M	(L+M)/2 or (H+M)/2	L or H	(VL+L)/2 or (VH+H)/2	VH

Figure 21. Rule inference for each connection type

5.4.3 Transfer Function

The resulted membership values $\mu_{i,k}$, as shown in Figure 20 for each linguistic variable *k*, found after the fuzzification of the crisp input value *x* are then combined according to the influence of each previous node (connection type). There are various methods used within fuzzy inference theory, such as product, minimization and maximization. However, after trying these different methods, the weighted sum is found to be the most appropriate as robustness of the whole system increases, as the model holds its generalization ability even after different input conditions are presented, avoiding the problem of overfitting (overtraining). The resulted fuzzy number $\mu_{o,k}$ after the combination through the transfer function of previous units is presented in equation (18).

$$\mu_{o,k} = \frac{\sum_{i=1}^{N} (W_{io} \cdot \mu_{i,k})}{\sum_{i=1}^{N} W_{io}} \qquad (18)$$

Where *N* is the total number of previous units and W_{io} is the weight between units x_i and y_o of the same layer structure, represented by indices *i* and *o*, respectively. *W* is also optimized during training; however, as all neurons are *a priori* identified, these weights may be also defined by users, if proper information is available. The fuzzy value for each linguistic variable *k* is obtained from the respective previous neurons *x* and the type of connection between *x* and *y*, resulting in the $\mu_{o,k}$ which can be graphically represented as in Figure 22.

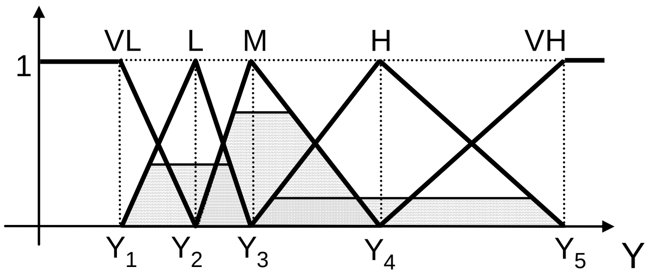


Figure 22. Graphic representation of the resulted fuzzy number $\mu_{y,k}$

5.4.4 Defuzzification of $\mu_{o,k}$

Knowing the fuzzy value $\mu_{o.k}$, it is possible to obtain a crisp value y_o through many available defuzzification methods. The most common defuzzification methods are the center of gravity, center of sums and center of areas (Jin, 2003). A method based on the center of gravity and described in Jin (2003) is adopted here, as shown in equation (19).

$$y_o = \frac{\sum_{k=1}^{K} (\mu_{o,k})^g \cdot Y_k}{\sum_{k=1}^{K} (\mu_{o,k})^g} \qquad (19)$$

Where Y_k is the kernel or the modal value, $g \in [0,\infty)$ is a power parameter responsible for "firing the rule", which can be properly adjusted to obtain a more efficient performance for the defuzzification method.

As already acknowledged before, the limits of the fuzzy inference variable can be defined as in the fuzzification process, such as automatically during the training process or based on expert experience and the data condition, e.g. taking into account the probability of the observed values. For the hidden layer of the CFNN, this may not be possible, so the automatic process may be used. Nevertheless, even if data are not available, as hidden units are supposed to be a fuzzy representation of real processes and variables, some knowledge about these processes may help the definition of such limits.

In general terms, some guidelines for defining the parameters of the membership functions for the (de-)fuzzification processes, when five linguistic variables are presented as in Figure 22, can be seen in Table 4. The center value for the *low* and *high* linguistic variables can be assigned as equal to the average between the mean value *M* and the minimum *VL* and maximum *VH* values, respectively, as shown in Table 4, line 1. In the same way, cumulative probability may also be used, line 2. Another way to define such limits may be based on other criteria, such as based on classes and their limits when information is already predefined and known, as shown in line 3, e.g. as the acceptable standard limits for water quality evaluation.

Table 4. Three methods for defining the center values of fuzzy membership of linguistic variables *k*

	Center Value for FIVE Linguistic Variables *k*				
	VL	L	M	H	VH
Based on Minimum, Averages & Maximum Values	Zero or a physical value	Average (VL & M)	Mean (Observed data)	Average (M & VH)	Maximum physical value
Based on Occurrence Probability of observed data	Zero or a physical value	25% of Occurrence Probability	50% of Occurrence Probability	75% of Occurrence Probability	Maximum physical value
Based on Evaluation Classes and their predefined limits	Zero or a physical value	Limit value representing a *low* class	Limit value representing a *medium* class	Limit value representing a *high* class	Maximum physical value

5.5 Performance Function

The most known performance functions for the training of ANN models are the functions based on error between target and output vectors, such as the mean square error (MSE), root percentage mean error (RPME) and the absolute error (AE) functions. Nevertheless, in the case where data are too scarce or present a high degree of uncertainty and the model has multiple outputs, the use of performance functions based solely on the error vector can be extremely inefficient and may result in a model with a set of improperly trained parameters.

The main reason for this is that with too few observations or target vector, the error of one single target value (which may or may not be the most representative one, for instance a common problem related to water quality observations) may be extremely large and can mislead the training process. For instance, the training algorithm may yield results that tend to the most frequently observation values, usually around the average, resulting that the few extreme values may not be at all considered, or vice-versa. Moreover, the use of the percentage error function clearly tends to favor the lower values. In order to overcome such problems of data adequacy, some other performance functions are proposed. All the functions proposed are determined within the interval [0,1], where 1 represents the best and 0 the worst performance or satisfaction regarding the model outputs.

To be able to address the training of multiple vague outputs, first the output values are fuzzified using linguistic variables (e.g. low, medium, high). But, what is a low or a high value? The answer to this question is not straightforward. But, basically, it can be said that the label and the limits defining it, should be defined according to the purpose of analysis, data conditions and expert experience. This fuzzification is extremely important as the degree of fitness between target and output values is highly dependent on the limits of the linguistic values. Let's take our case of water quality modeling, where small variations (uncertainties) or inaccuracy can be extremely important in some cases and less important in others, due to the non-linearity of water quality processes and evaluation functions. For example, an absolute or either a relative error in the model output may or may not represent a different evaluation of the water quality condition.

On the other hand, if actual values with actual units are of interest, the fuzzification should be based on the actual data, for example, a "medium" linguistic variable can be represented by the mean or average value, and "low" and "high" may be referred as the minimum and maximum observed values from the historical data, respectively. This discussion has been already summarized in Table 4. The actual limits used for calculation of representative number are shown is Table 5 below.

Table 5. Center values used for the definition of representative numbers

k	Linguistic Meaning	Symbol	Parameters				
			DO (mg/l)	BOD (mg/l)	TN (mg/l)	TP (mg/l)	CHA (10^{-3}mg/l)
1	Very Low	VL	0.0	0.0	0.0	0.0	0.0
2	Low	L	3.0	2.0	1.5	0.04	11.9
3	Medium	M	6.0	3.9	3.0	0.08	23.8
4	High	H	8.0	9.5	6.5	0.19	61.9
5	Very High	VH	10.0	16.0	10.0	0.30	100.0

After the fuzzification of target and output values, a "representative number" can be found and the performance functions for the training process may then be applied. The representative value r can be easily calculated using equation (20), based on the summation of the membership value $\mu_k(y)$ of the variable of interest, here y, multiplied by the integer value k : [1, 2,…, K] which also represent the index for each linguist variable and K is the maximum number of the linguistic variables.

$$r = \frac{\sum_{k=1}^{K} \mu_k(y) \cdot k}{\sum_{k=1}^{K} \mu_k(y)} \qquad (20)$$

In an attempt to handle and to assess qualitatively the uncertainty of data, various functions have been tested for the training process. Following, we introduce the functions developed, which presented satisfactory efficiency for training the CFNN water quality model.

The representative values are used hereafter for the calculation of the performance functions. This is done in an attempt to efficiently handle the uncertainties of the observation values, which may be originated by measurement itself or the temporal and spatial incompleteness of the dataset.

It could be very difficult to determine the uncertainties of the observation values directly. However, using the representative values makes it easier for defining a range or a credibility interval referent to the calculated values. For example, for the case of five linguistic variables ranging from 1 to 5, a range of one could be considered acceptable for a system with high uncertainties, such as the one presented as our case study. That is to say, a calculated representative value r of 2.3 could range between r_1 = 1.8 and r_2 = 2.8, where r_1 and r_2 are the lower and upper credibility limits, respectively, which are used for the calculation of one of the performance functions presented later.

5.5.1 Performance Function (PF_{mse}) of MSE of Linguistic variables

The percentage error (PE), such as the percentage mean square error (PMSE) and root percentage mean error (RPME), can be used for training ANN models with multiple output variables. Nevertheless, as already described before, the use of percentage error solely as a performance function, in the case of scarce data sets with high uncertainty, can be extremely inefficient. On the other hand, with some modifications, it can become a useful performance function, as it is still an important indicator of the fitness of the output in relation to the target vector. Therefore, for this performance function, the values used for calculation are based on the representative values *r*, where *r* is a real number within the interval [1, *K*] and *K* is the total number of linguistic variables *k*. Moreover, to address the importance of each target value, not only the most frequent ones, the mean square error (MSE) is calculated for each linguistic class *k*, as in equation (22) and then averaged, as shown in equation (23). This performance function tends to work as the functions used for the error calculation of minimal and maximal values.

$$MSE_k = \frac{1}{N_k}\sum_{i=1}^{N_k}\left(\frac{r_{t,i} - r_{o,i}}{K-1}\right)^2 \quad \text{for} \quad k-1 \le r_{o,i} \le k+1 \tag{22}$$

$$MSE = \frac{1}{K}\sum_{k=1}^{K} MSE_k \tag{23}$$

Where $r_{t,i}$ and $r_{o,i}$ are the representative values calculated for target *t* and output *o* vectors, respectively, for the points *i* of each class *k*. For example, a point $i \in$ class *k* if $k\text{-}1 \le r_{o,i} \le k\text{+}1$. N_k is the total number of points for each linguist class *k*, and *K* is the maximum number of classes (linguistic values). Finally, the performance function (P_{mse}) can be calculated based on equation (24). This function has the limits also between [0,1], which is important when combining the performance functions with the product aggregation type as shown later. A graphical representation of this equation is presented in Figure 23.

$$PF_{mse} = 1 - \sqrt{MSE} \tag{24}$$

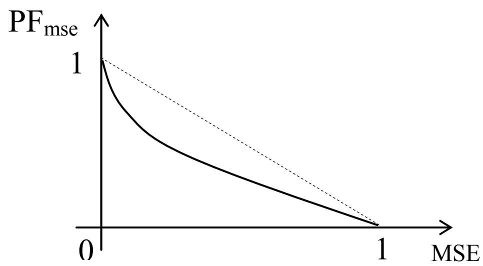


Figure 23. Performance function *PF* **PF_{mse} as a function of MSE.**

5.5.2 Performance Function (PF_{vague}) of Output Vagueness

The second performance function, named Output Vagueness and represented by PF_{vague}, is also based on the error between the observed and calculated representative values r, but this time, considering the credibility interval and considering the observed values as a whole (not for each linguistic class as performed for the previous function). The basic equation for PF_{vague} is presented below:

$$VAGUE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_o - r_1}{K-1} + \frac{r_o - r_2}{K-1}\right)^2 \quad (25)$$

$$PF_{vague} = 1 - \sqrt{VAGUE} \quad (26)$$

Where K is the total number of linguistic variables, T is the total number of target values; r_o, r_1 and r_2 are the representative values for observation, lower and upper limit of the credibility interval of the calculated result, respectively.

5.5.3 Aggregation of the Performance Functions

One of the biggest challenges in solving multi-objective problems is to decide how to combine all the objectives involved. In our case, this problem is also faced for the aggregation of the different proposed performance functions. There are various simple combination methods to handle multi-objective problems. Some of the basic methods include minimization or maximization and simple or weighted average and product. All of these techniques have been tested through a trial-and-error method and the results have been quantitatively and qualitatively analyzed. The combination, which has proven to present the best results, is the one which uses product aggregation.

The proposed performance functions are defined in the case where only one pair of target and output vectors is considered. For multiple output models, such as the one developed here for water quality simulation with five quality parameters, the same functions can be applied. The difference now is that individual results from each output must be evaluated between themselves. Therefore, the exponential aggregation is applied among the performance functions of the output variables. So, the final function for a certain PF_k considering variable n is described as:

$$PF_k = \left(\prod_{n=1}^{N} PF_{k,n}\right)^{\frac{1}{N}} \quad (27)$$

Where the index n represents the output variable (e.g. dissolved oxygen and total phosphorus) and index k the performance function (e.g. PF_{mse} and PF_{Vague}).

Then, the final objective function can be mathematically represented as in equation (28):

$$PF_{final} = PF_{mse} \cdot PF_{Vague} \tag{28}$$

Where PF_{final} is the final performance function, which combines all the other performance functions related to the fitting of the modal value of the output to the target vector. In conclusion, the use of multiple performance functions contributed to an increase in robustness of the calculated results when only scarce and vague data are available for training CFNN models.

5.5.4 Training Algorithm

The GA optimization model is constructed with five operators - elitism, crossover BLX-α, multi-points crossover, simple mutation and creep mutation applied to real numbers - instead of the more common binary formulation. Yao (1999) explains some of the advantages of evolving algorithms based on the real number formulation. Calculation is carried out for the optimization of the total number of fuzzy inference parameters (or center values of the linguistic variables) and connection weights to be trained. Moreover, The objective function used for GA optimization is shown in equation (28). Further information on the formulation of the developed GA model has been provided in section 3.6.

5.6 CFNN Water Quality Model

The proposed CFNN is applied in the development of a reservoir water quality simulation model. The application of ANNs for the simulation of water quality and ecological characteristics can be found in many studies, such as those by Maier and Dandy (1996), Schleiter *et al.* (1999), Maier and Dandy (2000), Recknagel (2001) and Lee *et al.* (2003). The objective of the proposed model is to simulate the overall monthly water quality condition of the reservoir system, considering the time dependence for the quality variables, that is to say, a recurrent neural network, where calculated outputs in the present stage are used as input information for the following time-step, in the same way performed by dynamic models.

The observation data available on water quality are very limited. Observations are carried out every one or two months, in shallow depths near to the water surface, during daytime and only a few years of observation are available (1995 and from 1997 to 2002). The available observation data are plotted later together with the final results of the CFNN model, as well as the length of the period considered for training and validation of the model. Two observation stations located within the reservoir water body are averaged and used as the target data. For the inflow quality two other observation stations are used representing the average loads from the two main tributaries, Tiete and Piracicaba Rivers, respectively. The uncertainty involved in this simulation can be considered extremely high, besides other sources of uncertainty, such as mis-measurement, the training process is carried out with incomplete data sets that are assumed to represent the whole month even though they are just daily observations.

Under these conditions, the use of traditional neural networks may result in a catastrophic simulation, and therefore, should be avoided. Even though physical models do not require long data sets, they usually demand a much greater variety of input data and information on physical parameters of the system. In his extensive review on the uncertainties related to water quality modeling, Beck (1987) had already suggested that the water quality modeling through linguistic means, such as based on fuzzy theory, could be an efficient way in dealing with such problems, due to the unpredictable uncertainties related to the quality modeling process. Therefore, after all of these considerations, the CFNN is assumed to be an appropriate model to achieve more efficient simulation of the reservoir water quality.

The water quality model is developed to simulate five basic quality parameters, considering only simple climatologic and hydrological variables. For the output layer, five water quality parameters are chosen for simulation: dissolved oxygen (DO), biochemical oxygen demand (BOD), total nitrogen (TN), total phosphorous (TP) and chlorophyll (CHA). The five parameters are fairly representative of the basic characteristics of water quality and are broadly used worldwide in the evaluation of water quality conditions. Moreover, they can represent the eutrophication process which is one of the main problems regarding the reservoir water quality.

5.6.1 Basic Concepts of Reservoir Water Quality

It is often useful to describe lakes and reservoirs under the assumption of completely mixed systems. This can be justified in the basis of wind stress on the water surface, reservoir geomorphologic and hydrologic characteristics and the long scale of the problem (monthly time-steps). For a more detailed explanation about the mathematical formulation of physical, biological and chemical of water quality processes, the readers are recommended to specific references, such as Jorgensen (1983), Orlob (1983) and Thomann and Mueller (1990). The general mass balance equation for the reservoir system is presented in equation (29).

$$C_{t+1}S_{t+1} = C_t S_t + W_t - C_t R_t - K S_t C_t \qquad (29)$$

where C represents concentration, S storage levels, R release and W the total mass input. The index t represents time. K is a first-order substance decay coefficient. For the right side of equation (29), the first term represents the actual mass in the reservoir, the third is the total mass output related to reservoir release and the last represents the substance decay. The total mass input W can be a function of various input sources, such as wastewater effluents, loads from rivers, precipitation and internal sources like releases of substances from sediments. However, for our case, all the mass input contribution are combined in a single factor according to the input originating from rivers. Moreover, the decay component may be substituted by the net loss settling due to the sedimentation of suspended solid, which in the case study seems to have a great influence over the quality concentrations, such as phosphorous in sediment and sediment oxygen demand. A schematic representation of the natural processes considered hereafter is presented in Figure 24.

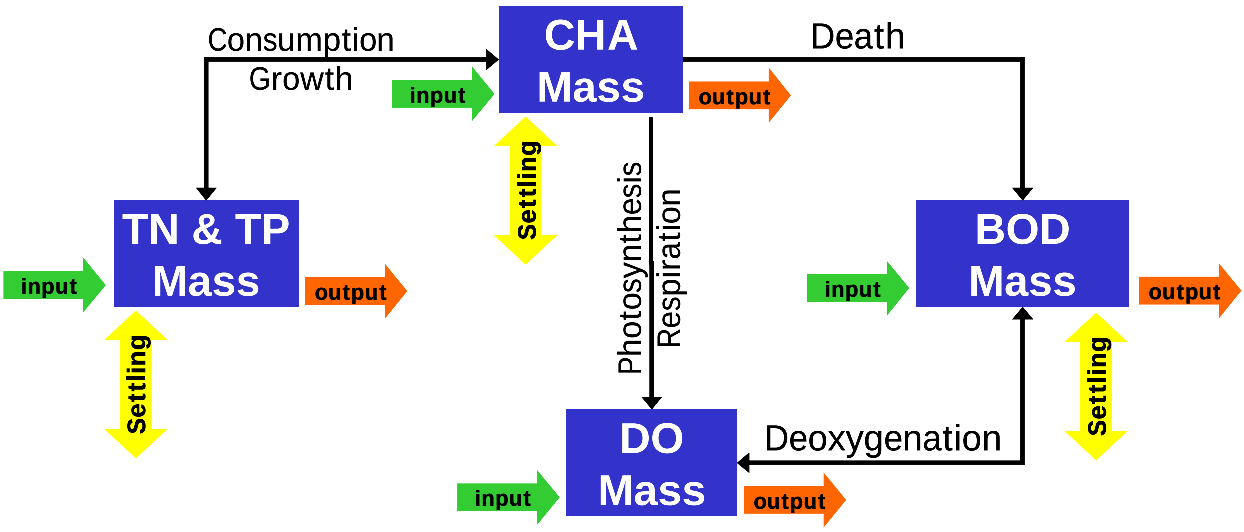


Figure 24. Basic processes and relations considered for the CFNN quality model.

For the case of phytoplankton mass balance, equation (28) presents some changes, as the net growth rate has to be considered. The settling rate in this case is also considered, even though in some cases it can present quite low values. The considered final mass balance equation for phytoplankton (chlorophyll-a) is presented as follows:

$$CHA_{t+1}S_{t+1} = CHA_t S_t + CHA_{I,t} I_t - CHA_t R_t + S_t(G - D)CHA_t - K_s S_t CHA_t \qquad (30)$$

Where K_s is the effective loss rate due to CHA setting and is, besides other factors, related to the retention time. G and D are the growth and death rates of phytoplankton, CHA are the phytoplankton chlorophyll concentration and the other variables are defined as previously. Growth depends on three principal components: temperature, light and nutrients. On the other hand, the death rate is associated with water temperature and zooplankton grazing. Note that, theoretically, all these factors have a positive influence on the growth and death rates.

Nevertheless, for the sake of simplicity, only temperature and nutrients are considered for our case study due to the lack of data regarding the other variables.

Knowing that phytoplankton growth rate depends on nutrients, it is expected that high growth rates result in more loss of nutrients from the water column. Therefore, it is important to include this relation in the mass balance equation of nutrients. The equation below presents the total phosphorous mass balance equation:

$$TP_{t+1}S_{t+1} = TP_tS_t + TP_{I,t}I_t - TP_tR_t - K_cS_tCHA_t - K_sS_tTP_t \tag{31}$$

Where the fourth term on the right side represents the reduction due to phytoplankton growth, K_c is the TP consumption coefficient by phytoplankton, which is a function of CHA concentration. K_s is the effective loss rate due to TP setting and *TP* is the total phosphorous concentration. Note that the total nitrogen (TN) mass balance equation can be written in the same way as the one for TP, as shown in the equation below:

$$TN_{t+1}S_{t+1} = TN_tS_t + TN_{I,t}I_t - TN_tR_t - K_cS_tCHA_t - K_sS_tTN_t \tag{32}$$

Where K_s is the effective loss rate due to TN setting and K_c is the TN consumption coefficient by phytoplankton, which is a function of CHA concentration.

The dissolved oxygen (DO) and its relation to phytoplankton and other oxygen demands can be presented as in the basic mass balance equation (33), which is similar to the one presented in equation (29), after the addition of other factors such as reaeration, deoxygenation by BOD and other sources and sinks of DO, including photosynthesis, respiration and sediment oxygen demand (SDO).

$$DO_{t+1}S_{t+1} = DO_tS_t + DO_{I,t}I_t - DO_tR_t + K_aA_t\left(DO_{sat} - DO_t\right) - K_dS_tBOD_t + K_rS_t(G-D)CHA_t \tag{33}$$

Where K_d is the effective deoxygenation rate and K_a is the reaeration coefficient. A_t is the surface area between the water and the atmosphere. However, for the proposed model, the reaeration component is not considered assuming that effects due to the reaeration may be ignored in the case of monthly time-scale. The photosynthesis and respiration effects may be considered by including a coefficient that can represent the increase of DO due to photosynthesis, defined by K_r. In the same way, effects due to SDO may be considered by the deoxygenation component.

The last mass balance equation to be present is related to the biochemical oxygen demand (BOD), considering as well a completely mixed system, the equation can be presented as follows:

$$BOD_{t+1}S_{t+1} = BOD_tS_t + BOD_{I,t}I_t - BOD_tR_t - K_dS_tBOD_{t-1} - K_sS_tBOD_t \quad (34)$$

Where K_d is the effective deoxygenation rate and K_s is the effective loss rate due to setting.

Nevertheless, most of the processes included in the mass balance equations and their parameters are extremely difficult to precisely determine or to even calculate. Attempts to develop a simple water quality model may fail to properly describe the system behavior. On the other hand, complex models may fail due to the great uncertainties originating from too many parameters and components. Therefore, the water quality model is proposed by linguistic means, where natural processes are represented by fuzzy linguistic variables. Figure 24 shows schematically the interaction among the water quality parameters and the considered natural processes.

5.6.2 Water Quality CFNN Model Architecture

The basic architecture of the CFNN model (Figure 25) is structured based on the mass balance equation of the reservoir water quality parameters, considering a completely mixed system, which can be seen in equations (30) to (34). As it can be observed in the CFNN quality model architecture in Figure 25, the right side of the mass balance equations and final storage are modeled as hidden units, where input units represent the influent components, such as the hydrological variables. Most of the input information can be easily obtained by dam operators. This can be an incentive for the practical use of such models in dealing with problems of water quality optimization such as in Chaves *et al.* (2004).

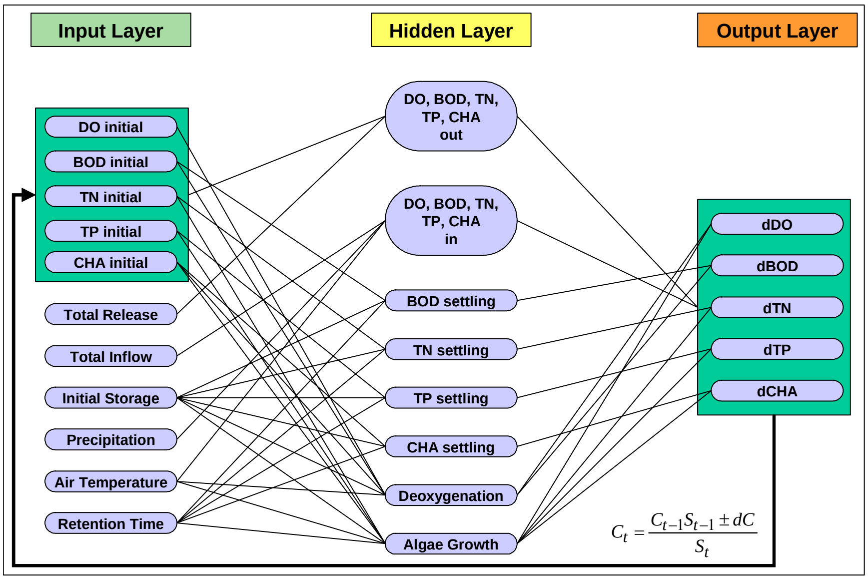


Figure 25. Architecture of the reservoir water quality CFNN based model.

Based on equations (30) to (34), we propose three basic components to represent the substance net loss by settling (BOD, TN, TP and CHA settling), phytoplankton net growth (Algae Growth) and Deoxygenation. Net Settlings are affected by the retention time, initial storage and initial concentrations. Algae Growth is related to nutrients in the water, actual CHA, retention time and initial storage. Finally, Deoxygenation is calculated based on the actual DO, BOD and CHA concentrations and air temperature.

The hydrological and climatological data used as input in the water quality model are presented in Figure 26. The input variables and some more information on their characteristics are provided below:

- Storage is a important factor for the solubility of organic matters and nutrients.
- Inflow and precipitation has great influence on nutrient and organic matter loads, where frequently, increase of inflow results in greater load values. Even though loads are responsible for most of the pollution problems in the reservoir water quality, their accurate estimation is a difficult task and is normally full of uncertainties.
- Release and previous concentrations contribute to the release from the reservoirs of nutrients, organic matter and algae (chlorophyll-a) from the reservoir water body.
- Retention time (RT) is defined as the relationship between storage volume and affluent discharge. Systems presenting higher retention time tend to have more deposition of sediment and organic matter; lower suspension from the bottom; and higher algae blooms.
- Air temperature influences the algae growth component and the DO saturation.

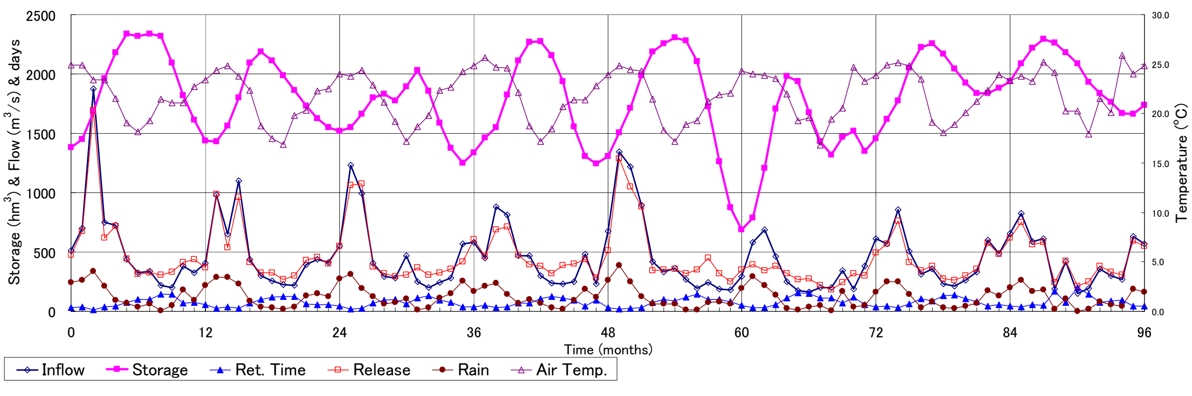


Figure 26. Climatological and hydrological input data.

Understanding the hidden neurons of a CFNN model makes it possible to define each connection type based on the accumulated knowledge about the system. Below, a brief explanation is given on how connection type has been defined for the quality model. Nevertheless, these are just unilateral relations and do not mean that output values should behave completely in such a way.

- *Input loads:* are positively (*connection type +1*) influenced by inflow volumes and precipitation, except for CHA, which increases with temperature, and DO, which tends to decrease with increase of precipitation (*connection type -1*) and decrease with increase of temperature– in Figure 27 lines 1 to 5.

- *Released mass:* representing the third term of the right side of the mass balance equations, is positively influenced by the released water and the initial concentration in the reservoir – in Figure 27 lines 6 to 10.
- *Algae growth*: tends to be positively influenced by nutrients and retention time. However, for most of the algae species there is an optimal growth temperature, for which smaller or higher temperature will present lower growth rates, therefore, for the temperature the connection is defined as a convex connection type (*+2*). Besides, the growth rate depends on the actual mass of chlorophyll within the reservoir, therefore CHA and initial storage are also considered (*type +1*) – in Figure 27 line 11.
- *Net Settling*: is basically influenced by the retention time. Higher retention time tends to reduce suspension from the bottom of the reservoir, and consequently results in nutrient and organic suspension into the water column. Moreover, the settling rate will also depend on the actual mass in the reservoir (*type +1*) – in Figure 27 lines 12 to 15.
- *Deoxygenation*: naturally the increase of BOD increases the oxygen consumption; also, the death of algae tends to increase organic matters (BOD). Moreover, higher temperatures decrease the oxygen saturation rate, here represented as increase function of consumed DO and lower levels of DO tend to decrease the deoxygenation process – in Figure 27 line 16.

The final output layer gives the actual variation of mass of a certain parameter, represented by *d* followed by the symbol of the respective quality parameter, as in Figure 28. Then, with little modification in the mass balance equation the final concentration values can be calculated as in equation (35):

$$C_{t+1} = \frac{C_t S_t \pm dM}{S_{t+1}} \tag{35}$$

here *dM* is the mass variation of a certain quality parameter, which has resulted from the balance of input loads, released mass and other processes such as settling and algae growth. The other symbols are as previously defined. The use of this arrangement of the mass balance equation showed better results than the direct use of equation (29), as this form is closer to the actual physical processes of mass conservation and mass dilution.

Hence, the connections between hidden and output layers are straightforward as they are based on the mass balance equation. The increase of the *settling* hidden unit is associated with the decrease of final concentrations of BOD, TN and TP, independently. Based on the previous explanation and evaluating the mass balance equations, the final transfer matrix for the reservoir water quality model can be proposed as shown in Figure 27 and Figure 28.

$C_{in,hid}$		INPUT										
		DO initial	BOD initial	TN initial	TP initial	CHA initial	Tair	Qtot	Rel	Rtim	So	Plu
HIDDEN	DO in	0	0	0	0	0	−1	1	0	0	0	−1
	BOD in	0	0	0	0	0	0	1	0	0	0	1
	TN in	0	0	0	0	0	0	1	0	0	0	1
	TP in	0	0	0	0	0	0	1	0	0	0	1
	CHA in	0	0	0	0	0	1	1	0	0	0	0
	DO out	1	0	0	0	0	0	0	1	0	0	0
	BOD out	0	1	0	0	0	0	0	1	0	0	0
	TN out	0	0	1	0	0	0	0	1	0	0	0
	TP out	0	0	0	1	0	0	0	1	0	0	0
	CHA out	0	0	0	0	1	0	0	1	0	0	0
	GROW	0	0	1	1	1	2	0	0	1	1	0
	BOD set	0	1	0	0	0	0	0	0	1	1	0
	TN set	0	0	1	0	0	0	0	0	1	1	0
	TP set	0	0	0	1	0	0	0	0	1	1	0
	CHA set	0	0	0	0	1	0	0	0	1	1	0
	DEOXY	−1	1	0	0	1	1	0	0	0	1	0

Figure 27. Transfer Matrix for the CFNN quality model: from input to hidden layers

$C_{hid,out}$		HIDDEN															
		DO in	BOD in	TN in	TP in	CHA in	DO out	BOD out	TN out	TP out	CHA out	GROW	BOD set	TN set	TP set	CHA set	DEOXY
OUTPUT	dDO	1	0	0	0	0	−1	0	0	0	0	1	0	0	0	0	−1
	dBOD	0	1	0	0	0	0	−1	0	0	0	0	−1	0	0	0	−1
	dTN	0	0	1	0	0	0	0	−1	0	0	−1	0	−1	0	0	0
	dTP	0	0	0	1	0	0	0	0	−1	0	−1	0	0	−1	0	0
	dCHA	0	0	0	0	1	0	0	0	0	−1	1	0	0	0	−1	0

Figure 28. Transfer Matrix for the CFNN quality model: from hidden to output layers

For this water quality model, a fuzzy inference system is constructed with five linguistic variables. The assumed values for the center of the linguistic variables of input layers are shown in Table 6 and they were defined based on the method presented in Table 6, line 1. Center values for the hidden and output layers are found automatically during the training process.

Table 6. Center values assumed for the membership functions of input variables

INPUT LAYER	Unit	1	2	3	4	5
		XL	VL	M	H	XH
DO initial	mg/l	0	3	6	8	10
BOD initial	mg/l	0	2	4	10	15
TN initial	mg/l	0	2	3	9	15
TP initial	mg/l	0	0.04	0.08	0.20	0.32
CHA initial	10^{-3} mg/l	0	12	24	62	100
Air Temperature	°C	15	18	20	23	26
Total inflow	m^3/s	0	200	400	1350	2300
Release	m^3/s	0	200	400	1350	2300
Retention Time	days	0	50	100	250	400
Initial Storage	hm^3	529	1465	2400	2825	3250
Precipitation	mm	0	60	120	310	500

5.7 Results

Figure 29 shows the evolution of the final performance function (28) used for training the CFNN model. It presents a fast increase (training) for the first generations, followed by slower improvements throughout the latter ones, which is usually observed in GA optimization. The results are shown in a 10-generations time-steps.

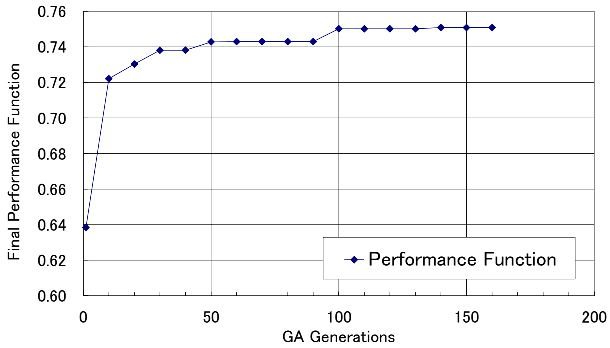


Figure 29. Evolution of the GA final performance function

Figure 30 to Figure 34 display the final results of training and validation by comparing the observed values, for all five simulated water quality parameters, with the respective credibility interval.

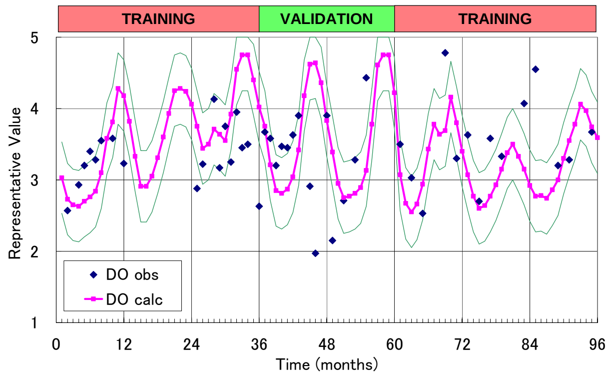


Figure 30. Results of DO for training and validation

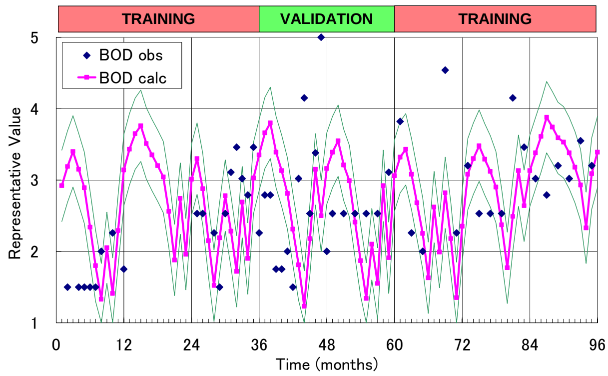


Figure 31. Results of BOD for training and validation

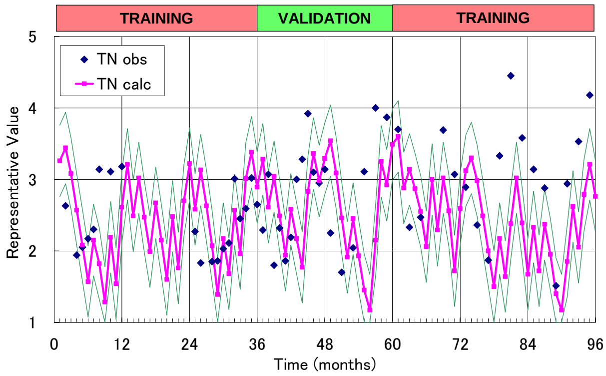


Figure 32. Results of TN for training and validation

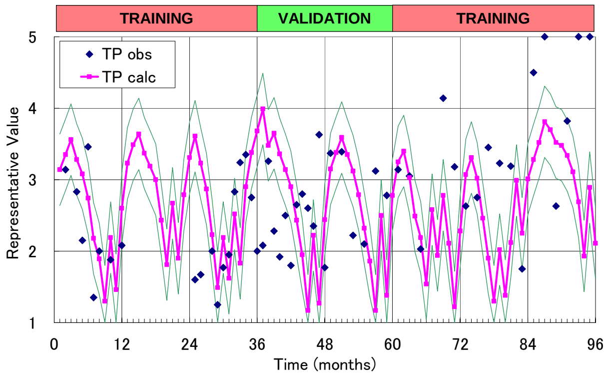


Figure 33. Results of TP for training and validation

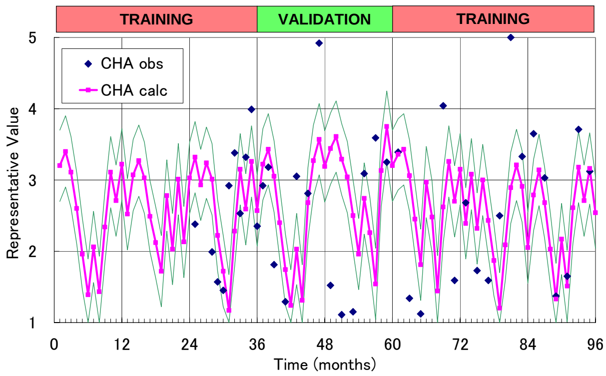


Figure 34. Results of CHA for training and validation

To demonstrate that the hidden nodes are representing the denominated variables and processes, the following charts show results from hidden units. Even though load data could have been used in the training process, it was not used here as net input information, but instead used for verification of the hidden fuzzy nodes of the CFNN model, representing the second term of the right side of the mass balance equations. Moreover, load data may not be available for future simulations, as they are often expensive and time-consuming to obtain. Figure 35 to Figure 37 show the fuzzy values of the hidden units, in contrast to the ones obtained using the observed data for loads of BOD, TN and TP, respectively. It is important to note that one is not interested in the absolute fuzzy values themselves, but more in the variation of these fuzzy values which, as can be seen, are correlated to the observed expected values.

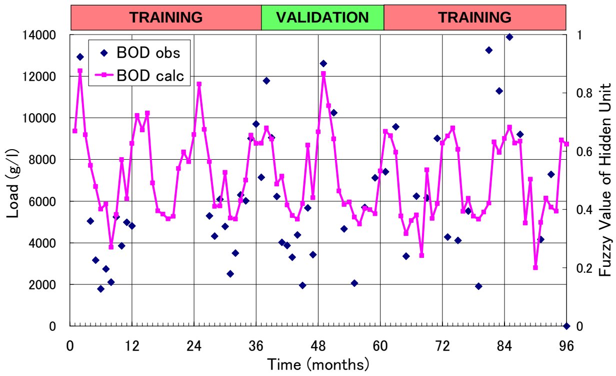


Figure 35. Results of hidden unit "BOD input" and the observed calculated values

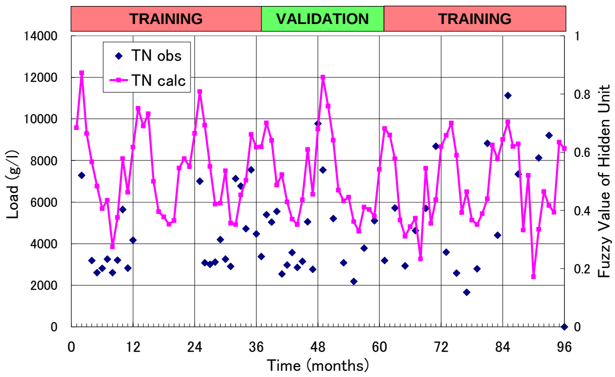


Figure 36. Results of hidden unit "TN input" and the observed calculated values

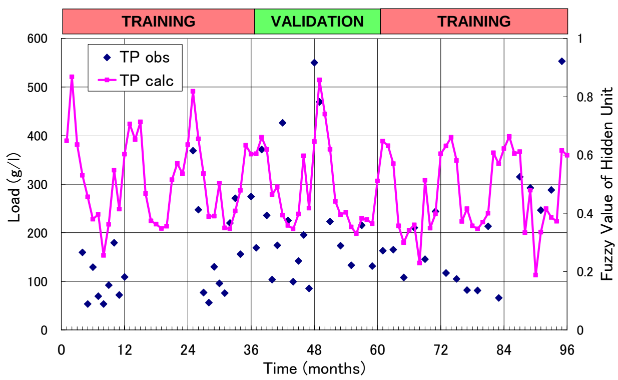


Figure 37. Results of hidden unit "TP input"and the observed calculated values

The next results are intended to show the response of the model to the *a priori* given knowledge. Figure 38 shows a plot of BOD concentrations in the reservoir versus BOD loads. Naturally, an increase in BOD loads must result in an increase of BOD concentration. However, the available data do not reflect this assumption clearly. This is probably because there is little available historical data and the observations are one or two months apart. Therefore, it can be assumed that some observations may not always represent what is really happening within the water body system in a more general way, as they may not express the daily changes or the spatial quality distribution, which have already been acknowledged by experts.

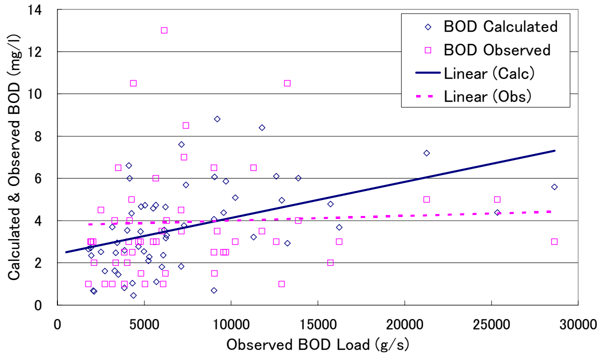


Figure 38. BOD concentration (observed and calculated values) x BOD load (observed)

The same idea can be applied to DO, as an increase in BOD concentrations should decrease DO, as shown in Figure 39. Therefore, the results are found to be representing the system as the results reflect the given knowledge. A traditional ANN model would tend to mimic the available data. However, as the CFNN model would not only learn from available data but also consider the provided knowledge, the CFNN gives what can be assumed to be more reliable results.

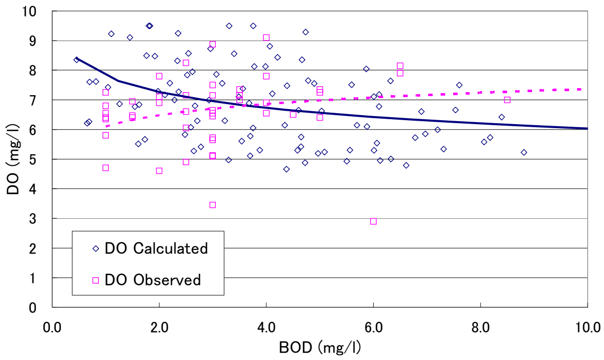


Figure 39. DO x BOD concentrations within the reservoir

In Figure 40, the "DO consumption" hidden node is plotted together with previous volumes of BOD present within the reservoir water body. High previous BOD tends to result in high oxygen consumption.

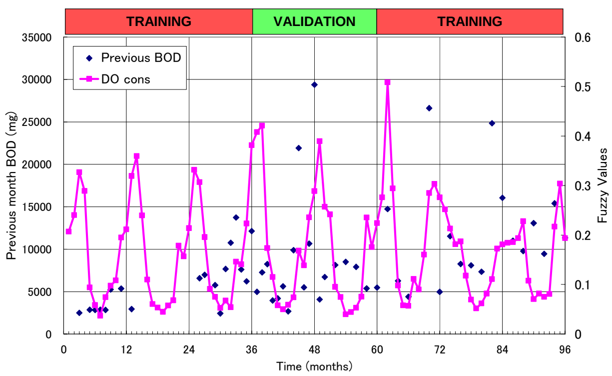


Figure 40. Results of the hidden unit "DO consumption".

Higher levels of algae growth are expected to occur when temperature is around the optimal growth temperature value. Figure 41 shows the results of the hidden unit representing this phenomenon. When plotted against temperature, it is also in accordance with the previous accumulated knowledge given to the model.

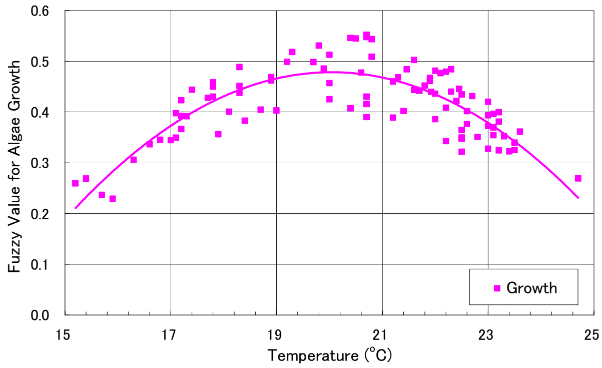


Figure 41. Results of the hidden unit "growth".

It could be seen that the hidden units are in agreement with the observed values and in fact represent the expected variations of the specified intermediate variables and processes. This is a valuable confirmation that the information given to the model through the definition of its connections is indeed taking effect.

5.8 Conclusions

CFNN has proved to be a successful example of human-machine collaboration, where the problem of data scarcity and vagueness is overcome by transferring some of the accumulated qualitative human-knowledge about the processes involved in the model.

The CFNN model can be an important tool for simulation and prediction of a system where information is considered to be vague or uncertain, and data are incomplete. The definition of connections based on accumulated knowledge was found to be an effective method to transfer the knowledge of experts to the model, thereby guiding the ANN to work not only with provided data but also with consideration of previous experiences and information, and therefore increasing its robustness. This can be said as CFNN did not show problems of over-fitting or over-training the model, a quite common problem relating to traditional ANN when training is carried out with scarce and vague data.

The CFNN model may be used when the application of the physical model is not possible, such as when there is a lack of data or when the model is too complex. Nonetheless, the CFNN can be applied when traditional ANN models are not recommended; when there are small data sets or when the data do not fully represent the system. The CFNN model can truly be seen as a “gray-box” model.

The definition of hidden neurons is not a problem for CFNN, as this is defined based on the physical characteristics of the modeled process and accumulated knowledge. Moreover, even though with fewer connections than an ANNs model with the same number of layers and units, CFNN was shown to remain flexible enough, keeping the capability of generalization which is one of the greatest advantages of ANNs based models.

6 Operation for Quantity Objectives

6.1 Introduction

One of the most important and applied techniques for the optimization of reservoirs is surely dynamic programming (DP), which was developed by Richard Bellman (1959). Applications of dynamic programming to water resources systems can be found in many works, such as Yakowitz (1982) and Esogbue (1989). DP presents various advantages over other methods to handle water resources management and optimization problems and it can be associated with other programming methods, such as stochastic techniques resulting in the so called stochastic DP (SDP). Some applications of SDP for water resources management is found in Torabi *et al.* (1973) and Kelman *et al.* (1990), where river discharge is treated as the stochastic variable. Moreover, DP may also be used for the optimization of multistage fuzzy control systems, called fuzzy DP (Fontane *et al.* 1997, Kojiri *et al.* 1993 and Hori *et al.* 1997), in which decision and state variables as well as constraints, can be set as fuzzy membership functions. The basic characteristics of water resources and reservoir operation that lead to the use of DP are: stage-wise structure and non-linearity of the system.

DP can be divided into two different approaches as continuous or discrete, with the latter the most commonly used for computational convenience. However, for reservoir operation, the rough discretization of state variables, such as storage levels in the case of reservoir optimization, may reduce the efficiency of the optimization model. Another important characteristic of DP is the direction in which the calculation is carried out, defined as forward-moving and backward-moving. For deterministic DP (DDP), calculation in either direction can be easily implemented. However, the application of forward-moving calculation for SDP models is not straightforward and may be extremely complex to model (Chaves *et al*, 2003b). The direction in which the calculation is executed becomes extremely important when DP based optimization models have to be combined or embedded with other models, such as rainfall-runoff and water quality simulation models (Chaves, 2002 and Chaves *et al*, 2003b), which require a time-wise calculation as actual values depend on previous ones. Moreover, the most important drawback of using DP is certainly the curse of dimensionality, caused by the large number of state and decision variables.

To overcome these problems related to SDP, a new optimization method is proposed here to handle the discretization of state variables through fuzzy theory and the limitations of the direction of calculation through a neuro-fuzzy system. The new proposed method is also able to handle the stochastic characteristics of variables, such as river discharge, in a forward-moving calculation direction. Moreover, the problem resulting from the increase in the number of state variables is easily overcome, as new state variables can simply be introduced as new neuron units. The proposed method is named stochastic fuzzy neural network (SFNN). For validation of the proposed method, a real storage reservoir optimization problem considering the stochastic characteristics of inflow discharge information is presented. Then, results using the SFNN are compared to the ones obtained by other dynamic programming formulations.

6.2 Methodology

There are few applications of ANN models for optimization problems, whereas the ANNs are usually applied as simulation or prediction models. It was only recently that some attempts to solve the storage reservoir optimization using ANN have appeared. Raman and Chandramouli (1996) and Chandramouli and Raman (2001) proposed the use of ANN to generate operating rules based on the optimal results from a deterministic DP model, for the case of a single and multiple reservoir system, respectively. In a similar way, Cancellier *et al.* (2002) developed a model to derive the operating rules for an irrigation supply reservoir. In an attempt to consider the stochastic characteristics of inflow, Ponnambalam *et al.* (2003) trained an adaptive neuro-fuzzy inference system (ANFIS; Jang, 1993) based on the optimal results obtained by a stochastic optimization model. Chang *et al.* (2001) combines genetic algorithm (GA) and ANFIS, in which GA is applied to search the optimal reservoir operating histogram, which is used as the training pattern of the ANFIS model, intended to estimate the optimal water release with current storage level and inflow conditions as input information.

There are some important characteristics of the proposed method, which differ from other applications of neuro-fuzzy systems for reservoir operation. First, most of the ANN based models has the ANN trained based on already-optimized results from other models to derived operating rules. However, for the proposed SFNN model, the optimization is carried out directly considering the objective functions. The second and most important point is that the SFNN is trained directly under stochastic input conditions, presenting a probability of occurrence depending on the known previous value, defined as conditional probability. The most common approach to define the conditional probability is the Markov chain technique (Howard, 1960) which considers one previous stage of the conditional probability. The Markov chain is applicable for continuous processes or, for computational convenience, discrete processes. Some applications of optimization problems using the Markov chain technique can be found in Butcher (1971) and Yakowitz (1982).

The inflow into the reservoir can be considered as a Markov chain process, where the system state, within a certain stage, is considered to be dependent on the state of the previous stage and the known probabilities. Correlograms were constructed to evaluate the correlation of inflows. Figure 42 shows the correlation coefficient for the inflow into the reservoir for two months, representing the wet season (January) and the dry season (June). The months of the wet season seem to be events with short memory. On the other hand, the dry season months can be considered as middle memory, which is expected as base flow plays an important role in the stabilization of river flow. Nevertheless, it is assumed here that the one-previous-period Markov chain is effective to represent the conditional probability between every two consecutive months.

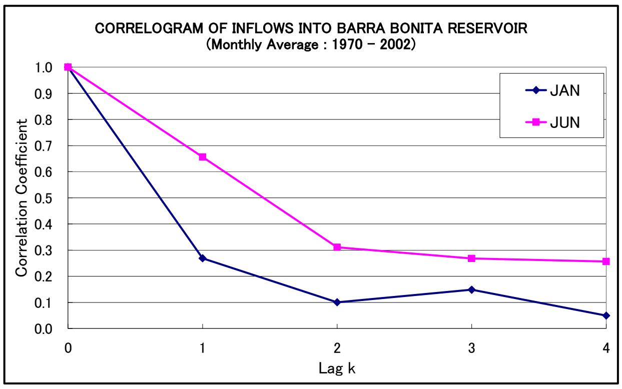


Figure 42: Correlogram for the inflow discharge

Depending on the available data and the number of assumed intervals to pursue the calculation of the Markov chain or one previous stage conditional probability (CP), the resulting probabilistic curves may present unrealistic characteristics due to incomplete data set. The use of such results in stochastic models may greatly affect its efficiency. To reduce the uncertainties of the discretization process when calculating the conditional probabilities, due to incomplete data set and rough assumption of crisp discrete intervals, the introduction of conditional probability of a fuzzy event is proposed for the calculation of the conditional probabilities of the representative inflow discharges between two consecutive months.

6.3 Conditional Probability of a Fuzzy Event

For the (nonfuzzy) conditional probability of a fuzzy event, the difference is that instead of crisp discrete intervals, the domain is divided into fuzzy linguistic variables as can be seen in the lower part of Figure 43, where L, M and H stand for *low*, *medium* and *high*, V and X for *very* and *extreme*, respectively, with the center values of each linguistic variable described in the axis of the bottom part of Figure 43.

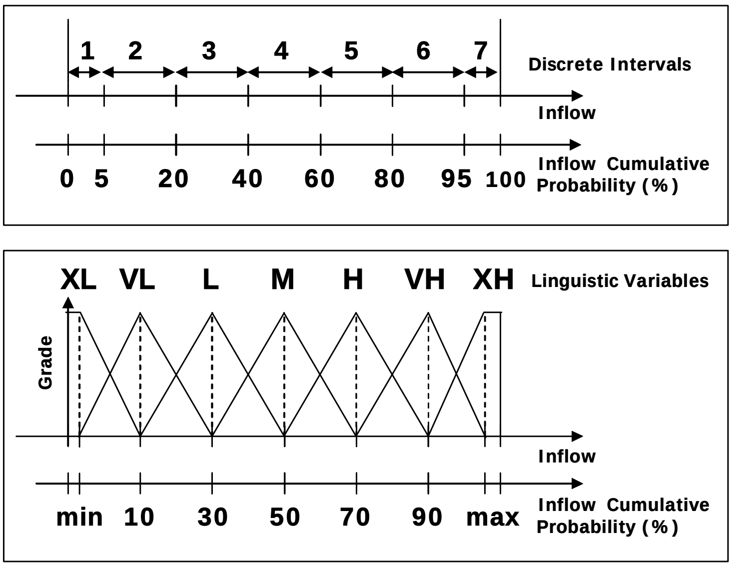


Figure 43. Discretization of discharge: discrete (upper) and fuzzy (lower) intervals

The percentage scale in Figure 43 represents the chosen limits for the discrete intervals and center values of linguistic variables, corresponding to the monthly inflow (unconditioned) cumulative probabilities. This increases the efficiency of the stochastic optimization model as already carried out by Chaves (2002), reducing to a certain extent null-probability values, as a result of many discretized interval or small time-series data. The actual limits and representative stochastic inflow values used for the calculation of both conditional probabilities are shown in Table 7. The numbers in bold represent the representative stochastic inflow values for a certain interval, and in the fuzzy case, the modal value of the membership functions.

Table 7. Cumulative (unconditioned) probabilities of monthly inflows

CDF	**Cumulative Probability and respective Monthly Inflow Values (m^3/s) for Barra Bonita Reservoir**											
	JAN	**FEB**	**MAR**	**APR**	**MAY**	**JUN**	**JUL**	**AUG**	**SEP**	**OCT**	**NOV**	**DEC**
Min	**182**	**87**	**179**	**118**	**115**	**88**	**79**	**75**	**56**	**91**	**106**	**129**
0.05	233	173	193	123	127	96	82	76	57	93	109	167
0.10	**337**	**264**	**226**	**136**	**140**	**112**	**92**	**78**	**75**	**98**	**119**	**224**
0.20	407	309	296	162	159	139	129	123	104	136	148	249
0.30	**426**	**370**	**354**	**236**	**170**	**168**	**153**	**135**	**117**	**162**	**184**	**326**
0.40	517	468	403	278	211	175	170	150	149	188	212	373
0.50	**530**	**598**	**467**	**311**	**224**	**201**	**189**	**175**	**181**	**195**	**234**	**402**
0.60	628	668	537	342	241	223	214	203	202	212	265	426
0.70	**745**	**911**	**684**	**372**	**294**	**316**	**254**	**220**	**224**	**260**	**339**	**502**
0.80	848	1008	733	481	444	483	288	239	243	467	406	608
0.90	**1087**	**1214**	**1009**	**712**	**564**	**711**	**530**	**455**	**484**	**685**	**504**	**781**
0.95	1187	1435	1184	955	760	1497	718	533	808	749	642	956
Max	**1341**	**1872**	**1193**	**1032**	**920**	**2269**	**760**	**551**	**941**	**795**	**738**	**1106**

Note: Numbers in bold are the stochastic inflow ($I_{t,k}$) values used in the proposed optimization schemes

Conditional probability is traditionally defined as the probability of an event *A* under the condition that an event *B* occurs. This probability is denoted by *P(A/B)* and defined as shown in equation (36).

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{P(A \cap B)}{P(B)} \qquad (36)$$

for *P(B)* > 0 and where *P(AB)* is the intersection of events *A* and *B*, and *P(B)* is the probability of event *B*. As stated by Kaufman (1975) and Kacprzyk (1997), the above concept is valid for both cases of nonfuzzy and fuzzy events.

For the case of the conditional probability of inflow discharge, we are interested to find the probability of inflow *t* when inflow *t-1* occurs, which would correspond to events *A* and *B*, respectively. The *P(AB)* can be calculated based on the observed time series data. For the case of discrete intervals, the frequency of occurrence *F* can be found through equation (37) as shown below:

$$\text{if and only if } (I_t \in A_k) \text{ when } (I_{t-1} \in B_j) \text{ then :}$$

$$F(A_k / B_j) = \sum_{n=1; c_0=0}^{\mathrm{N}} c_{n-1} + 1 \qquad (37)$$

Where c_n is a counter operator for which the term +1 accounts for each occurrence between consecutive events (intervals) A_k and B_j, whereas *k* and *j* are indices representing the intervals [1,2,…7] in the upper part of Figure 43. Note that the summation of all $F(A_k/B_j)$ is equal to the total number of pairs of observations *N*.

$$P(A_k \cap B_j) = \frac{F(A_k / B_j)}{N} \qquad (38)$$

Finally, as the probability $P(B_j)$ of the inflow event B_j corresponding to interval *j* can easily be found by dividing the frequency of inflows contained within interval *j* by the total number of observations *N*. Then, it is possible to obtain the final one-previous-stage or Markov chain conditional probability using equation (39):

$$P(A_k / B_j) = \frac{P(A_k \cap B_j)}{P(B_j)} \qquad (39)$$

For the case using the fuzzy method, the discharge values are fuzzified using the membership functions as shown in the lower part of Figure 43. Then, also based on the observed data, frequency of conditioned occurrence of a fuzzy event, hereafter denoted by $F\mu(A_k/B_j)$ can be analogously calculated for each linguistic variable using equation (40):

$$F\mu(A_k / B_j) = \sum_{n=1}^{N} \mu_{A,k}(I_{t,n}) \cdot \mu_{B,j}(I_{t-1,n}) \qquad (40)$$

Where *A* and *B* are now fuzzy consecutive events, for example inflows of January and February; $\mu_{A,k}$ and $\mu_{B,j}$ are the fuzzy membership function of these events for the linguistic value *k* and *j*, respectively, *k* and *j*: [XL,VL…XH], for each pair *n* of observed data (for example, for every year). The membership function here represents the degree to which an observed value corresponds to a certain class or linguistic variable. For example, if both consecutive values are equal 1, this would indicate that the observed values are equal to the representative inflows and equation (40) would yield the same results of frequency for the nonfuzzy method, equation (37). Values different than 1, could be roughly said to represent a certain degree of "frequency" for two consecutive observed values. Note that the summation of all $F\mu(A_k/B_j)$ for k and j, is equal to total number of pair of observations N, analogous to the nonfuzzy counterpart method. Hence, the $P\mu\,(AB)$ of the fuzzy events can be then calculated using equation (41):

$$P\mu(A_k \cap B_j) = \frac{F\mu(A_k / B_j)}{N} \qquad (41)$$

The (nonfuzzy) probability of a fuzzy event A is denoted by $P\mu\,(A)$ and is defined as in equation (42), as described by Kaufman (1975) and Kacprzyk (1997), based on the classic definition of Zadeh (1968), which is by far the most popular and widely used.

$$P\mu(B) = \sum_{n=1}^{N} \mu_B(x_n) \cdot p(x_n) \qquad (42)$$

Where x is the actual value of the variable in question for occurrence n. Hence, for our case of inflow discharge:

$$P\mu(B_j) = \sum_{n=1}^{N} \mu_{B,j}(I_{t-1,n}) \cdot p(I_{t-1,n}) \qquad (43)$$

Where $p(I_{t-1,n})$ is the probability of a certain inflow I_{t-1} and $\mu_{B,j}(I_{t-1,n})$ is the value of the fuzzy membership function for the inflow corresponding to observation n for event (linguistic variable) j. That is to say, $P\mu\,(B_j)$ is the probability of a fuzzy event B to occur. Note that the summation of $p(I_{t-1,n})$ is equal to 1 and (in our case) for each pair of observation n, the probability of a certain observed inflow presents the same value $1/N$. Finally the conditional probability of a fuzzy event A_k, $P\mu(A_k/B_j)$, is calculated according to equation (44).

$$P\mu(A_k / B_j) = \frac{P\mu(A_k \cap B_j)}{P\mu(B_j)} \qquad (44)$$

The summation of $P\mu(A_k/B_j)$ in respect to j is also equal to 1 (one). Note that the summation is equal to 1 only if the linguistic variables are arranged in a way such as illustrated in Figure 43, where the minimum values of membership functions for each class coincide with the center values of adjacent classes. In any other case, for example using a different membership function, the probability of each month has to be compensated by the total summation for all classes.

It is important to mention that the proposed method to define the Markov chain conditional probabilities is not dealing with the vagueness and uncertainties related to the probabilities themselves, but only the uncertainties originating from the discretization process and the fuzziness of discharge observations. For more sophisticated analyses to handle such probabilistic problems, the readers are referred to other methods such as possibility theory, Dempster-Shafer theory of evidence and fuzzy measures.

6.4 Stochastic Fuzzy Neural Network (SFNN)

Figure 44 shows the flowchart of the overall calculation procedures for the SFNN optimization model where inflow values are considered as stochastic variables based on conditional probabilities. According to Figure 44, first the SFNN model is initiated with a set of parameters, which can be randomly created.

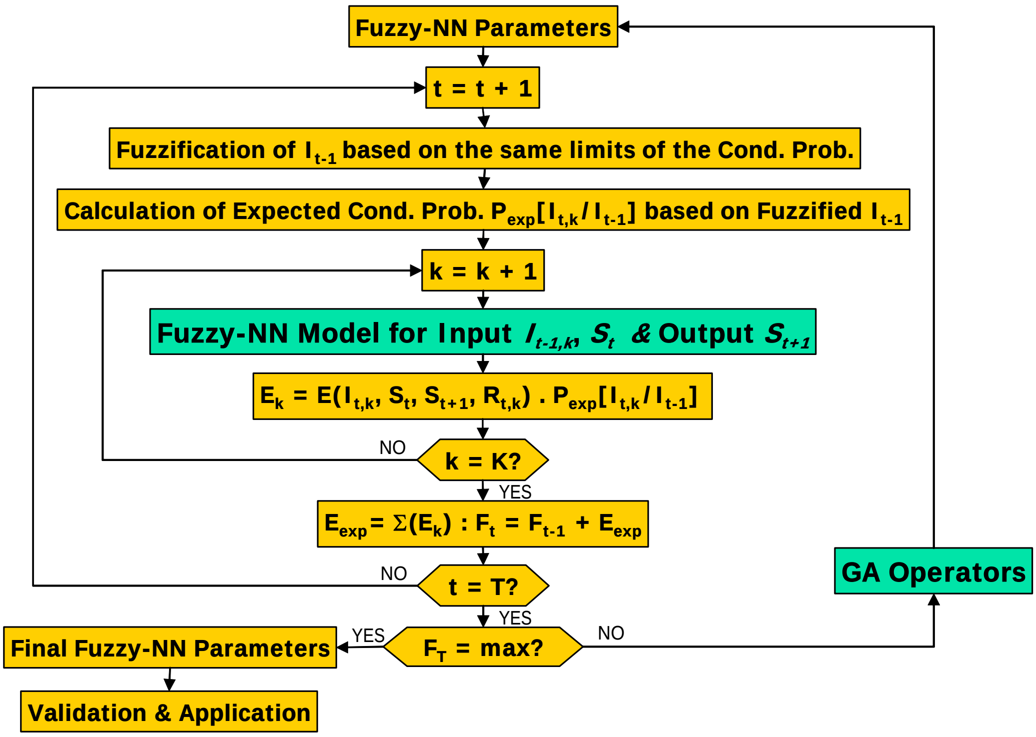


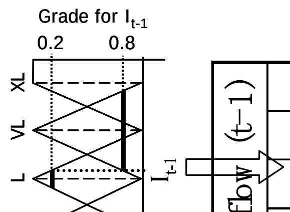
Figure 44. Flowchart of the calculation procedures for the SFNN

The SFNN can be applied to both types of conditional probabilities presented in the previous section. The values found through the traditional conditional probability, which uses the crisp intervals, can be directly used for the SFNN optimization model or they may be interpolated according to the actual previous inflow and the representative inflows. As for the CP of fuzzy events, the fuzzy representative inflows may be used. However, in doing so, the optimization model may result in optimum fuzzy values of release and end-of-period storage that are too difficult for users to understand and too vague for their practical consideration. The propagation of the fuzziness associated with these values, when applied to the optimization model, would tend to generate greater vagueness, resulting in less efficient operation rules. Therefore, a method that could be considered analogous to interpolation is proposed, in which only the center values are considered and according to the actual previous inflow, the conditional probability values is changed.

As the previous inflow value is known, based on the already calculated conditional probabilities of a fuzzy event, the new expected value can be obtained for each of the representative stochastic inflow $I_{t,k}$, which is also defined as the center value of the fuzzy event k with actual values presented previously in Table 7. The expected conditional probability for the SFNN model is calculated based on equation (45). The center values of the fuzzy membership function for the fuzzification of the previous inflow are the same as the ones used for the calculation of the conditional probabilities themselves, having the center and limit values defined as the (unconditioned) probability for each month as shown in the lower part of Figure 43 and the actual values are presented in Table 7.

$$P_{\exp}(I_{t,k} / I_{t-1}) = \sum_{j=1}^{J} \left[P\mu(A_k / B_j) \cdot \mu(I_{t-1}) \right] \qquad (45)$$

Where $P\mu$ is the conditional probability of a fuzzy event, μ is the membership function or fuzzy grade value of previous inflow I_{t-1} and A and B two consecutive fuzzy events. An example of the calculation of the expected probabilities is shown below in Figure 45, where values in the center table represent the conditional probabilities, calculated as in the previous section, between two consecutive fuzzy events, i.e. previous and actual inflows. The stochastic fuzzy neural network (SFNN) is then trained considering the representative inflow values $I_{t,k}$ and the respective expected conditional probabilities $P_{exp}(I_{t,k}/I_{t-1})$.



		Actual Inflow (t)							
		XL	VL	L	M	H	VH	XH	Sum
Previous Inflow (t−1)	XL	0.35	0.39	0.26	0.00	0.00	0.00	0.00	1.00
	VL	0.32	0.37	0.25	0.06	0.00	0.00	0.00	1.00
	L	0.00	0.22	0.32	0.38	0.07	0.01	0.00	1.00
	M	0.01	0.03	0.02	0.38	0.42	0.12	0.01	1.00
	H	0.03	0.15	0.25	0.17	0.21	0.12	0.06	1.00
	VH	0.00	0.11	0.26	0.16	0.21	0.16	0.10	1.00
	XH	0.00	0.00	0.00	0.00	0.17	0.31	0.53	1.00
Expected Probability		0.06	0.25	0.30	0.32	0.06	0.01	0.00	1.00

Figure 45. Example for the calculation of the expected probability P_{exp}

Usually for the training of neuro-fuzzy models, the target vector is previously defined and usually accounts for observed values. However, in the case of optimization using the ANN based models; the training is proposed to be based directly on the maximization of the recursive function F itself. The following formulation are proposed for the recursive equations (46) and (47), considering two different types of fuzzy decision:

- *Weighted average type*

$$F(S_t) = \max_{S_{t+1}} \left[E_{\exp}(E_{t,k}) \cdot \alpha + F_{t-1}(S_{t-1}) \right] \quad (46)$$

- *Minimization type*

$$F(S_t) = \max_{S_{t+1}} \left[\min\left[E_{\exp}(E_{t,k}), F_{t-1}(S_{t-1}) \right] \right] \quad (47)$$

Where α is the weight, here we consider all the weights to be the same and equal to *1/T*. Nevertheless, different weights may be incorporated to increase the importance of a particular period in time or as a discount factor, where T is the total number of stages (time-steps). The summation of α is equal to 1, which guarantees the final value of the recursive function (46) to be between [0, 1]. This does not affect the final optimum storage or release sequence, but instead facilitates the comparison of the final results with other decision types, such as the minimization type (47). E_{exp} is the expected value of the evaluation function, as shown in equation (48), which depends on the expected probability.

$$E_{\exp} = \sum_{k=1}^{K} E_k \cdot P_{\exp}(I_{t,k} / I_{t-1}) \quad (48)$$

where the P_{exp} is the expected conditional probability presented in equation (45), k represents the index of conditional probabilities, referring to the representative stochastic inflows $I_{t,k}$. E_k is defined according to the kind of aggregation operator to be used, such as shown in equations (49), (50) and (51), for three aggregation operators: weighted-average, product and minimization, respectively.

$$E_k = \sum_{n=1}^{N} E_n \cdot \beta \quad \text{where} \quad \sum_{n=1}^{N} \beta = 1 \quad (49)$$

$$E_k = \left(\prod_{n=1}^{N} E_n \right)^{1/N} \quad (50)$$

$$E_k = \min_n (E_n) \quad (51)$$

Where β is the weight, here assume the same for all objectives but which could vary depending on the importance of each objective. N is the total number of fuzzy objectives, and index n represents each fuzzy objective. All of them are represented by a fuzzy objective membership functions E_n (also represented later by μ in section 6.4.2). The fuzzy objective membership functions used here are shown later in Figure 48. The calculation of the expected value E_{exp} is used in the calculation of the recursive function shown in equation (46) and (47). Optimization results using the above fuzzy decisions and aggregation operators are shown later.

Having a set of parameters for the neuro-fuzzy system, it is possible to obtain the end-of-period storage level S_{t+1}, as we know S_t and I_{t-1} values. For each stochastic inflow value $I_{t,k}$, we can obtain a release value $R_{t,k}$ based on the reservoir mass balance equation (52), which neglects other losses such as evaporation and infiltration that also could be considered.

$$R_{t,k} = S_t - S_{t-1} + I_{t,k} \qquad (52)$$

This operation is carried out for the whole training period *T*. If the parameters of the neuro-fuzzy system are considered to be optimal, giving the maximum or quasi maximum value for the recursive function then the training process can be concluded. After validation the model can be finally applied as the network is now representing the reservoir stochastic optimal operation rules.

It is important to note that in the case of SDP models all alternatives are tested for a period of one year, or the time representing one cycle of the stochastic variable, here being the monthly inflow discharges. The results of a SDP model are presented as guidecurves that can be used to obtain the end-of-period storage when the actual month, previous inflow and initial storage values are provided. On the other hand, the SFNN model is training using directly observed data, considering the conditional probability between stages and also the long-term characteristics of river discharge. In the case that observations are scarce or considered to be incomplete, the SFNN model can be trained with artificially generated random or synthetic time series data. On the other hand, only pure random values can be used, however the long-term characteristics of inflow are no longer considered, which could decrease part of the SFNN model efficiency. It is important to keep in mind that the computational effort increases with an increase in the data set size. Nevertheless, for longer training data, more efficient operation rules may be expected. The optimal results found by SFNN are no longer presented as guidecurves, but instead as a fuzzy neural network model, stochastically trained, which has the end-of-period storage represented by its output neuron, and initial storage and previous inflow represented through its input neurons.

6.4.1 Fuzzy Neural Network Model

Artificial neural network (ANN) models have been widely used in various fields, such as aerospace, finance, robotics, environmental assessment and hydrology, for different purposes such as classification, pattern identification, simulation and prediction (Hagan *et al.*, 1995). The combination of ANNs with other techniques, such as fuzzy theory, has been proposed by some authors. The use of fuzzy theory with the ANN aims to combine the ability of fuzzy sets to represent knowledge that is understandable to human beings with the learning capability of ANNs (Jin, 2003). Therefore, a neuro-fuzzy system is a fuzzy system that can learn from data as well as from experts' experience. Application of fuzzy theory with ANNs may change the model from being a "black-box" into a "gray-box" type. As described by Jin (2003), there have been some attempts to develop neuro-fuzzy models and the most well known example is the adaptive-network-based fuzzy inference system (ANFIS) proposed by Jang and Sun (1993, 1995). The ANFIS model has been successfully applied to many problems, including storage reservoir operation and it can serve as a basis for the construction of fuzzy if-then rules. However, this method can face great computational effort, as the increase in the number of parameters is proportional to its number of linguistic variables and input nodes.

Therefore, a simpler neuro-fuzzy network is proposed, whereas the basic architecture of the model is briefly presented as follows. The training process is carried out similaryl to the one used for the neuro-fuzzy system of Chapter 5, by a genetic algorithm (GA) model which has been described in 3.6.1. The difference here is that the performance functions of the former are substituted by the objective function used for the reservoir operation, such as the ones shown later in Figure 48. The developed neuro-fuzzy system architecture is shown in Figure 46.

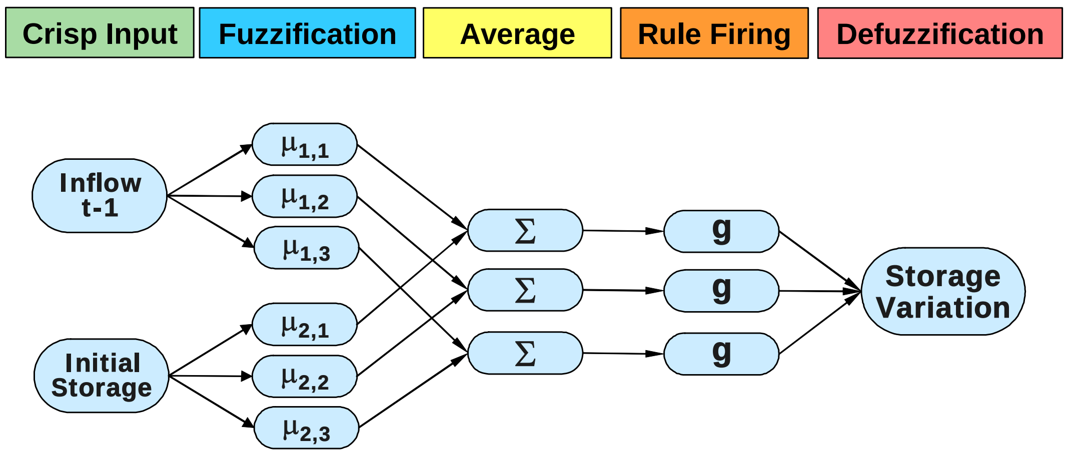


Figure 46. Architecture of the proposed neuro-fuzzy inference system

Input and Output Values

Previous inflow (*t-1*) and initial storage are represented by two input nodes, whereas the output node gives the variation of the end-of-period storage value, which is the consequence of the operating rules. Each month has its own network, which proved to significantly increase the optimization efficiency. Hence, the operation rules represented by the neuro-fuzzy system are analogous to the operation guide-curves derived by the SDP models, where by three types of information is required (actual month, previous inflow and initial storage).

Fuzzification of Inflows and Storage values

Considering a crisp input value x_i, such as previous inflow and initial storage, the fuzzification function is carried out as shown in Figure 47. The parameters X_k are defined as the modal value of the fuzzy membership function of each linguistic variable k. For the case of previous inflow (t-1), the parameters X_k, where k[1,2,…7], are defined based on the cumulative probability function for each month, accounting for the same values as the ones used for the calculation of the conditional probability of a fuzzy event, as shown in Figure 43 and Table 7 presented in the previous section. For the case of initial storage, the parameters X_k are optimized during the training process. The extreme values X_1 and X_7 are set equal to the minimum and maximum expected values, i.e. constrain values use for the reservoir operation here assumed equal to 600 hm^3 and 3250 hm^3, respectively. Seven linguistic values are used as they proved to be enough for modeling the system. Moreover, this number has also been used widely in other applications of fuzzy inference. The resulted membership values are represented by the symbol $\mu_{i,k}$, which is represented graphically in Figure 47.

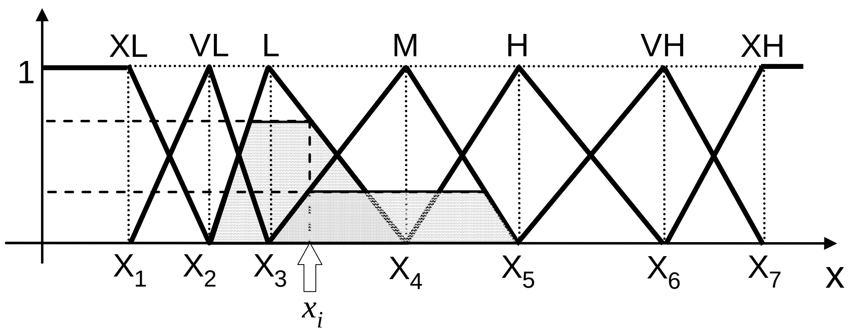


Figure 47. Membership function for the fuzzification of input variables x_i

Average

The resulted membership values $\mu_{x,k}$ for each linguistic variable k, found after the fuzzification of the crisp input value x, are then combined. There are various methods used within fuzzy inference theory, such as product, minimization and maximization. However, after trying these different methods, the weighted sum is found to be the most appropriate as robustness of the whole system increases. Probably, this is due to the use of triangular functions that present only two membership values other than zero, which in a product type aggregation could null all the membership values. The resulted fuzzy number $\mu_{y,k}$ after the combination through the transfer function of previous units i is presented in equation (53).

$$\mu_{j,k} = \frac{\sum_{i=1}^{N} (W_{ij} \cdot \mu_{i,k})}{\sum_{i=1}^{N} W_{ij}} \quad (53)$$

Where N is the total number of previous units and W_{ij} is the weight between neurons x_i and y_j, W is also optimized during training. The resulting fuzzy inference is represented by $\mu_{j,k}$.

Firing the Rule and Defuzzification

Knowing the fuzzy value $\mu_{i.k}$, it is possible to obtain a crisp value y_j , which in our case represent the variation of final (end-of-period) storage, through many available defuzzification methods. The most common defuzzification methods are the center of gravity, center of sums and center of areas (Jin, 2003). A method based on the center of gravity and described in Jin (2003) is adopted here, in equation (54).

$$y_j = \frac{\sum_{k=1}^{K} (\mu_{j,k})^g \cdot Y_k}{\sum_{k=1}^{K} (\mu_{j,k})^g} \quad (54)$$

Where Y_k is the kernel or the modal value and is identified during training; $g \in [0,\infty)$ is a power parameter responsible for "firing the rule", which can be properly adjusted to obtain a more efficient performance for the defuzzification method. This parameter is also optimized during the training period. After calculating the variation of storage ($dS = y_i$) and knowing the initial storage ($S_{initial}$), the final storage (S_{final}) can then be calculated by equation (54):

$$S_{final} = S_{initial} + dS \quad (55)$$

6.4.2 Operation Objectives

The proposed methodology is applied in the development of a simple storage reservoir optimization model considering fuzzy objectives related to water quantity characteristics of the system. These functions are defined here as flow stabilization $\mu_{stab}(R_t)$ or STAB, as a function of operated release, and hydropower generation (in KW) $\mu_{pow}(R_t, S_t)$ or POW, as a function (indirectly) of operated release and storage level. Both objectives are represented by fuzzy membership functions, defined between 0 and 1, representing the low and high satisfaction of operation, respectively. The fuzzy membership functions represent the degree of satisfaction or conformity in attaining the operation objectives. Due to a lack of realistic information about the reservoir operation, the two objectives considered here for optimization have been hypothetically formulated and their fuzzy membership functions are shown in Figure 48.

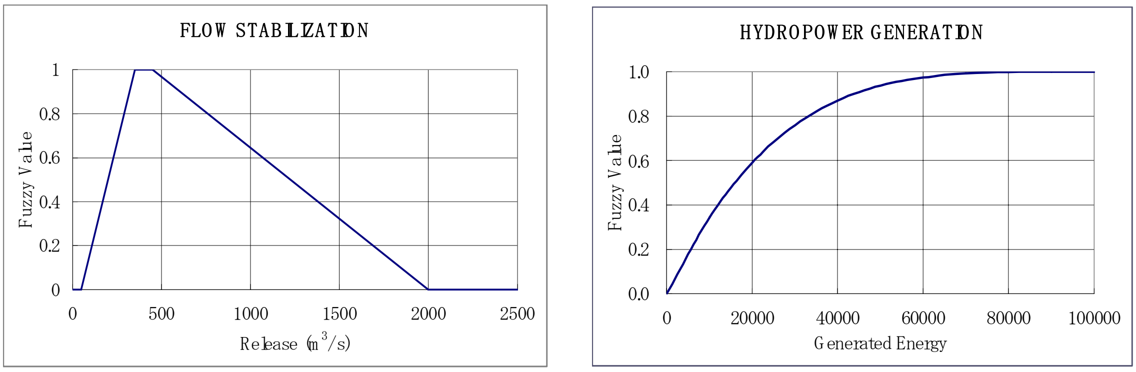


Figure 48. Fuzzy objective membership functions for the reservoir operation

The release values of 350 and 400 m^3/s are assumed to represent maximum satisfaction in the flow stabilization membership function, which is intended to stabilize the release around the historical average inflow for the reservoir. It is very important, particularly in the case of Brazil, to consider power generation in the optimization. The membership function is based on the produced and demanded energy. Due to a lack of data regarding the energy production, an average for all months is used to represent energy demand. It may be improved with realistic data and generation targets for the specific reservoir. Another way to improve the power generation analysis is by using statistical analysis to predict demand. However, for the purpose of this study, a value of 100MW is assumed as the monthly average demand for all months. The generated energy (KW) is calculated based on the release R (m^3/s) and water head H (m) by equation (56), where H is a function of the average of initial- and end-of-period storage values of stage t.

$$Generated\ Energy = 8.319 \cdot R_t \cdot H_t \qquad (56)$$

6.5 Validation of the SFNN Model by Comparison with DP

To test the proposed methodology, the Barra Bonita reservoir is optimized through five different optimization schemes, including the Stochastic and deterministic FNN method and three SDP formulations, all for monthly time-steps. The total time series is divided into two data sets, where 22 years of observations are used to identify the conditional probabilities for all models and to train the SFNN model, whereas 11 years is used to validate the proposed methodology in a practical reservoir optimization. All the optimization schemes use the same fuzzy objective membership functions presented in Figure 48, which are combined using the product aggregation method and the weighted-average type of decision, for the same length of optimization horizon. The first scheme is a simple deterministic DP model where the inflow in stage t is considered to be known. Therefore, the DDP is expected to give the maximum satisfaction for the operation objectives. The recursive equation for the DDP is presented in equation (57).

$$F_t(S_t) = \max_{S_{t+1}} \{E_t(.) + F_{t+1}(S_{t+1})\} \qquad (57)$$

Where F is the recursive equation, E is the evaluation function and S, R and I represent storage, release and inflow, respectively, for time step t.

To exemplify the importance of storage discretization within the DP optimization, some levels of storage discretization were tested for the optimization of the Barra Bonita reservoir applying the deterministic formulation, considering the whole inflow data set (33 years). The final results of the recursive equation for the whole optimization horizon (referred to as *average* in the legend of Figure 49), the minimum values among all E_t values and its standard deviation are presented in Figure 49.

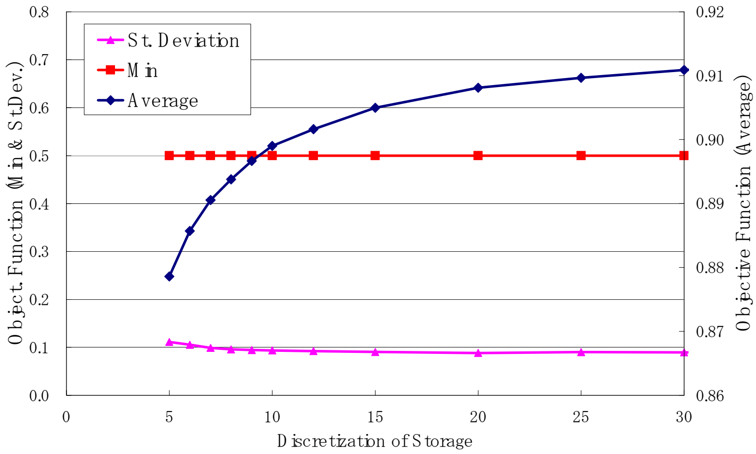


Figure 49. Relation between storage discretization and DP optimization results

Then, stochastic characteristics of inflow are considered by applying two SDP models. The first one is based on discrete conditional probability and the second uses the proposed fuzzy formulation. This is to confirm the efficiency of the proposed conditional probability of a fuzzy event. The recursive equations for the two stochastic optimization schemes using SDP are presented in equations (58) and (59), respectively.

$$F_t(S_t, I_{t-1}) = \max_{S_{t+1}} \left\{ \sum_{k=1}^{K} [E_t(\cdot) + F_{t+1}(S_{t+1}, I_t)] \cdot P(I_{t,k} / I_{t-1}) \right\} \qquad (58)$$

Where P is the conditional probability for the discrete inflow intervals k, $I_{t,k}$ is the stochastic inflow which refer to the cumulative probability values of Table 7 and $E_t(.)$ is the evaluation function.

$$F_t(S_t, I_{t-1}) = \max_{S_{t+1}} \left\{ \sum_{k=1}^{K} \left[E_t(\cdot) + F_{t+1}(S_{t+1}, I_t) \right] \cdot P\mu\left(I_{t,k} / I_{t-1} \right) \right\} \qquad (59)$$

Where $P\mu$ is the fuzzy conditional probability which has been already defined in equation (44). The state transition function has been presented in equation (52). Note that this is the inverted form of the mass balance equation (Labadie, 1993). This form is chosen for its coding simplicity. The optimization is subjected to the constraint of minimum and maximum acceptable values of storage and release. Moreover, $F_T = 0$, where T is the last chronological time, and therefore, for the backward calculation, the first stage used for calculation.

Since the transition probabilities repeat every 12 months, the undiscounted stochastic DP calculations are repeated via successive approximations to confirm and guarantee that the optimum guidecurves of end-of-period storage for each month are converging to stationary values. As a result, optimal guidecurves may be applied to each year over the entire operational horizon for any sequence of inflow. As referred by Labadie (1993), if this procedure converges, then the solution must be optimum. Finally, the reservoir is optimized through the SFNN proposed method, also considering the conditional probability of a fuzzy event. The SFNN recursive equation has already been shown in equation (46). Note that both recursive equations for the SDP models are presented as backward-moving while the SFNN can be formulated for the forward-moving.

6.6 Results

Figure 50 shows the results of cumulative conditional probability curves for the month of February carried out for the monthly inflow discharge of the Barra Bonita reservoir using the CP for discrete intervals of inflow, whereas Figure 51 shows the results after using the conditional probability of fuzzy inflow. It can be seen that the results of the fuzzy based method (Figure 51) are much smoother than the ones from the formulation based on discrete intervals. Therefore, they are much more in keeping with what is expected of realistic conditional probability curves. Moreover, results after optimization using a SDP model showed to be more efficient than the ones using the traditional method, as can be seen in Figure 52.

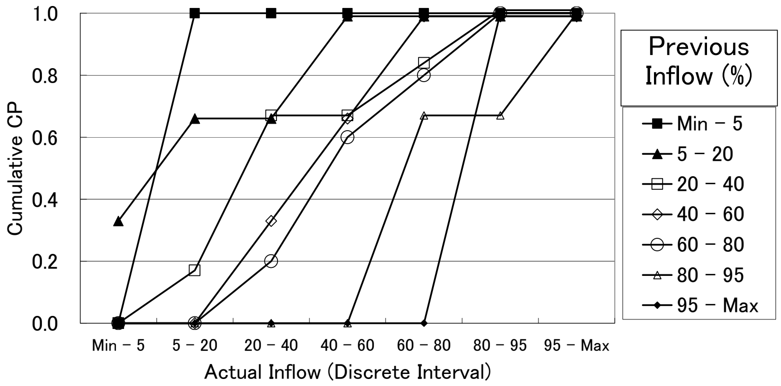


Figure 50. Cumulative CP assuming discrete intervals

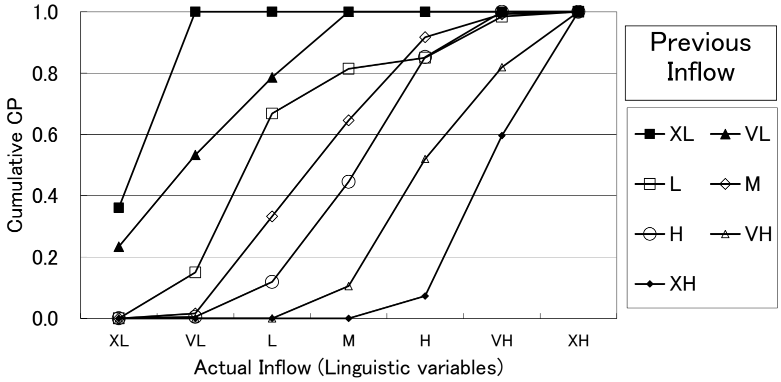


Figure 51. Cumulative CP assuming fuzzy linguistic variables

Figure 52 shows a comparison of evaluation function results after the five different optimization schemes using the weighted-average aggregation operator for the objective functions, equation (49): a deterministic DP (DDP) model (maximum expected value, for which inflows are considered to be known), a deterministic FNN model (maximum expected value, for which inflows are also considered to be known), SDP model using both method (discrete and fuzzy interval) for the calculation of conditional probabilities and the proposed SFNN model (considering the CP of a fuzzy event).

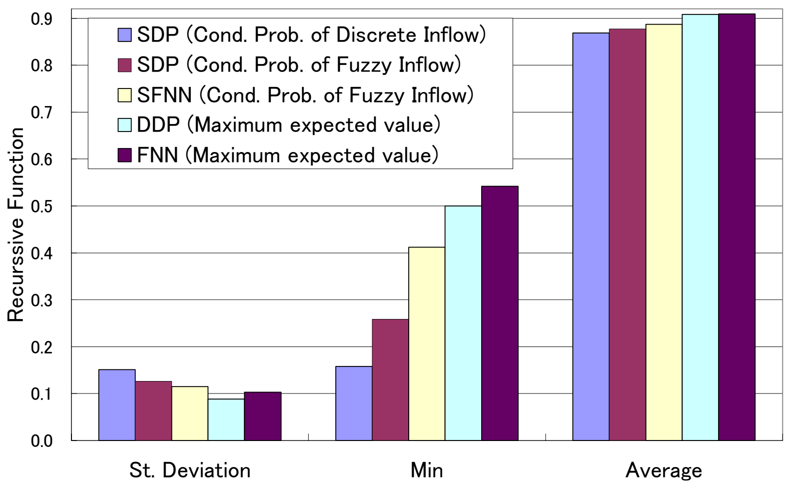


Figure 52. Results after five optmization schemes (weighted-average aggregation)

It can be seen that the SFNN model presents a mean value slightly superior to the ones found with the SDP models, showing the increase in the efficiency of optimization when the SFNN model is compared to the other two SDP formulations (as values near to 1 mean higher satisfaction and consequently more efficient operation rules). Moreover, the minimum value of the evaluation function of the SFNN is greater, and its standard deviation is lower than the SDP models, also confirming the efficiency of the proposed method. Within the two SDP models, it is possible to conclude that the use of CP of fuzzy inflow improved the operation performance, as the "average" value was higher (closer to 1) when using the SDP based on the CP pf fuzzy event.

The DDP and FNN formulations are only intended to show what would be the maximum expected results that could be obtained if future inflows were known exactly. The determinist optimization using the same FNN architecture also showed excellent results compared to its counterpart DDP. Figure 53 shows the results of useful-storage for the validation period for three (SDP with CP of fuzzy inflow, SFNN and DDP) of the optimization schemes.

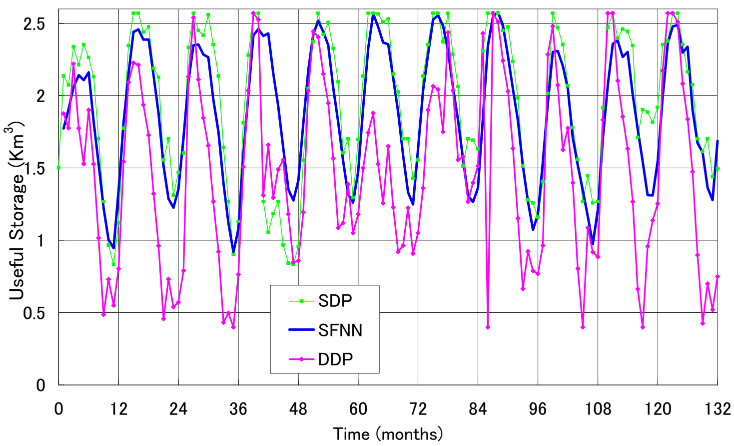


Figure 53. Resuts of optimal useful storage after three different optimization schemes

The final storage and release sequences after the SFNN optimization are shown in Figure 54. It can be seen that the proposed SFNN operation model is flexible enough due to discretization of storage levels based on fuzzy inference, which can calculate any storage level independent of limitations originating from the discretization process, usually observed in SDP models. Also in Figure 54, the training and validation periods used for all schemes are shown, which correspond to 264 and 132 months, respectively.

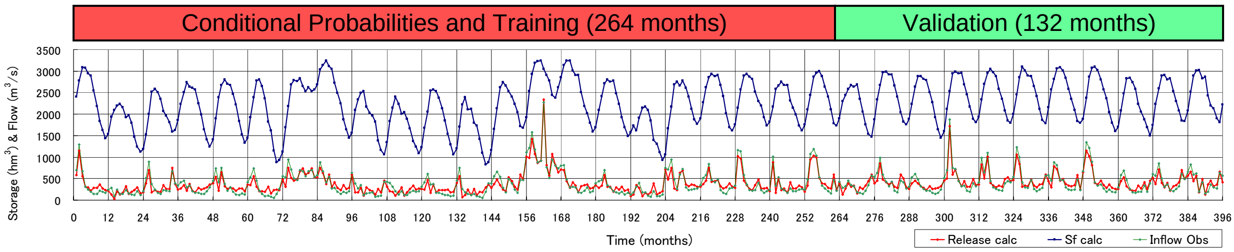


Figure 54. Results of operation using the SFNN model

Figure 55 shows the behavior of the objective function throughout the generations of the GA optimization model. As expected, there is a great increase within the first generations and slower improvements in the optimal values towards the last ones.

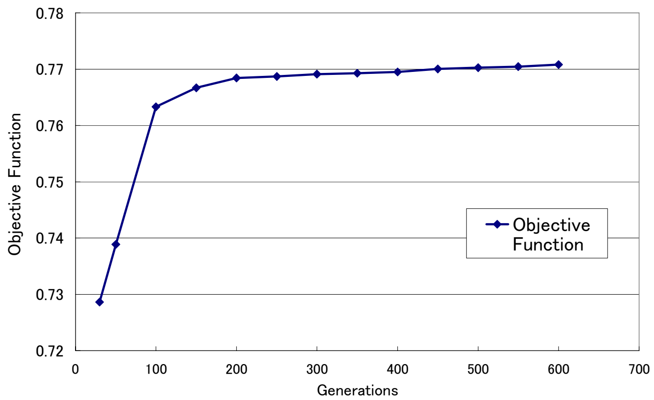


Figure 55. Results of objective function versus GA generations

Besides the previous formulation, another one was carried out for the same optimization schemes as for the previous case, but now the product aggregation operator, equation (50), was considered. This second test was necessary to guarantee that the proposed model is robust enough, independently of the operator being used. In this case as well, the SFNN model showed to be superior to the SDP schemes as can be seen in Figure 56.

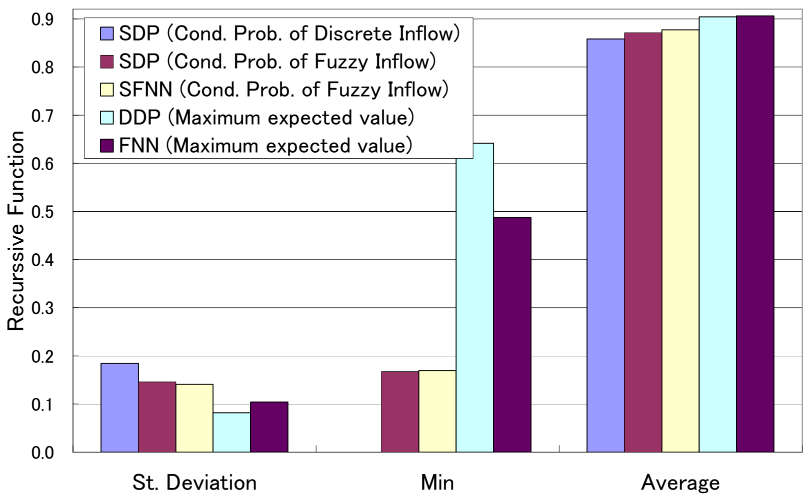


Figure 56. Results from five different optmization schemes (product aggregation)

Within the two deterministic schemes, where inflows are assumed to be known even for the future (validation period), the FNN formulation also showed more efficient results for the final recursive function ("average" in Figure 56). The standard deviation ("st. deviation") and minimum value ("min") are calculated based on the whole validation horizon (132 months) as done before in Figure 52.

Finally, we investigate how the proposed model responds to different recursive functions. The optimization is carried out using six different SFNN formulations depending on the combination of two decision types (minimization and weighted-average) and three aggregation operators (minimization, weighted-average and product). The operation results for the training and validation periods are condensed as cumulative probability function, as shown in Figure 57 and Figure 58, to create a better visualization of the final results after finding the stochastically trained optimized operation rules.

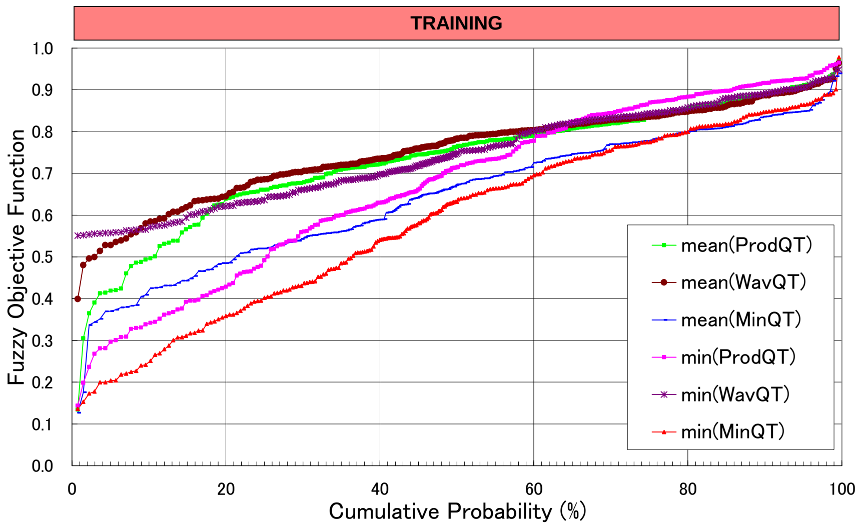


Figure 57. Objective function results after six SFNN formulations - training

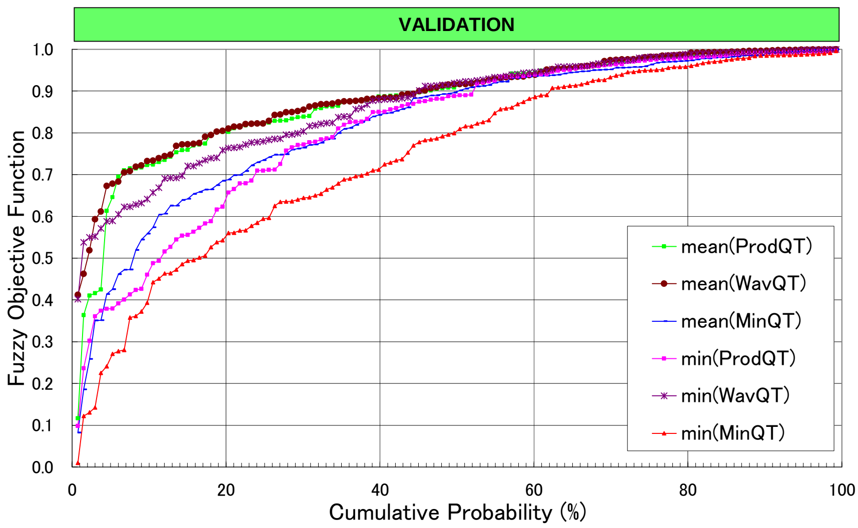


Figure 58. Objective function results after six SFNN formulations - validation

Figure 59 shows the results for both the stabilization (STAB) and hydropower generation (POW), fuzzy objectives for the validation period. This may assist operators to decide on which aggregation operator and decision type to use, depending on their preference for a specific fuzzy objective.

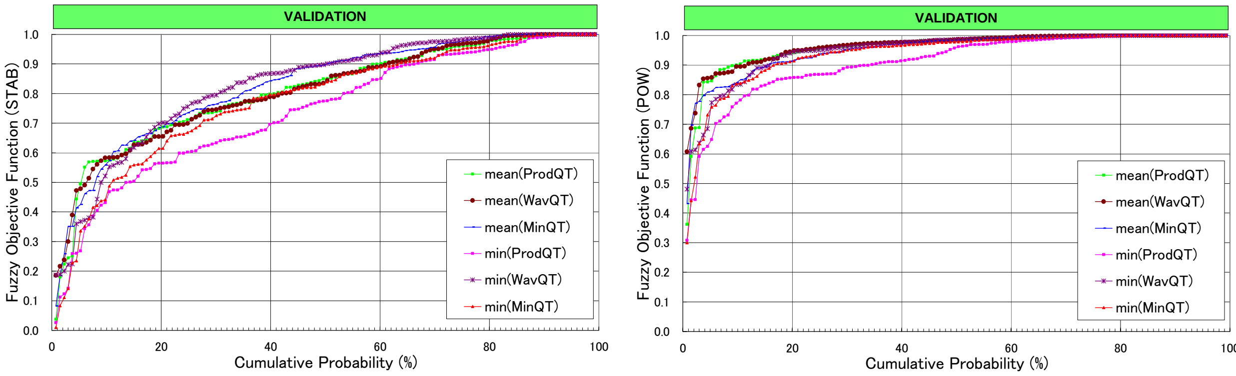


Figure 59. Results of stabilization (left) and hydropower (right) evaluation functions

It can be seen that when using the minimization operator, the overall results were moderately decreased. For our case study, the stochastic training of the neuro-fuzzy system using such operators seems to be less robust, because, even though the minimum value could be maximized for the training period, in the validation period the operation policy represented by the SFNN was not able to return the maximum of the minimum among the other results. That is to say, for the validation period, the resulted minimum value was not the minimum when using the minimization operator, which was not really expected, as this operator was supposed to maximize the minimum values. This problem may also occur with other stochastic optimization techniques as well, such as SDP. Figure 60 shows a comparison of the minimum results found after training and validation for the six SFNN formulations. When looking at Figure 60, the results should be compared within the same aggregation operator for the two periods, training and validation.

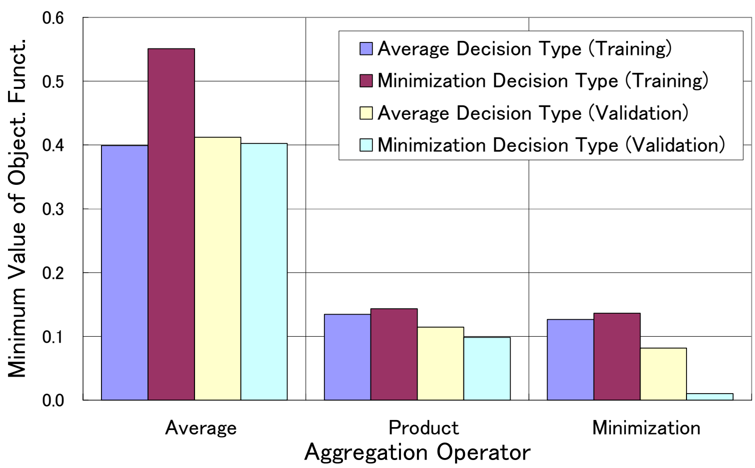


Figure 60. Minimum value of objective function for six SFNN formulations

Basically, it may be said that the decision on which aggregation method and fuzzy type of decision to use depends significantly on the judgment of operators. Even though in our problem the minimization decision type did not show great robustness, in other problems with other operation purposes and different shapes for the fuzzy objective membership functions, the minimization can also finds its use. This decision type demonstrated an inability to reflect any correlation among intermediate evaluation values within the optimization (training) period, as it always chose the smallest among different evaluation values, which in some operation problems may be insufficient for properly achieving efficient operation rules. This may be aggravated when a greater number of objectives and longer training periods are used. The reason is simple, a water resources system such as reservoir operation like many other systems, is designed to present the least risk of failure as possible. However, systems presenting risk “zero” of failure may be physically and economically impossible to achieve. Therefore, if one objective in a certain time of the operation horizon, cannot be satisfied at all (membership function value equals zero) the whole system is assumed to fail. That is why the minimization type is frequently called the pessimistic decision. It is recommended that when using the SFNN for deriving operation rules for practical use, various types of objective functions, aggregation operators and decision types should be tested. Nevertheless, it could be said that as harder objectives are difficultly satisfied during the training period, then the calculated operation rules would be less efficient.

6.7 Conclusions

First, the conditional probability of a fuzzy event proved to increase the efficiency of the stochastic DP. This was possible as a result of the use of more realistic conditional probability curves, found through the fuzzy methodology presented here. This was verified after a comparison of the results from the two SDP optimization schemes. The CP considering stochastic representative inflows as fuzzy events is believed to reduce some of the uncertainties related to the incomplete data set and the crisp discretization process when calculating the conditional probability of discharge values.

The results from the proposed SFNN model showed improvements in optimization efficiency, which could be obtained as a result of the introduction of fuzzy inference to deal with storage discretization. Moreover, the SFNN model has the ability to consider, not only the conditional probability but also, the long-term characteristic of inflow, as the model is trained with actual inflow observation data. This is different from the SDP model which tries all combinations related to the discretized storage levels for a twelve-month period only, ignoring the long-term characteristics of inflow. Moreover, the SFNN also addresses the uncertainties of the conditional probability as the model is prepared to calculate a new expected conditional probability for each time stage *t*, through a kind of interpolation (based on the membership functions) of the CP curves.

Different from the SDP, the forward-moving calculation scheme for the SFNN model was successfully carried out, making the combination of the latter with other models straightforward, such as runoff and water quality models. The possible combination with other models may increase the efficiency of optimization models and may expand its applications for optimization of more complex systems, which for instance may involve more state variables, such as the one presented in the next chapter.

Even though further research is necessary, having the whole simulation being possible at once after the parameters of the neuro-fuzzy system are defined, the application of SFNN seems to present great potential to overcome the problem related to the curse of dimensionality, commonly faced by SDP models, in the case of many state and decision variables. For example, in the case of multi-reservoir systems, new state variables could simply be added as new neuron units.

The results, using different fuzzy decision and aggregation operators, indicated that the SFNN is flexible enough to optimize different types of recursive functions, demonstrating great potential for its application to a variety of other kinds of stochastic optimization problems. As could happen to other techniques as well, the minimization decision type may present less robust results when optimization is carried out for systems with a greater number of operation goals and longer optimization horizon. Moreover, the shape of the fuzzy objective function may also influence the performance of the model when using the minimization and product decision types.

After proving the efficiency of the proposed model for the stochastic optimization of the proposed problem, considering only quantity objectives for the reservoir operation, in the next chapter, a more complex problem, considering water quality objectives is considered.

7 Operation for Quality Objectives

7.1 Introduction

Water quantity and quality in the environment are affected by many different factors. Such factors include increasing water demand, multiple-use, water pollution due to rapid urbanization, high development of water resources, eutrophication and degradation of water bodies and an increase in costs related to water treatment. Recently, assessment of seasonal, social and ecological changes has become more relevant to reservoir management.

There are many in-lake techniques that may be used for the improvement of water quality within the reservoir, such as aeration, artificial mixing, sediment removal and hydraulic regulation (Straskraba and Tundisi, 1999). Hydraulic regulation may present some limitations, such as the requirement of knowledge about complex stochastic, physical, chemical and biological processes, the limitations regarding flushing rates and the increase of negative downstream impacts. However, this method is of great interest as the cost involved in its implementation may be extremely low, which could be an alternative for solving reservoir water quality problems, particularly in developing countries, where investments for the improvement of the environment are rather limited. Moreover, it can be considered a sustainable way to deal with the problems related to reservoir water quality (Zalewski *et al.*, 1997).

Basically, water quality management can focus on one or more of the four components related to a reservoir system, which are inflow, release, storage and withdrawn water from the reservoir. All of these four components are interrelated in respect to their quantity and quality. Table 8 shows some of the methods and variables to be considered when optimizing the water quality for each of the four components of a reservoir system.

Management models play a crucial role for achieving better and more efficient management options. For most of them, some research has been already carried out. From the different types of models available for use in the management of water quality, two are of greater importance: the predictive and optimization models. The first is intended for off-line analysis which intends to investigate several management alternatives and scenarios. Optimization models can be used on-line for actual operation or may be applied for planning optimization when connected to simulation models.

Table 8. Water quality optimization considering the hydraulic components and methods

Component to be optimized when Quality is considered	Method and decision variables	Examples
- Inflow (and consequently the reservoir and rivers water quality)	- Control of point source - Control of non-point source - Watershed management	- Discharge from industries and wastewater treatment plants - Agriculture and urban runoffs - Land use and socio-economic activities
- Withdrawn water from reservoir	- Operation of outtakes	- Selective outtakes depending on the vertical water quality situation of the water column.
- Released water from reservoir (downstream areas)	- Operation of multiple outlets	- Bottom and spillway discharges - Selective outlets
- Storage volumes	- Operation of release, storage and inflow volumes	- Bottom and spillway discharges - Selective outlets - Diversion of inflow (if applicable) - Optimal operation rules considering the dilution capability and retention time.

As predictive models, Wu *et al*. (1996) investigate the effect of reservoir operation on downstream water quality through simulation techniques. The Ta-Chi reservoir and watershed in Taiwan are simulated through a distributed rainfall-runoff and reservoir water quality (WASP) model, respectively. Results demonstrated that operation rules may improve the outlet water quality by releasing water from a certain depth in the reservoir. Another similar work considering different watershed management and pollution control as decision options with the simulation of its impacts on the reservoir carried out by a water quality model (CE-QUAL-W2) can be found in Gunduz *et al*. (1998).

For the last 40 years, there have been only few attempts to optimize and improve water quality condition in rivers, frequently using linear programming and dynamic programming. Most of the works in this field have focused on the optimization of few (one or two) water quality variables, such as for the optimal strategies of wastewater treatment related to the DO-BOD relation. Ellis (1987) showed an example of river quality stochastic optimization considering DO-BOD and a list of other relevant works in the field of river water quality optimization. Another attempt to improve water quality of rivers (and indirectly reservoirs) is by means of watershed management, where land use management and pollution control plays an important role. Simulation models for land use impacts on river water quality combined with optimization models has already been developed such as by Das and Haimes (1979) and Goicoechea and Duckstein (1976). However, little attention has been applied to the water quality in reservoir when storage volumes are optimized. Even though reservoir and lake water quality simulation models have been well developed, only a few of the works are simultaneously introducing such models into the optimization framework.

Nandalal *et al.* (1995) developed a deterministic dynamic programming (DDP) optimization model combined with a rather simple model for reservoir salinity, with the objective to decrease salinity in release and storage volumes applied for the Jarreh Reservoir, Iran. The system was optimized by controlling release and inflow volumes. However, Nandalal did not address some important issues, such as stochastic characteristics of inflow, multi-objective analysis considering quantity and quality objectives, multiple water quality parameters, uncertainties related to the salinity model and uncertainties related to the objectives themselves. Another application of reservoir optimization by deterministic DP, considering the minimization of downstream turbidity through the operation of selective intakes, can be found in Ikebuchi and Kojiri (1992). Other applications of DP, for the optimization of downstream water quality, can be found in Fontane *et al.* (1981), Sale *et al*. (1982) and Dandy and Crawley (1992).

Recently, Chaves *et al.* (2003a, b) and Chaves *et al.* (2004) successfully combined DP optimization models with simultaneous water quality simulation for multiple parameter, which was carried out through a physical (PAMOLARE) and an artificial neural network (ANN) model, respectively. Nevertheless, to be able to do so, some inconvenient assumptions had to be made. For the combination of DP with the physical model, stochastic dynamic programming (SDP) was applied for the quantity optimization and then quality objective was optimized within an "optimal" path through a deterministic DP model. For the ANN based quality model, SDP was applied for quality optimization. However, quality parameters were assumed to be time-independent, as the optimization was implemented for monthly time-steps. However, it would not be valid for short-term operation, as time-dependence should be considered. On the other hand, if the quality variables are considered time dependent, the number of state variables increases and the methodology may not be feasible due to the curse of dimensionality, a common problem associated to DP models. Besides the problem of dimensionality, as already shown previously in Chapter 3, the backward-moving calculation in which SDP had to be realized became another obstacle for the combination of SDP with simulation models (Chaves *et al.*, 2003a, b).

It would appear that there is still a lot of research needed in this field to address topics such as simultaneous combination of stochastic optimization and water quality simulation models, multiple water quality parameters, uncertainty analysis, multiobjectives, multiple purpose reservoirs and integrated quality management considering inflows, storages and releases. Our objective is to address some of these topics, where quality in storage and stochastic characteristics of inflow are considered in a combined optimization framework to obtain the optimal planning operation rules, which is intended to achieve satisfaction for multiple quantity and quality objectives.

The application of stochastic and artificial intelligence (AI) techniques, such as fuzzy set theory and neural networks, in the development of decision support systems can provide suitable ways to analyze all these complex connections and uncertainties, as we have already shown in the previous chapters. After testing the proposed stochastic optimization model (Chapter 6) and calibration of the water quality simulation model (Chapter 5) we are ready to carry out the operation of the reservoir considering water quality objectives through the simultaneous integration of both proposed models.

The reservoir water quality is represented by the central part of the water body, just after the junction of the two main tributaries. Storage, inflow and release act as the key factors influencing system behavior. The analysis is conducted considering quantity and quality at each time-step in the optimization process, since quality parameters are time-dependent. Although, a neural-network model is used for quality simulation, the ability to carry out the optimization, considering time-dependence for quality parameters, allows the proposed methodology to also be executed using a physical simulation model. For calculation, a monthly time-step is used, since the analysis is intended to find the operation rules for the monthly stationary operation.

7.2 Simultaneous Combination of Simulation and Optimization Models

In an attempt to overcome the limitations of the stochastic DP encountered by Chaves *et al.* (2003a, b) and Chaves *et al.* (2004), this chapter proposes the integration of the stochastic fuzzy neural network (SFNN) optimization model, which has been introduced in Chapter 6, with the water quality simulation model based on conceptual fuzzy neural network (CFNN), which is proposed in Chapter 5. A schematic diagram showing the combination of both ANN based models is displayed in Figure 61.

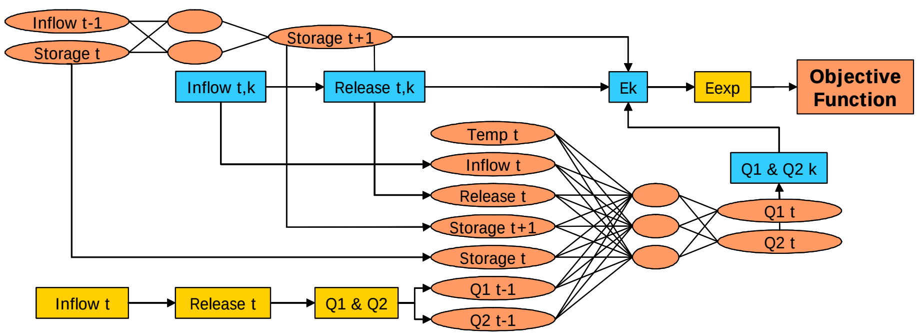


Figure 61. Conceptual integration of ANN based optimization and simulation models

Figure 62 shows the flowchart of the overall calculation procedures for the SFNN optimization model, where inflow values are considered as stochastic variables based on the conditional probability of a fuzzy (inflow) event. Note that the water quality simulation is to be carried out simultaneously to the optimization process.

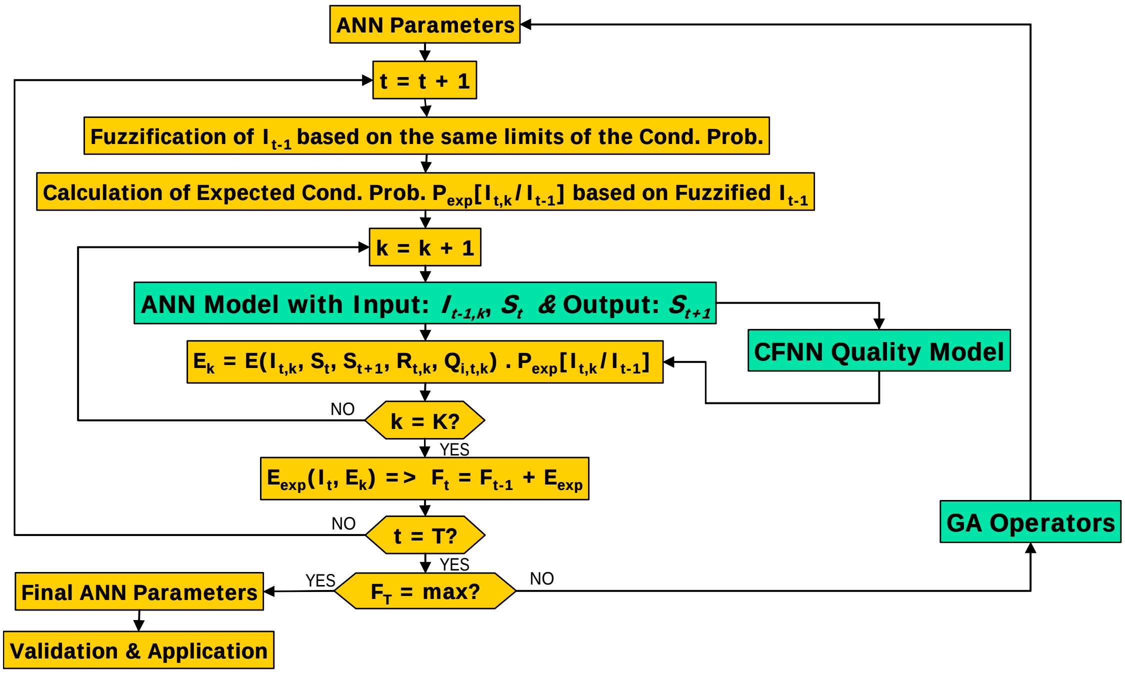


Figure 62: Flowchart for the SFNN model considering water quality objectives

In the case of optimization using the ANN based models; the training is based on the maximization of the recursive function *F*, which could be formulated as shown in equations (60) and (61) for two different types of fuzzy decision:

- *Weighted-average type*

$$F(S_t, I_{t-1}) = \max_{S_{t+1}} \left[E_{\exp}(E_{t,k}) \cdot \alpha + F_{t-1}(S_{t-1}) \right] \tag{60}$$

- *Minimization type*

$$F(S_t, I_{t-1}) = \max_{S_{t+1}} \left[\min \left[E_{\exp}(E_{t,k}) \right] \right] \tag{61}$$

where E_{exp} is the expected value of the evaluation function, represented by equation (48), being a function of the evaluation functions E_k and α is the weight, here all the weights are considered the same and equal to *1/T*. The recursive function for the water quality optimization is analogous to the formulation presented in the previous chapter.

7.3 Water Quality Objectives

The quantity objectives are the same as the ones considered in the previous chapter for the quantity optimization case. However, for quality analysis it is also necessary to elaborate functions that represent the objectives related to the water quality. Due to a lack of more realistic and specific information about the actual reservoir operation, lets just assume our objectives are the overall improvement of the water quality within the reservoir through means of the storage operation, considering the main present water quality problem and common quality parameters. Five quality parameters are then considered, dissolved oxygen (DO), biochemical oxygen demand (BOD), total nitrogen (TN), total phosphorous (TP) and chlorophyll (CHA). The evaluation functions are constructed based on limits and standards of water quality parameters that are actually being used in Brazil for the evaluation of the State of Sao Paulo reservoirs, including our case study, the Barra Bonita reservoir. The evaluation functions for each water quality parameter are shown in Figure 63. More general information on the elaboration of water quality parameters can be found in Bollman and Marques (2000). For more information on the actual water quality indices used in Sao Paulo State, readers are referred to CETESP homepage (in Portuguese).

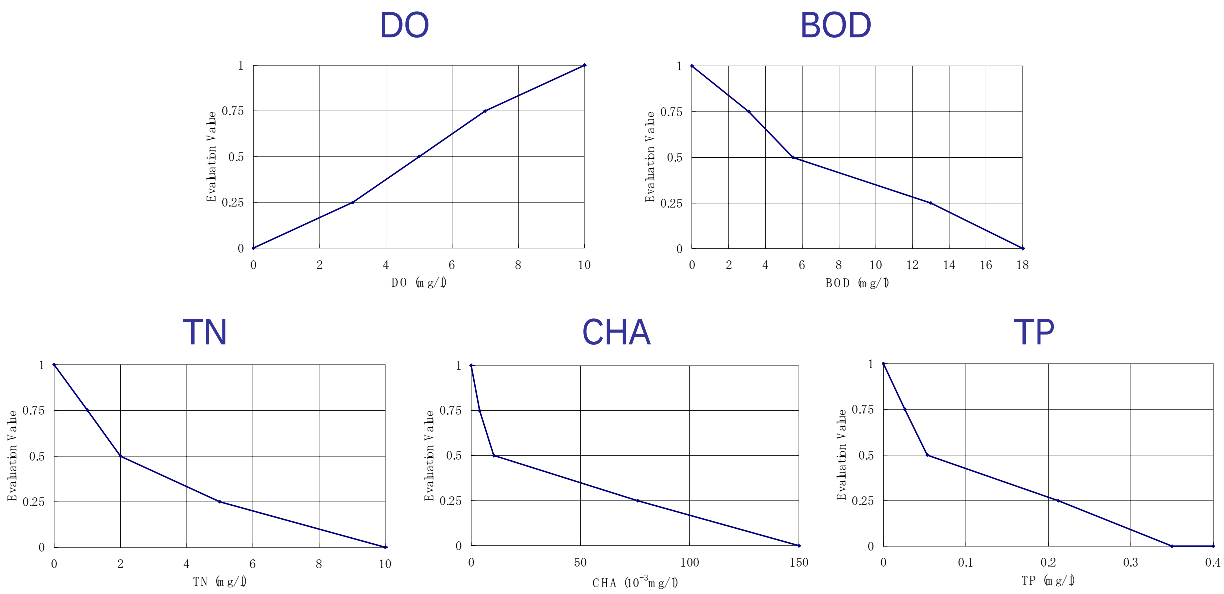


Figure 63: Fuzzy membership functions for reservoir water quality evaluation

The five water quality parameters considered here can be divided here into two categories: parameters regarding the trophic state (TN, TP and CHA) and aquatic life preservation (DO and BOD). Therefore, similar to what is done by the authorities in the state of Sao Paulo, where different indices are used for different water uses, out of these five evaluation functions, two indices are developed, defined by EUTRO and OXYG. These indices are of main concern regarding the reservoir trophic state and its aquatic life preservation condition, respectively. Table 9 and Table 10 show the limits assumed in the identification of the quality evaluation function and the linguistic representation for each of the two indices, respectively.

Note that the simplification resulting from the combining of five quality parameters into two indices is not a requirement for the proposed optimization model. This is carried out in an attempt to improve the water quality evaluation. From the optimization viewpoint, the proposed method could easily handle more quality parameters and quality indices if it was necessary.

Table 9. Attributes for quality parameters in the evaluation of trophic state

Index	Linguistic Meaning	Evaluation Value	Parameters TP (mg/l)	CHA (10^{-3}mg/l)	TN (mg/l)
EUTRO	Oligotrophic	1.00	0	0	0
	Mesotrophic	0.75	0.027	3.8	1.0
	Eutrophic	0.50	0.053	10.3	2.0
	Very Eutrophic	0.25	0.212	76.1	5.0
	Hypereutrophic	0.00	0.35	150	10.0

Table 10. Attributes for quality parameters in the evaluation of aquatic life preservation

Index	Linguistic Meaning	Evaluation Value	Parameters DO (mg/l)	BOD (mg/l)
OXYG	Very Good	1.00	10.0	0
	Good	0.75	7.0	3.1
	Acebtable	0.50	5.0	5.5
	Poor	0.25	3.0	13.0
	Very Poor	0.00	0	18.0

EUTRO and OXYG indices are calculated as the average of the values given by the evaluation functions of the corresponding water quality parameters. This method (average) is actually used in the elaboration of many quality indices used worldwide, including the State of Sao Paulo. The final equation for the calculation of EUTRO and OXYG are shown in (62) and (63), respectively:

$$EUTRO = \frac{E(TP) + E(CHA) + E(TN)}{3} \qquad (62)$$

$$OXYG = \frac{E(DO) + E(BOD)}{2} \qquad (63)$$

As already shown in Chapter 5 regarding the water quality simulation, results from the CFNN models are provided with some “credibility interval”, which can be also seen as a fuzzy number. There are many ways to evaluate fuzzy numbers when using fuzzy evaluation functions such as the ones shown in Figure 63. As the fuzzy numbers originate from the CFNN water quality model and the evaluation functions present the same domain, it is possible to plot both in the same graph as shown in Figure 64, which demonstrates a hypothetical example for dissolved oxygen. The method used here to obtain the final evaluated value is based on the expected valued calculated using equation (64).

$$E_n = \frac{\sum_{i=1}^{I} y_i \cdot q_i}{\sum_{i=1}^{I} y_i} \qquad (64)$$

Where E_n is the final evaluation value for each water quality parameter n (e.g. DO, BOD, CHA); q_i is the quality ith value of the fuzzy number and y_i is the ith value of the evaluation function when considering domain discretization with I intervals.

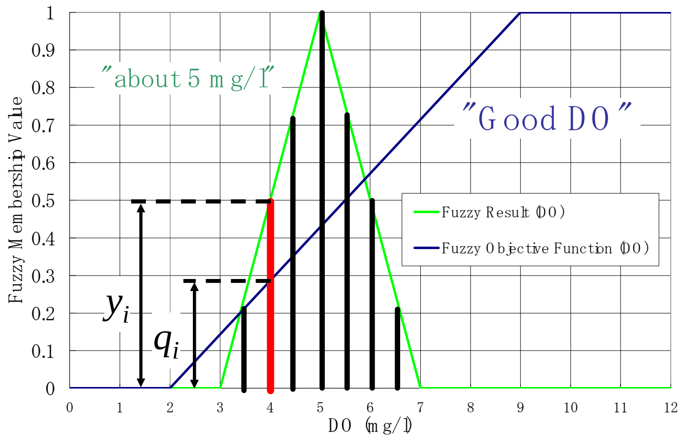


Figure 64. Evaluation of fuzzy results

7.4 Considering Quality Objectives in the Reservoir Optimization

In Chapter 2, some reasons as to why water quality is still left out of the operation and management framework have already been provided. Table 11 summarizes some of these reasons comparing the quantity and quality objectives. It is clear that in most cases, objectives relating to water quality variables, such as release and storage, are much easier to quantify. Moreover, quantity variables present much less uncertainty in their values, as measurements and even simulation can be achieved with much more accuracy and with much less effort and at lower costs.

Acknowledging these facts, an incautious consideration of quality aspects into the operation framework could result in catastrophic results. Two basic ways to include the quality objectives into the operation framework could then be suggested. First, if the uncertainties relating to the calculation of quality parameters are believed to be reduced and quality objectives can be clearly defined through evaluation functions that realistically represent the goals of all stakeholders affected by the operation of the system, it could be done through directly mathematical combination, using the same aggregation operators used to handling the quantity objectives mentioned in the previous chapter. This is exemplified later in what is called *Case 1*.

Table 11. Comparison between Quantity and Quality objectives

Characteristics	Objectives	
	QUANTITY	**QUALITY**
Data Collection	Low Cost Faster	High Cost Slower
Data Availability	Plenty	Scarce
Relation to Hydrological variables	Straightforward	Complex
Assessment by momentary units	Easy	Difficult
Uncertainty	Low	High
Acceptance by Operators	High	Low

However, this situation is still far from becoming a reality, especially due to the resistance from system operators who, regardless of the pressure to improve environmental quality, still have to cope with strong demands for socio-economic development and water quantity volumes, such as industrial and domestic water supply and hydropower generation. Nevertheless, in some cases, for example, during low water demand or high-water periods, operators can afford to use some of the water for the improvement of water quality. Therefore, the second way that water quality related objectives could be optimized is to consider the quantity volume (or quantity objectives) constraints in the quality optimization process (The opposite could also be considered, when quality reaches a state that it is no longer acceptable and the focus turns to the prioritization of the quality objectives). This second way to see the combination problem seems to be more realistic and could be easier for operators to accept. On the other hand, the possible benefits that could be achieved from the optimization of quality objectives may be limited and even meaningless. This is exemplified later in what is called *Case 2*.

7.4.1 Case 1: Quantity and Quality Objectives are Considered Equally Important

As carried out in the previous chapter for quantity optimization, here we test some aggregation operators. The minimization type decision is not tested because, as shown before, training with this operator showed unstable (not robust) results. This may even be aggravated more due to the nature of the water quality condition, which presents limitations on the ration of improvement, independent of the adopted policy. Equations (65) and (66) below show the weighted-average aggregation operator, considering the same weights for the evaluation functions and the recursive equation applied in this case, respectively.

$$E_t = \frac{1}{N}\sum_{n=1}^{N} E_{t,n} + \frac{1}{K}\sum_{k=1}^{K} E_{t,k} \tag{65}$$

$$F = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K} E_t . P_{\exp,k} \tag{66}$$

Where E is the fuzzy objective function, indices n and k represent the quantity (STAB and POW) and quality (EUTRO and OXYG) objectives, respectively. N and K are the maximum number of quantity and quality objectives, which is equal to 2 objectives for both. T is the total number of time-steps t, which is equal to 132 months for the validation period. P_{exp} is the expected probability already defined in the previous chapter.

7.4.2 Case 2: Quantity Objectives are Given Priority over Quality Ones

In *Case 2*, quality condition (objectives) is only considered under certain constraints. Here, we hypothetically chose to account for quality objective only when both quantity objectives (STAB and POW) are higher than 0.8 and the minimum value between the two quality objectives (EUTRO and OXYG) is smaller than 0.5, as in equation (67). In this case, the weighted average type decision is also applied as shown in equation (68).

$$\begin{aligned} & \mathit{if}\ (\min(STAB, POW) > 0.8\ \&\ \min(EUTRO, OXYG) < 0.5) \\ & \Rightarrow E_t = \min(EUTRO, OXYG) \\ & \mathit{else} \\ & \Rightarrow E_t = average(STAB, POW) \end{aligned} \tag{67}$$

$$F = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K} E_t . P_{\exp,k} \tag{68}$$

7.5 Results

7.5.1 Results for Case 1

The results of the quantity and quality objectives before and after a consideration of the water quality related objectives (*Case 1*) are shown in the following figures. In Figure 66, the results are supplied as cumulative probability functions of the calculated results of the evaluation functions, referent to the validation period (132 months). For the case of Figure 66, with average fuzzy type decision for weighted average aggregation operator, as shown in equation (65), results before and after operation considering water quality showed improvements on the water quality condition. Water quality results, after consideration of water quality objectives were greater than 0.5, whereas without considering the quality objectives around 25 % was below this level. Note that 0.5 means the membership value after the minimization operator of two (EUTRO and OXYG) quality objectives, representing the degree of satisfaction in fulfilling these objectives.

On the other hand, quantity objectives showed some decrease. As mentioned before, operation considering quality objectives is a multiobjective problem, where tradeoffs between quanity and quality objectives are expected. Therefore, improvements on some objectives are expected to drive losses in others. These results can give the operators a better way to understand the problem, resulting consequently in better decisions. These results are shown as cumulative probability curves in Figure 65 and Figure 66 for the training and validation periods, respectively.

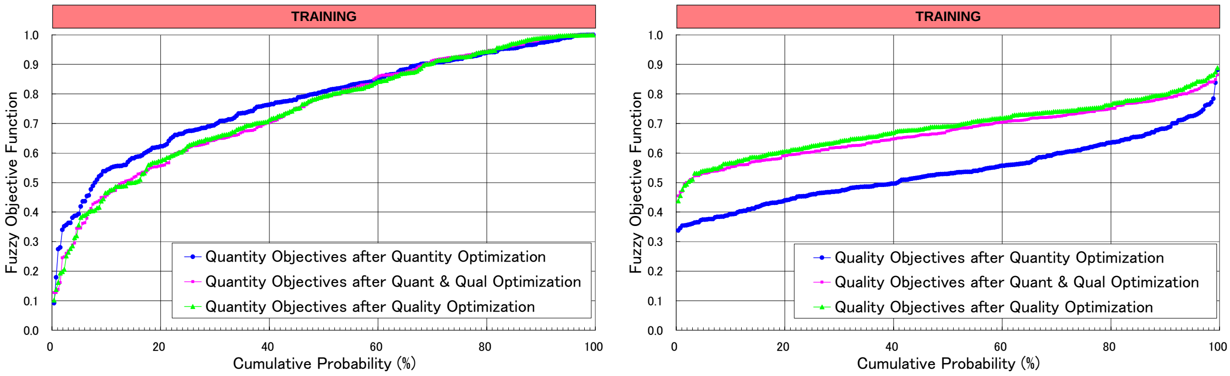


Figure 65: Quantity (left) and Quality (right) objectives after Training for Case1

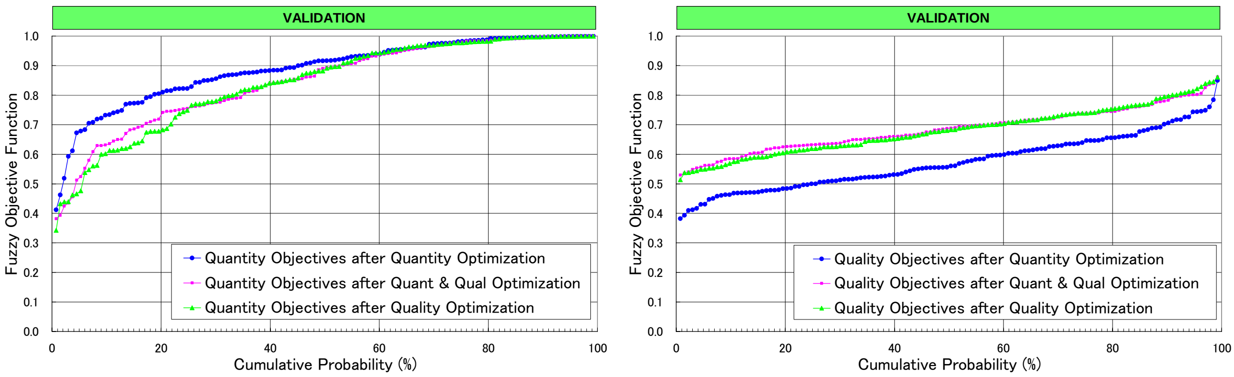


Figure 66: Quantity (left) and Quality (right) objectives after Validation for Case1

The results for *Case 1*, can also be seen for each individual water quality parameter, either with their actual unit or by their fuzzy membership functions as shown in the next figures. Figure 67 shows the results for BOD and Figure 68, the ones for chlorophyll-a. Note that the pick concentrations could be fairly reduced. The reduction could be even greater if a single parameter had been considered or pick reduction was applied as an operation objective.

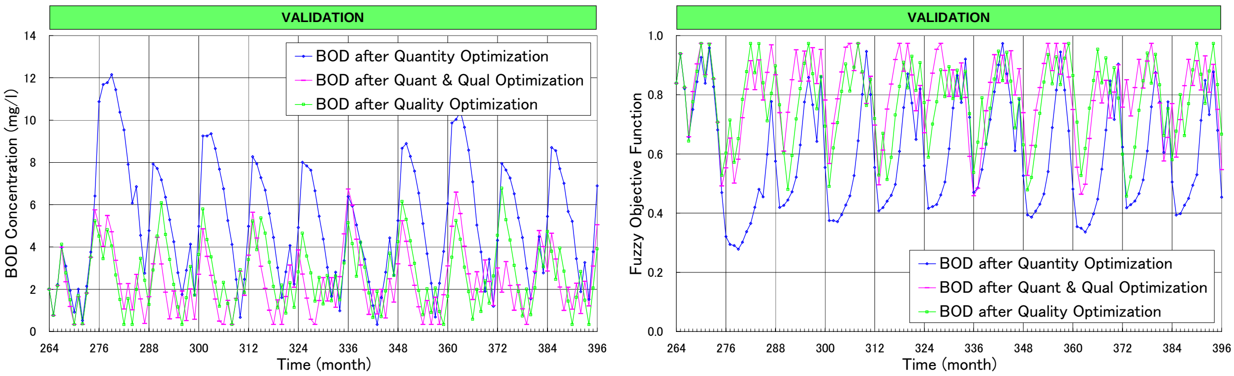


Figure 67: BOD concentration (left) and BOD evaluation function (right) for Case 1

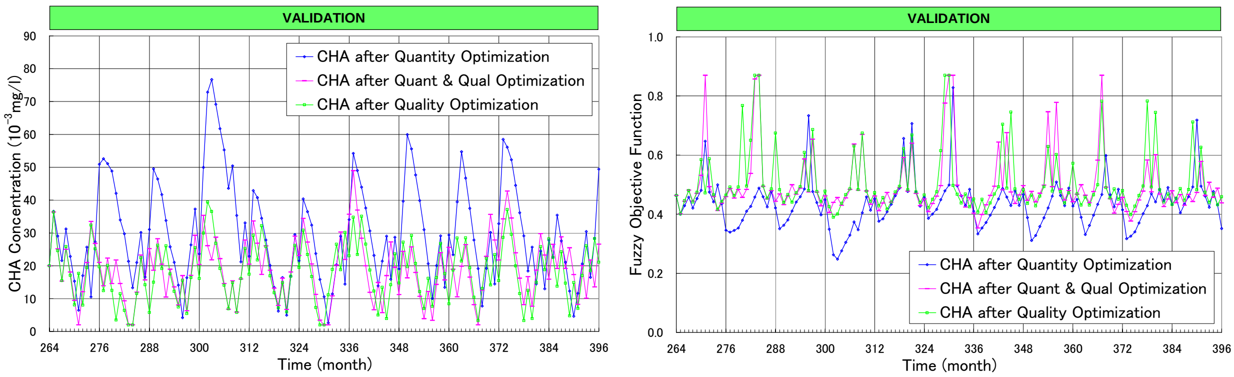


Figure 68: CHA concentration (left) and CHA evaluation function (right) for Case 1

Figure 69 shows the optimal results of end-of-period storage for the training and validation periods of *Case 1*, respectively.

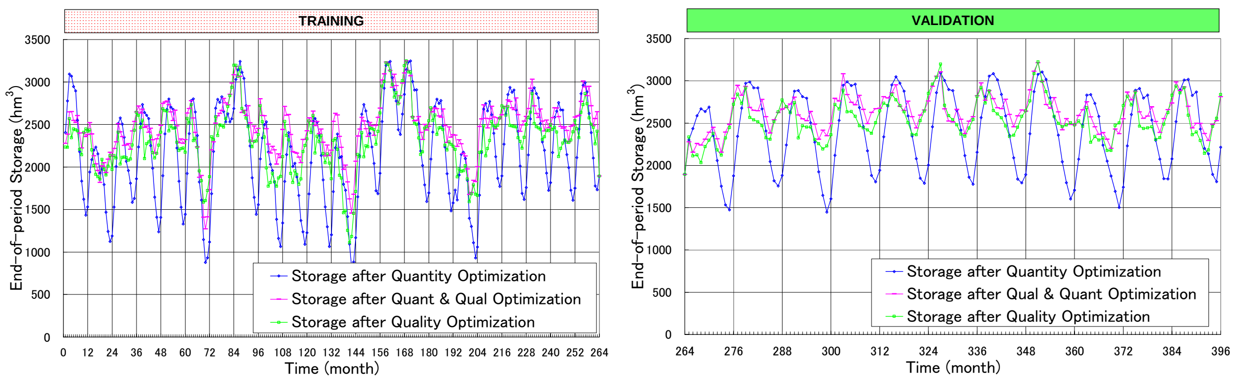


Figure 69: Optimal storage sequence for training (left) and validation (right) for Case 1

7.5.2 Results for Case 2

For *Case 2*, with the formulation presented in equation (67) and (68), as expected, the results of the water quality objectives showed much less improvement. On the other hand, the losses related to the quantity objectives seemed to compensate for the improvements in water quality, much more clearly than in the previous case.

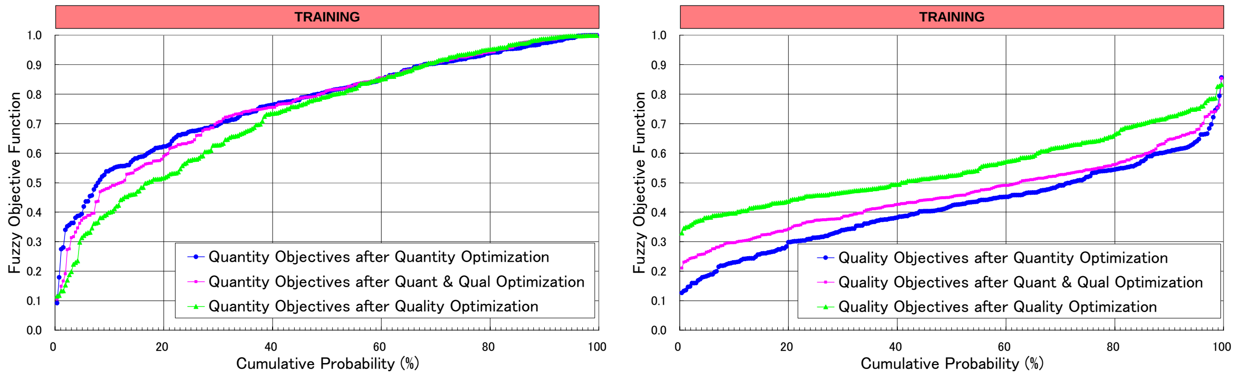


Figure 70: Quantity (left) and Quality (right) objectives after Training for Case2

However, again it will be a matter of trade-off between the multiple objectives involved, particularly quantity and quality, which will lead the final decision regarding the actual dam operation. These results shown in Figure 70 and Figure 71 are for the training and validation periods, respectively.

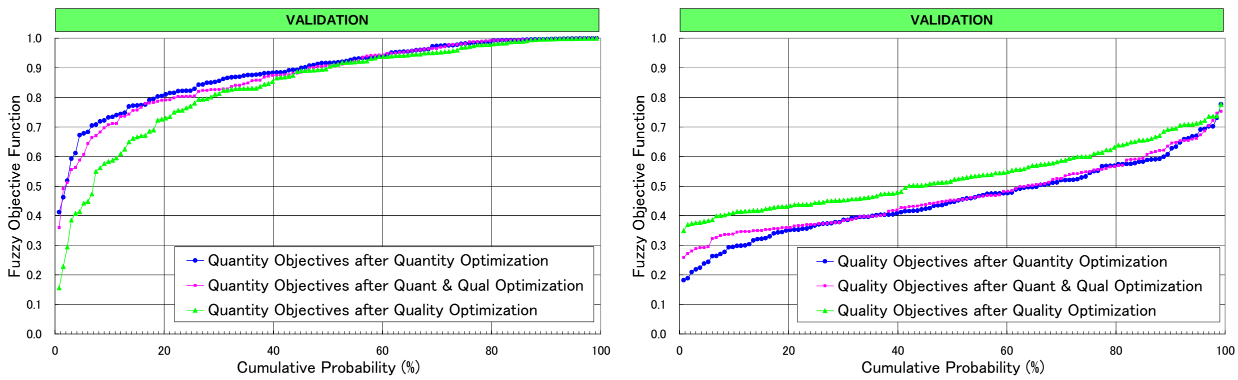


Figure 71: Quantity (left) and Quality (right) objectives after Validation for Case2

Also for *Case 2*, the same results displayed in Figure 71 can also be given against time, as plotted in Figure 72.

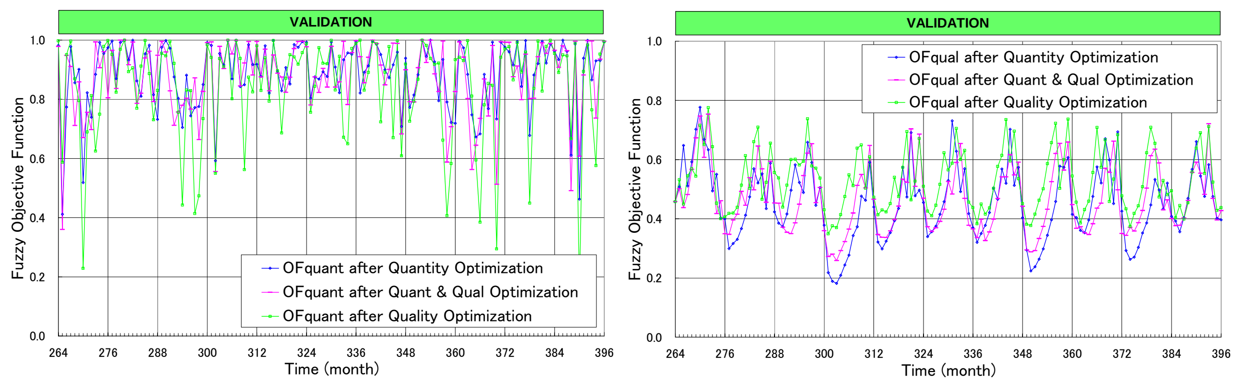


Figure 72: Quantity (left) and Quality (right) objectives against time after Validation for Case 2

Moreover, the results of end-of-period storage for training and validation periods are presented in Figure 73.

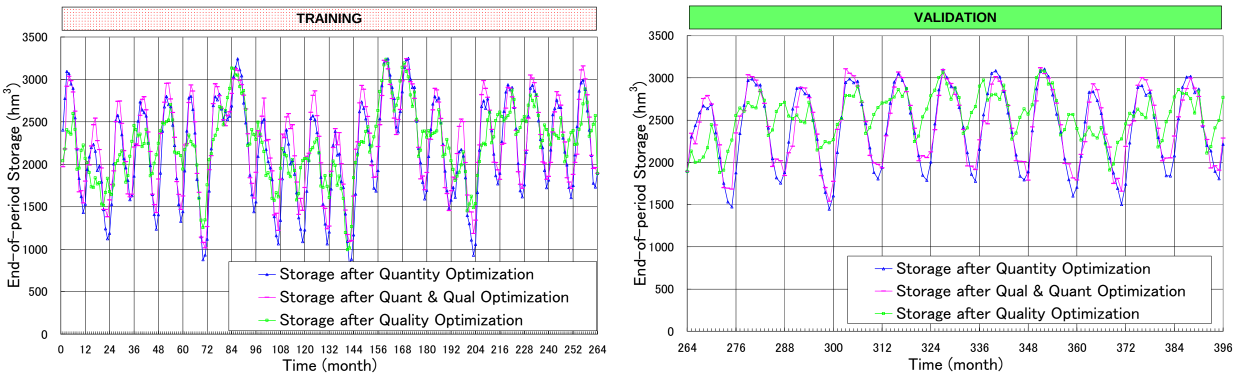


Figure 73: Optimal storage sequence for training (left) and validation (right) for Case 2

7.6 Conclusions

The results indicated that water quality condition could be improved with the proposed methodology. However, it is very important to clearly and accurately define the operation objectives and water quality evaluations before any change in the operation rules can be carried out, as such modification can greatly influence the final results.

It was also demonstrated that even though better quality could be achieved, water quality improvement only by means of hydraulic control has its limitations, due to the system characteristics such as the quality of inflow water. The gains to be achieved with such approach has to be studied in advance, as carried out in this research, and only then should the operation focus on quality improvement be implemented. In the case that hydraulic control is not sufficient, other techniques such as watershed management should be also considered. The latter combined with the proposed approach has great potential to fully satisfy the stated objectives.

Moreover, it is important to note that the accuracy of simulation models will also affect the derived final operation rules. Here, accuracy does not only mean the error between actual observed and calculated values with actual units, but based on the importance of these differences and the actual quality objectives. For example, an error of +1 mg/l of TP when the calculated value is over 1 mg/l is much less important than an error of +0.1 when the calculated value is equal to 0.1 mg/l, in terms of environment impacts. Therefore, when analyzing the water quality simulation model for the purpose of operation, this analysis becomes extremely important. An increase in the models accuracy and confidence and proper definition of quality objectives tends to guide the operation to Case 1, as presented here. On the other hand, if analysis of water quality and formulation of quality objectives are assumed to be too vague and inaccurate, the proposed methodology may be more efficient when carried out as indicated in Case 2.

To increase the credibility of the calculated operation rules used for water quality improvement, as carried out in this chapter; other water quality models may also be applied. This is possible since the proposed formulation presents a forward calculation direction, which makes feasible the combination of the proposed optimization model with other simulation and prediction models.

8 Conclusions

A new methodology to achieve better water quality by taking advantage of the close relationship between the quality (limnological) and quantity (hydrological and climatological characteristics) of water for reservoir operation was successfully carried out. Better management of stored volumes to yield optimal benefits for water utilization while at the same time improving water quality has been effectively proposed and the potential of water quality improvements have been clearly presented. The application of various AI techniques, such as artificial neural networks (ANN), fuzzy theory and genetic algorithm (GA), has shown to be an effective way to achieve the research objectives, whereas state of the art methods have also been proposed, such as the CFNN for water quality simulation and the SFNN for the stochastic operation of the reservoir. The analysis focused on the planning operation of the Reservoir for monthly time-steps considering the simultaneous combination of the optimization and simulation models. The relationship between storage volume, release from a single outlet, and quality in the reservoir were considered as the main driving forces of the system.

It is obvious that operation procedure influences water quality and that changes in quality can be used to seek the most suitable water conditions in the system through the operation rules. The proposed methodology was successfully tested and applied for our case study for the improvement of water quality after derivation of optimal operation rules represented by a neuro-fuzzy system. However, the proposed methodology can only improve the water quality condition to a certain degree. For further improvements, it is necessary to address the problem, considering the whole watershed with an integrated quality management viewpoint. That is to say, there are still other issues inside the management scope which have a great influence over water quality and quantity, such as control of point and non-point source of pollutions, land use management and sustainable economic development, which were not directly considered here. Therefore, the greater the connection between storage and water quality, the more applicability and success the proposed methodology will have.

The efficiency of the resulted operation, including quantity and quality objectives, will depend most heavily on the capability of decision-makers in defining proper objectives for the optimization models and also on the tradeoffs between gain in water quality improvements and losses related to water quantity related activities, when this is the case. Nevertheless, during some periods of the year, for example high water periods, these losses related to quantity objectives may be minimized, and quantity objectives may be satisfied at a low cost.

There is a great degree of uncertainty regarding inflow quality, since it is closely related to human activities and land use. However, those aspects are not directly considered in this work. A more detailed study to effectively assess inflow quality is recommended. One possibility is, for example, through the use of distributed models, which can consider land use and other physical characteristics of the basin to calculate inflow quality. Other models, based on techniques such as the one presented in Chapter 5, may also be applied to assess and predict inflow quality.

The application of ANN based models for the water quality simulation present more flexibility when compared to physical models. Due to the straightforward way that ANN model can be constructed, as input and output information can easily be chosen and modified, its application for the water quality simulation may present much more flexibility when compared to physical models. Moreover, physical models tend to require much more knowledge about their parameters, which is frequently not available, and may substantially increase the calculation time when combined with optimization models, such as the one proposed here. This may not be applicable in some cases, such as for real-time operation.

The proposed methodology has great potential to be applied to a system of reservoirs to obtain maximum benefits for the whole system. Moreover, the potential applicability for real-time and online operation can be easily seen. Of course, for both cases, some adjustments should be carried out. For example, in the case of real-time operation, the objectives may be changed and prediction models may be incorporated. Nevertheless, as both simulation and optimization models are based on ANN, it is quite simple to add in or take out neurons to represent new components of the systems. Another example when considering real-time operation regards the water quality in release, as discharged water in a short scale will influence more downstream areas than for the monthly operation.

To effectively be able to apply the proposed methodology for the real case of reservoir operation or to increase the confidence of its results, some points may be proposed, as follows:

i) An application of multiple water quality models may increase the credibility of water quality analysis. Also, quality models should be applied concerning the purposes of analyses.
ii) The methodology should be tested in other case studies for increasing its credibility.
iii) The effects of release quality should be studied.
iv) Socio-economic analysis may indeed be necessary.

v) Uncertainties regarding the variables, such as inflow quality and radiation, and changing conditions may also be considered.

vi) The collection of more data, particularly related to water quality and climatologic characteristics, is recommended.

vii) Evaluation functions have to be developed with the participation of all stakeholders involved in water management taking into consideration the present situation of the reservoir and the operation objectives.

REFERENCES

Azevedo, G.T.A., *et al*. Integration of water quantity and quality in strategic river basin planning. Journal of Water Resources Planning and Management, pp. 85-97, March/April. USA, 2000.

Banerjee, M.; Mitra, S. and Pal, S.K.; Rough fuzzy MLP: knowledge encoding and classification, IEEE Transactions of Neural Network 9 (6): pp 1203 – 1216, 1998.

Beck, B.M., Water quality modeling: a review of the analysis of uncertainty. Water Resources Research, Vol. 23, No. 8, pp. 1393-1442, August, 1987.

Bellman, Richard. Dynamic Programming. Princeton University Press. New Jersey, 1959.

Bernardo, J.M. and Smith, A.F.M., Bayesian Theory, John Wiley and Sons, Ltd., Chichester, 1994.

Biswas, A.K., A History of Hydrology. North Holland Publishing, Amsterdam, 1970.

Biswas, A.K., Jellali, M. and Stout, G., eds., Water for Sustainable Development in the Twenty-first Century. Oxford University Press, Delhi, 1993.

Bollmann and Marques. Bases para a estruturação de indicadores de qualidade de aguas. RBRH vol 5 n.1 Jan/Mar 2000, pp. 37-60. ABRH. Brazil 2000. (in Portuguese)

Butcher, W., Stochastic dynamic programming for optimum reservoir operation. Water Resources Bulletin, Vol. 7, No. 1, pp. 115-123, 1971.

Cancelliere A., Giuliano, G., Ancarani, A. and Rossi, G., A neural networks approach for deriving irrigation reservoir operating policies. Water Resources Research, Vol. 16, pp. 71-88, 2002.

Cardwell, H., Ellis, H., Stochastic dynamic programming models for water quality management, Water Resources Research, Vol. 29, No. 4, pp. 803-813, April 1993.

Chandramouli, V. and Raman, H., Multireservoir modeling with dynamic programming and neural network. Journal of Water Resources Planning and Management, Vol. 127, No. 2, 2001.

Chang N.B., Wen, C.G., Chen, Y.L. and Yong, Y.C., A grey fuzzy multiobjective programming approach for the optimal planning of reservoir watershed. Part A: theoretical development, Water Research, Vol. 30, pp. 2329-2334, 1996.

Chang, L.C. and Chang, F.J., Intelligent control for modeling of real-time reservoir operation. Hydrological Processes, Vol. 15, pp. 1621-1634, 2001.

Chang, N.B., Chew, H.W. and Ning, S.K., Identification of river water quality using the fuzzy synthetic evaluation approach, Journal of Environmental Management, Vol. 63, pp. 293-305, 2001

Changchit and Terrell. “CCGP model for multiobjective reservoir systems”. Journal Water Resources Planning and Management, ASCE, Vol. 115, No. 5, 658-670. 1989.

Chaves, P., Planning Operation of Storage Reservoir for Water Quantity and Quality, Master Thesis, Graduate School of Engineering, Kyoto University, Japan, 2002.

Chaves, P. and Kojiri, T., Multi-objective storage reservoir operation under uncertainty. Annuals of Disaster Prevention Research Institute, Kyoto University, No.46 B, Japan, 2003a.

Chaves, P., Kojiri, T and Yamashiki, Y. Optimization of storage reservoir considering water quantity and quality, Hydrological Processes, Vol. 17, pp. 2769-2793, 2003b.

Chaves, P., Tsukatani, T. and Kojiri, T., Operation of storage reservoir for water quality by using optimization and artificial intelligence techniques. Mathematics and Computers in Simulation, Vol. 67, No. 4-5, pp. 419–432, 2004.

Chow, Ven Te and Maidment, David. Stochastic state variable dynamic programming for reservoir systems analysis. Water Resources Research, Vol. 17, No. 6, pp. 1578-1584, December, 1981.

Chow, Ven Te, “Applied hydrology”. McGraw-Hill. USA, 1964.

Cozzio-Bueler, Rico A. The Design of Neural Networks using a priori Knowledge, A dissertation, Swiss Federal Institute of Technology, Swiss, 1995.

Dandy, G. and Crawley, P., Optimum operation of multiple reservoir system including salinity effects. Water Resources Research, Vol. 28, No. 4, pp. 979-990, 1992.

Dandy, G.C., and Warner, R.F. Planning and Design of Engineering Systems. Unwin Hyman. London, 1988.

Das, F and Haimes, Y.Y., Multiobjective optimization in water quality and land management. Water Resources Research, Vol. 15, No. 6, pp. 1313-1322, 1979.

Dombi, J., Membership function as an evaluation. Fuzzy Sets and Systems, Vol. 35, pp. 1-21. 1990.

El-Swaify and Yakowitz, eds., Multiple Objective Decision Making for Land, Water, and Environment Management. Proceedings of the First International Conference on Multiple Objective Decision Support Systems (MODSS) for Land, Water and Environment Management. CRC Press LLC. Hawaii, September 1998.

Ellis, J.H., Stochastic water quality optimization using imbedded chance constraints. Water Resources Research, Vol. 23, No. 12, pp. 2227-2238, December 1987.

Esogbue, Augustine O., Dynamic Programming for Optimal Water Resources Systems Analysis, Prentice Hall Inc. USA 1989.

Fogel, D.B., An introduction to simulated evolutionary optimization, IEEE Transactions on Neural Networks: Special Issue on Evolutionary Programming, D.B. Fogel and L.J. Fogel (eds.), Vol. 5, No. 1, pp. 3-14, January 1994.

Fontane, D.G., Gates, T.K., Moncada, E., Planning reservoir operation with imprecise objectives. Journal of Water Resource Planning and Management, Vol. 123, No. 3, pp. 154-162. May/June 1997.

Fontane, D.G., J.W. Labadie and B. Louis, Optimal control of reservoir discharge quality through selective withdrawl. Water Resources Research, Vol. 17, pp. 1594-1604, 1981.

Fu, L.M.: Introduction to knowledge-based neural networks, Knowledge-Based Systems, Vol. 8, No. 6, pp. 299-300, December, 1995.

Galvao, C. O. and Valenca, M.S. (editors). Sistemas Inteligentes: Aplicacoes a Recursos Hidricos e Ciencias Ambientais. ABRH. Brazil, 1999. (in Portuguese)

Goicoechea, A. and Duckstein, L., Multiobjective programming in watershed management: a study case of the Chaleston watershed. Water Resources Research, Vol. 12, No. 6, pp. 1085-1092.

Goldberg, D.E. and Kuo, C.H., Genetic algorithms in pipeline optimization. Journal of Computation in Civil Engineering, ASCE, Vol. 1, No. 2, pp. 128-141, 1987.

Gunduz, O., S. Soyupak, C. Yurteri, development of water quality management strategies for the proposed Isikli Reservoir. Water Science and Technology, Vol. 37, No. 2, pp. 369-376, 1995.

Hagan, M. T., Demuth, H.B., Beale, M. H., Demuth, B. H., Neural Network Design, PWS Publishing Company, Boston, 1995.

Harboe, R. “Explicit Stochastic Optimization”. Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization, pp. 295-306. NATO ASI Series. Kluwer Academic Publishers. The Netherlands, 1993.

Hori, T. and Takasao, T., Drought control under the interaction between uncertainty of inflow prediction and the fuzziness of decision criteria. Proc. Int. Conf. On Large Scale Water Resources Development in Developing Countries: New Dimensions of Prospects and Problems, pp. 53-60, 1997.

Howard, R. A., Dynamic Programming and Markov Process. MIT Press, Cambridge, Mass, 1960.

Huang, W.C., Yuan, L.C. and Lee, C.M., Linking genethic algorithm with stochastic dynamic programming to the long-term operation of multiresvoir system. Water Resources Research, Vol. 38, No. 12, 2002.

Huttunen, M and Vehvilainen, B. Coding a conceptual model into a neural network in modeling ice-correction. Proceedings of the ANNIE (Artificial Neural Networks in Engineering) Conference, USA, 1996.

Ikebuji, S. And Kojiri, T., Multiobjective reservoir operation includig turbidity control. Water Resources Bulletin, Vol. 28, No. 1, pp. 223-231, 1992.

Jang, S.R. and Sun, T., Functional equivalence between radial basis functions and fuzzy inference system. IEEE Transaction on Neural Networks 4, pp 156 – 158, 1993.

Jang, S.R. and Sun, T., Neuro-fuzzy modeling and control. Proceedings of IEEE, 83, pp 378 – 405, 1995.

Jin, Yaochu, Advanced fuzzy systems design and applications, Physica-Verlag, New York, 2003.

Jorgensen, S.E., Application of ecological modeling in environment management. Elsevier, Amsterdam, 1983

Kacprzyk, Janusz. Multistage Fuzzy Control, A model based approach to fuzzy control and decision making. John Wiley & Sons, England, 1997.

Kaufmann, A., Introduction to the theory of fuzzy subsets, Vol. I. Fundamental theoretical elements. Academic Press, Orlando, 1975.

Kelman, J. and Stendinger, J.R., Sampling stochastic dynamic programming applied to reservoir operation. Water Resources Research, Vol. 26, No. 3, pp 447-454, March 1990.

Kiely, Gerard, Environment Engineering. McGraw-Hill, Berkshire, 1997.

Klir, G. and Yuan, B, Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice-Hall, Inc., Englewood Cliffs. USA, 1995.

Kojiri, T., Sugiyama, Y., Paudyal, G.N., Fuzzy reservoir operation with multi-objective and few monitoring data. Proceedings of the International Conference on Environment Sound Water Resources Utilization, Bangkok, Thailand. November 1993.

Kojiri, Toshiharu. Planning of reservoir construction through fuzzy optimization. DPRI. Japan, 1992.

Labadie, J.W. Combining simulation and optimization in river basin management. Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization, pp. 345-371. NATO ASI Series. Kluwer Academic Publishers. The Netherlands, 1993.

Labadie, J. W., Optimal operation of multireservoir systems: state-of-the-art review. Journal of Water Resources Planning and Management, Vol. 130, No. 2, pp. 93-111, 2004.

Lamb, J.C., Water Quality and Its Control. John Wiley and Sons, New York, 1985.

Le Cun, Y.; Denker, J. S. and Solla, S. A.: Optimal brain damage. Advances in Neural Information Processing Systems 2, pp 598 – 605, 1990.

Lee, J.H.W.; Huang, Y.; Dickman, M. and Jayawardena, A.W.: Neural network modelling of coastal algal blooms, Ecological Modelling 159, pp 179 – 201, 2003.

Lohani, B.N. and K.B. Hee, A CCDP model for water quality management in the Hsintien River in Taiwan. International journal of Water Resources Development, Vol. 1, pp. 91-114, 1983.

Maidment, David R., editor. “Handbook of Hydrology”. McGraw-Hill. USA, 1993.

Maier, H.R. and Dandy, G.C., The use of artificial neural networks for the prediction of water quality parameters, Water Resources Research Vol. 32, No. 4, 1013-1022, 1996.

Maier, H.R. and Dandy, G.C.: Neural networks for the prediction and forecast of water resources variables: a review of modeling issues and applications, Environment Modelling & Software 15, pp. 101 – 124, 2000.

Mitra, S.; Mitra, P.; Pal, S.K.; Evolutionary modular design of rough knowledge-based network using fuzzy attributes, Neurocomputing 36, pp 45 – 66, 2001.

Mutreja, K.N. “Appplied Hydrology”. Tata McGraw Hill. New Delhi, 1986.

Méllo, Arisvaldo, and Mastos, Leonardo. “Optimização da Operação de Reservatório pela programação dinâmica em um processo de decisão Markoviano”. Revista Brasileira de Recursos Hidricos, vol 4, n.3, pp. 5-15, Jul/Sept. Brazil, 1999.

Nandalal, K.D.W., J.J. Bogardi, Optimal operation of a reservoir for quality control using inflows and outflows. Water Science and Technology, Vol. 31, No. 8, pp. 273-280, 1995.

Norwick, A. and Turksen, I., A model for the measurement of membership and consequences of its empirical implementation. Fuzzy Sets and Systems, Vol. 12, No. 1, 1-25. 1984.

Optiz, D.W. and Shavlik, J.W., Dynamically adding symbolically meaning nodes to knowledge-based neural networks, Knowledge-Base Systems 8 (6), pp 301 – 311, December 1995.

Orlob, G.T., Mathematical modeling of water quality: streams, lakes and reservoirs. Wiley, Chichester, 1983.

Oussar, Y. and Dreyfus, G.: How to be a gray box: dynamic semi-physical modeling, Neural Networks 14, pp 1161 – 1172, 2001.

Pedrycz, W. (1993). “Fuzzy control and fuzzy systems, 2nd Ed., John Wiley & Sons, inc., New York, NY, 1993.

Pike, Ralph W. “Optimization for Engineering Systems”. Van Nostrand Reinhold Company. New York, 1986.

Ponnambalam, K., Karray, F. and Mousavi, S.J, Minimizing variance of reservoir system operations benefits using soft computing tools. Fuzzy Sets and Systems, Vol. 139, pp. 451-461, 2003.

Raman, H. and Chandramouli, V., Deriving a general operating policy for reservoir using neural network. Journal of Water Resources Planning and Management, Vol. 122, No. 5, 1996.

Rea, L. and Parker, R. “Designing and conducting survey research: a comprehensive guide”. Jossey-Bass Pub., San Francisco, CA, 1993.

Reckhow, K.H., The importance of scientific information in decision making. Environment Management, Vol. 18, pp. 161-166, 1994b.

Reckhow, K.H., Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecological modeling, Vol. 72, pp. 1-20, 1994a.

Recknagel, F.: Applications of machine learning to ecological modeling, Ecological Modeling 146, pp 303 – 310, 2001.

Renner, G. and Ekart, A, Genetic algorithms in computer aided design. Computer-aided Design 35, pp 709 – 726, 2003.

Russel, Samuel and Campbell, Paul. "Reservoir Operation Rules with Fuzzy Programming". Journal of Water Resources Planning and Management, pp. 165-170, May/June 1996.

Saggio, A.A., Estudo da eutrofização do reservatório de barra bonita (rio tietê - sp) através de simulação numérica. Dissertação de Mestre, Escola de Engenharia de São Carlos, USP. São Carlos, Brazil 1992.

Sale, M.J., E.D. Brill Jr., E.E. Herrics, An approach to optimizing reservoir operation for downstream aquatic resources. Water Resources Research, Vol. 18, pp. 705-712, 1982.

Schleiter, I.M. *et al*., Modelling water quality, bioindication and population dynamics in lotic ecosystems using Neural Networks, Ecological Modelling, No. 120, pp. 271-286, 1999.

Shavlik, J. W.; A framework for combining symbolic and neural Learning. Technical Report 1123, Computer Sciences Department, University of Wisconsin – Madison, Nov. 1992.

Shrestha, Bijaya. "Fuzzy Rule-Based Modeling of Reservoir Operation". Journal of Water Resources Planning and Management, pp. 262-269, July/August 1996.

Simonovic, S. "Coping with changing objectives in managing an existing multipurpose reservoir". Hydrology of Natural and Manmade Lakes, Publication No. 206, G. Schiller *et al.*, eds., International Association of Hydrological Science, Wallingford, UK, pp. 181-189. 1991.

Straskraba, M. and Tundisi, J.G., Guidelines of lake management – Reservoir water quality management, Vol. 9, ILEC. Japan, 1999.

Suharyanto and Goulter, I., Generating flexible reservoir operation strategy: a fuzzy reasoning approach. Proceedings of the 6th International Conference on Hydro-Science and – Engineering, Vol. VI, Brisbane, Australia, May, 2004.

Thanh, N.C. and Biswas, A., Environment-sound water management, Oxford University Press, Oxford, 1990.

Thomann R.V. and Mueller, J.A., Principles of surface water quality modeling and control. Harper Collins Publishers, New York, 1990.

Torabi, M. and Mobasheri, F. A stochastic dynamic programming model for the optimum operation of a multi-purpose reservoir. Water Resources Bulletin, American Water Resources Association, Vol. 9, No. 6, December 1973.

Towell, G.G. and Shavlik, J.W., Knowledge-based artificial neural networks, Artificial Intelligence Vol. 70, pp. 119-165, 1994.

Tundisi, J.G. and Tundisi, M.: Limnology and eutrophication of Barra Bonita Reservoir, Sao Paulo, Brazil. Archiv für Hydrobiologie Beiheft/Ergebnisse der Limnologie 33, pp. 661 – 676. Stuttgart, May 1990.

Tundisi, J.G. "Limnologia no Século XXI : Perspectivas e Desafios". International Institute of Ecology (IIE). Brazil, 2000.

Wardlaw, R. and Sharif, M., Evaluation of genetic algorithms for optimal reservoir system operation. Journal of Water Resources Planning and Management, Vol. 125, No. 1, pp. 25-33, 1999.

Welch, E.B. and C.R. Patmont, Lake restoration by dilution, Moses Lakes, Washington. Water Research, Vol. 14, pp. 1317-1325, 1980.

Wells, Dan M., *et al.* Stochastic dynamic programming for optimum reservoir operation. Water Resources Bulletin, Vol. 7, No. 1, pp. 113-123, February 1971.

Wilby, R.L., et al Detection of conceptual model rainfall-runof processes inside an Artificial Neural Network. Hydrological Sciences Journal, Vol. 48, No. 2, April 2003.

Wu, R.S., W.R. Sue, C.H. Chen, S.L. Lian, Simulation model for investigating effectof reservoir operation on water quality. Environment Software, Vol. 11, No. 1-3, pp. 143-150, 1996.

Yakowitz, Sidney. Dynamic programming applications in water resource. Water Resources Research, Vol. 18, No. 4, pp. 673-696. August 1982.

Yao, X., Evolving artificial neural networks. Proceeding of the IEEE, Vol. 87, No. 9, September 1999.

Yasdi, R., Combining rough sets learning- and neural learning-method to deal with uncertain and imprecise information. Neurocomputing 7, pp. 61-84, 1995.

Yen, B. C., System and component uncertainties in water resources in J.J. Bogardi and Z.W. Kundzewicz, eds., International Hydrology Series – Risk, Rreliability, Uncertainty and Robustness of Water Resources Systems. Cambridge University Press. pp. 133-142, 2002.

Yiyao, L.; Venkatesh, Y. V. and Ko, C.C., A knowledge-based neural network for fusing edge maps of multi-sensor images, Information Fusion 2, pp. 121 – 133, 2001.

Zadeh, L.A., Probability measures of fuzzy sets, Journal of Mathematical Analysis and Applications 23, pp. 421 – 427, 1968.

Zalewski, M., *et al.* Ecohydrology: a new paradigm for sustainable use of aquatic resources. UNESCO. Paris, France, 1997.

APPENDIX I: Total inflow assumed into Barra Bonita Reservoir (from AES)

Year	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
1969	123	126	133	100	61	82	61	58	53	132	410	229
1970	700	1294	674	305	248	195	145	137	225	193	163	251
1971	283	87	245	145	138	231	161	124	113	213	112	204
1972	522	885	375	251	164	140	238	214	203	668	391	337
1973	417	449	301	335	208	172	164	138	141	210	294	448
1974	737	409	754	372	219	291	191	132	116	159	125	411
1975	529	737	458	220	142	88	100	75	58	142	257	554
1976	531	937	712	498	470	698	675	551	675	702	546	581
1977	883	675	407	546	272	286	193	144	200	164	209	635
1978	390	270	345	118	154	171	175	78	95	100	421	422
1979	396	313	206	128	213	119	118	177	223	192	213	315
1980	418	611	362	371	165	174	140	122	117	95	106	450
1981	753	258	179	143	115	138	84	77	56	196	383	392
1982	556	661	538	289	179	486	384	242	169	471	461	1106
1983	1211	1576	1174	878	920	2269	760	514	941	795	738	806
1984	812	475	291	316	296	176	163	226	292	183	203	392
1985	433	461	693	348	291	207	187	155	193	130	164	129
1986	182	304	399	178	175	104	79	212	91	91	132	755
1987	709	946	475	260	599	724	303	235	251	253	273	353
1988	512	585	843	464	417	479	235	172	136	267	285	342
1989	1162	1133	582	357	228	214	237	395	235	168	210	244
1990	1012	330	535	266	233	164	270	202	206	239	252	247
1991	478	1070	1193	1032	529	341	272	203	157	463	215	430
1992	283	321	446	343	325	176	162	142	218	349	568	547
1993	636	983	588	426	329	420	217	186	455	415	276	306
1994	527	605	477	323	252	223	214	165	137	172	246	515
1995	702	1872	751	728	439	323	336	217	200	384	326	409
1996	982	650	1101	436	302	257	228	219	392	437	410	547
1997	1231	992	407	296	279	467	252	203	245	279	566	580
1998	453	884	813	469	467	302	239	234	248	481	229	678
1999	1341	1217	892	418	330	364	266	196	241	189	180	295
2000	579	688	464	253	175	162	199	197	341	189	381	613
2001	568	854	509	310	359	231	212	260	328	602	487	658
2002	824	589	615	190	418	149	193	359	297	267	631	572
Average	*643*	*713*	*557*	*356*	*297*	*324*	*231*	*205*	*237*	*294*	*319*	*463*
Minimum	*123*	*87*	*133*	*100*	*61*	*82*	*61*	*58*	*53*	*91*	*106*	*129*
Maximum	*1341*	*1872*	*1193*	*1032*	*920*	*2269*	*760*	*551*	*941*	*795*	*738*	*1106*

NOTE: From AES Tiete.

APPENDIX II: Transition Matrix of conditional probabilities (Discrete Inflow Event)

		Actual Inflow						
	1	1	2	3	4	5	6	7
Previous Inflow	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.250	0.000	0.250	0.250	0.250	0.000
	3	0.000	0.000	0.670	0.170	0.000	0.170	0.000
	4	0.000	0.250	0.250	0.250	0.250	0.000	0.000
	5	0.000	0.330	0.000	0.330	0.330	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.670	0.330	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	2							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.330	0.330	0.000	0.000	0.000	0.000
	3	0.000	0.000	0.400	0.400	0.000	0.200	0.000
	4	0.000	0.000	0.000	0.250	0.750	0.000	0.000
	5	0.000	0.200	0.200	0.200	0.200	0.200	0.000
	6	0.000	0.000	0.330	0.000	0.330	0.330	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	3							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.000	0.670	0.000	0.000	0.000	0.000
	3	0.000	0.200	0.200	0.200	0.200	0.200	0.000
	4	0.000	0.250	0.250	0.000	0.250	0.250	0.000
	5	0.000	0.000	0.200	0.600	0.200	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.670	0.000	0.330
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	4							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.670	0.000	0.330	0.000	0.000	0.000
	3	0.200	0.000	0.400	0.200	0.200	0.000	0.000
	4	0.000	0.000	0.750	0.000	0.000	0.250	0.000
	5	0.000	0.000	0.000	0.400	0.400	0.200	0.000
	6	0.000	0.000	0.000	0.000	0.670	0.330	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	5							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.330	0.000	0.330	0.000	0.000	0.000
	3	0.000	0.200	0.400	0.200	0.000	0.200	0.000
	4	0.000	0.000	0.500	0.000	0.500	0.000	0.000
	5	0.000	0.000	0.200	0.400	0.400	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.330	0.330	0.330
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	6							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.000	0.330	0.000	0.330	0.000	0.000
	3	0.000	0.200	0.600	0.000	0.000	0.200	0.000
	4	0.000	0.250	0.250	0.250	0.250	0.000	0.000
	5	0.000	0.000	0.000	0.600	0.400	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.330	0.670	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000

	7	Actual Inflow						
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.670	0.000	0.000	0.000	0.000	0.000
	3	0.000	0.000	0.400	0.200	0.400	0.000	0.000
	4	0.000	0.000	0.500	0.250	0.250	0.000	0.000
	5	0.000	0.000	0.200	0.400	0.400	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	8							
Previous Inflow	1	0.000	0.000	0.000	0.000	1.000	0.000	0.000
	2	0.330	0.330	0.000	0.330	0.000	0.000	0.000
	3	0.000	0.200	0.600	0.000	0.200	0.000	0.000
	4	0.000	0.250	0.500	0.250	0.000	0.000	0.000
	5	0.000	0.000	0.000	0.600	0.200	0.200	0.000
	6	0.000	0.000	0.000	0.000	0.330	0.330	0.330
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	9							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.330	0.330	0.000	0.000	0.000	0.000
	3	0.000	0.000	0.600	0.200	0.200	0.000	0.000
	4	0.000	0.000	0.200	0.400	0.400	0.000	0.000
	5	0.000	0.250	0.000	0.000	0.250	0.500	0.000
	6	0.000	0.000	0.000	0.330	0.330	0.000	0.330
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	10							
Previous Inflow	1	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	2	0.330	0.330	0.330	0.000	0.000	0.000	0.000
	3	0.000	0.200	0.200	0.200	0.400	0.000	0.000
	4	0.000	0.250	0.250	0.000	0.250	0.250	0.000
	5	0.000	0.000	0.200	0.400	0.200	0.200	0.000
	6	0.000	0.000	0.330	0.000	0.330	0.330	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	11							
Previous Inflow	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.330	0.000	0.330	0.000	0.000	0.330	0.000
	3	0.000	0.200	0.600	0.200	0.000	0.000	0.000
	4	0.000	0.000	0.250	0.250	0.500	0.000	0.000
	5	0.000	0.200	0.000	0.400	0.400	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.330	0.670	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	12							
Previous Inflow	1	0.000	0.000	0.000	0.000	1.000	0.000	0.000
	2	0.000	0.330	0.000	0.330	0.000	0.330	0.000
	3	0.200	0.200	0.200	0.200	0.000	0.200	0.000
	4	0.000	0.250	0.250	0.000	0.500	0.000	0.000
	5	0.000	0.000	0.600	0.200	0.200	0.000	0.000
	6	0.000	0.000	0.000	0.330	0.330	0.000	0.330
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000

NOTE: Values in the first column represent last period inflow (I_{n-1}) and, first line, current month inflows ($I_{n,k}$).

APPENDIX III: Transition Matrix of conditional probabilities (Fuzzy Inflow Event)

		Actual Inflow						
	1	1	2	3	4	5	6	7
Previous Inflow	1	0.975	0.013	0.000	0.003	0.010	0.000	0.000
	2	0.106	0.130	0.131	0.305	0.200	0.127	0.000
	3	0.008	0.056	0.439	0.336	0.025	0.114	0.022
	4	0.000	0.064	0.319	0.298	0.205	0.088	0.027
	5	0.000	0.083	0.122	0.277	0.450	0.069	0.000
	6	0.000	0.053	0.078	0.109	0.611	0.145	0.005
	7	0.000	0.000	0.000	0.000	0.102	0.561	0.338
	2							
Previous Inflow	1	0.257	0.463	0.280	0.000	0.000	0.000	0.000
	2	0.407	0.359	0.156	0.076	0.002	0.000	0.000
	3	0.000	0.193	0.315	0.364	0.074	0.055	0.000
	4	0.000	0.000	0.021	0.383	0.493	0.099	0.004
	5	0.006	0.193	0.258	0.204	0.172	0.151	0.016
	6	0.000	0.114	0.213	0.142	0.109	0.316	0.106
	7	0.000	0.000	0.000	0.000	0.101	0.599	0.300
	3							
Previous Inflow	1	0.031	0.825	0.144	0.000	0.000	0.000	0.000
	2	0.332	0.124	0.363	0.147	0.034	0.000	0.000
	3	0.045	0.194	0.223	0.138	0.348	0.053	0.000
	4	0.000	0.068	0.336	0.272	0.231	0.093	0.000
	5	0.000	0.000	0.217	0.412	0.245	0.045	0.082
	6	0.000	0.000	0.000	0.173	0.461	0.106	0.260
	7	0.000	0.000	0.000	0.008	0.173	0.265	0.554
	4							
Previous Inflow	1	0.126	0.825	0.049	0.000	0.000	0.000	0.000
	2	0.128	0.471	0.032	0.286	0.083	0.000	0.000
	3	0.172	0.106	0.192	0.195	0.282	0.053	0.000
	4	0.000	0.078	0.483	0.257	0.134	0.049	0.000
	5	0.000	0.000	0.072	0.329	0.509	0.089	0.000
	6	0.000	0.000	0.000	0.008	0.463	0.272	0.257
	7	0.000	0.000	0.000	0.000	0.000	0.336	0.665
	5							
Previous Inflow	1	0.000	0.369	0.386	0.245	0.000	0.000	0.000
	2	0.300	0.328	0.212	0.160	0.000	0.000	0.000
	3	0.019	0.266	0.341	0.171	0.027	0.160	0.015
	4	0.000	0.009	0.199	0.370	0.353	0.063	0.006
	5	0.000	0.029	0.182	0.366	0.292	0.131	0.000
	6	0.000	0.000	0.000	0.087	0.350	0.298	0.266
	7	0.000	0.000	0.000	0.000	0.085	0.612	0.303
	6							
Previous Inflow	1	0.000	0.496	0.430	0.055	0.019	0.000	0.000
	2	0.321	0.057	0.264	0.272	0.086	0.000	0.000
	3	0.079	0.266	0.398	0.056	0.122	0.080	0.000
	4	0.005	0.150	0.300	0.324	0.209	0.013	0.000
	5	0.000	0.002	0.223	0.396	0.243	0.136	0.000
	6	0.000	0.000	0.002	0.001	0.367	0.593	0.038
	7	0.000	0.000	0.000	0.000	0.000	0.150	0.850

Previous Inflow	7	Actual Inflow						
	1	0.232	0.670	0.098	0.000	0.000	0.000	0.000
	2	0.347	0.302	0.140	0.046	0.164	0.002	0.000
	3	0.048	0.084	0.425	0.208	0.225	0.010	0.000
	4	0.000	0.032	0.394	0.423	0.150	0.000	0.000
	5	0.000	0.000	0.051	0.432	0.428	0.085	0.004
	6	0.000	0.000	0.000	0.041	0.476	0.285	0.198
	7	0.000	0.000	0.000	0.000	0.007	0.142	0.851

Previous Inflow	8							
	1	0.095	0.286	0.000	0.110	0.509	0.000	0.000
	2	0.332	0.257	0.126	0.248	0.037	0.000	0.000
	3	0.020	0.155	0.563	0.109	0.149	0.004	0.000
	4	0.000	0.139	0.513	0.199	0.109	0.040	0.000
	5	0.000	0.000	0.024	0.272	0.587	0.118	0.000
	6	0.000	0.000	0.000	0.035	0.487	0.170	0.308
	7	0.000	0.000	0.000	0.000	0.000	0.409	0.591

Previous Inflow	9							
	1	0.875	0.125	0.000	0.000	0.000	0.000	0.000
	2	0.374	0.250	0.376	0.000	0.000	0.000	0.000
	3	0.000	0.017	0.532	0.202	0.248	0.001	0.000
	4	0.000	0.029	0.237	0.335	0.399	0.000	0.000
	5	0.000	0.084	0.119	0.299	0.438	0.060	0.000
	6	0.000	0.000	0.011	0.045	0.470	0.139	0.334
	7	0.000	0.000	0.000	0.000	0.000	0.420	0.580

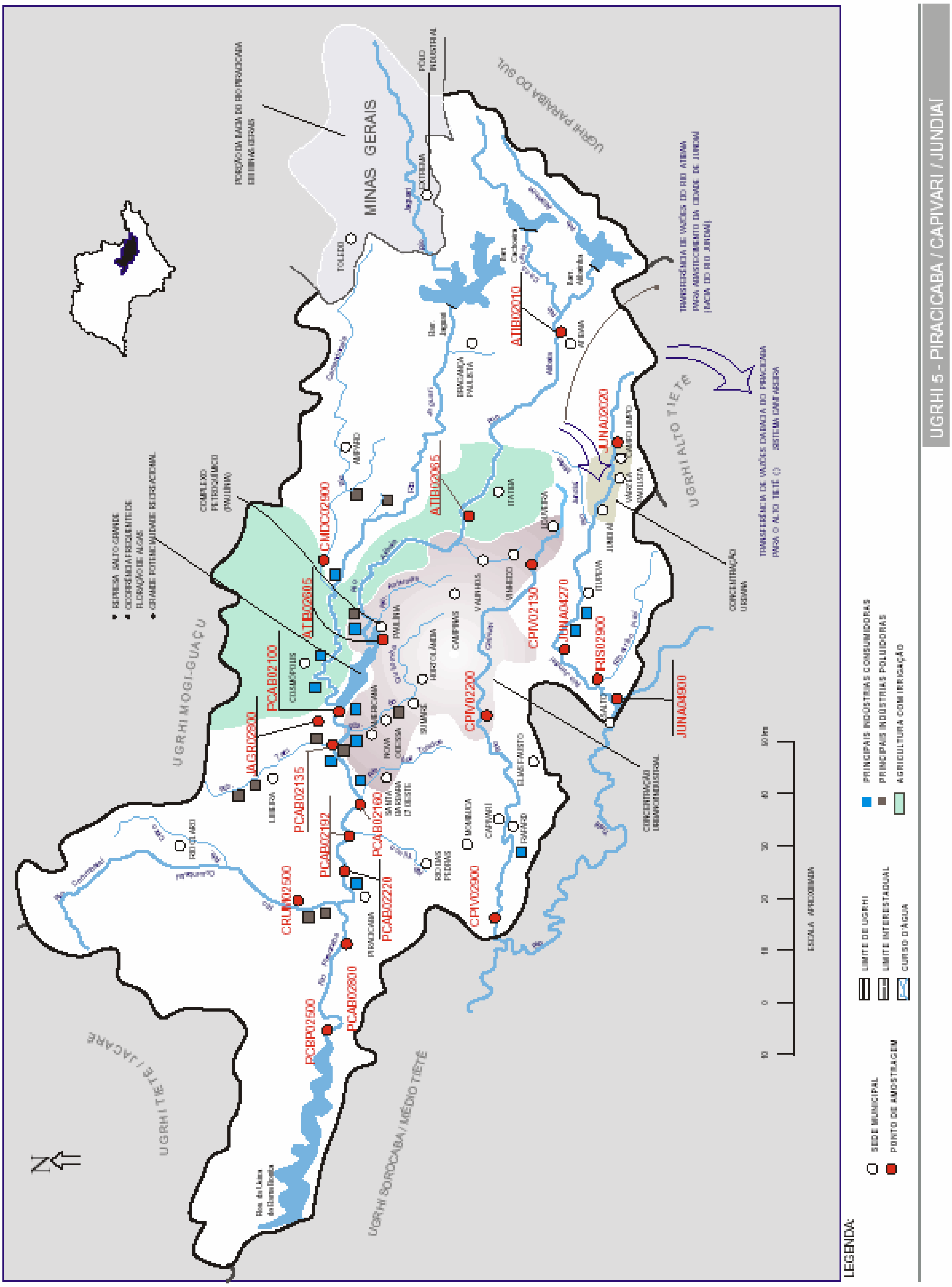
Previous Inflow	10							
	1	0.000	0.148	0.325	0.520	0.008	0.000	0.000
	2	0.371	0.433	0.097	0.081	0.019	0.000	0.000
	3	0.126	0.210	0.168	0.202	0.245	0.050	0.000
	4	0.000	0.084	0.205	0.111	0.328	0.272	0.000
	5	0.000	0.022	0.275	0.436	0.190	0.077	0.000
	6	0.000	0.000	0.115	0.168	0.082	0.511	0.123
	7	0.000	0.000	0.000	0.000	0.000	0.449	0.551

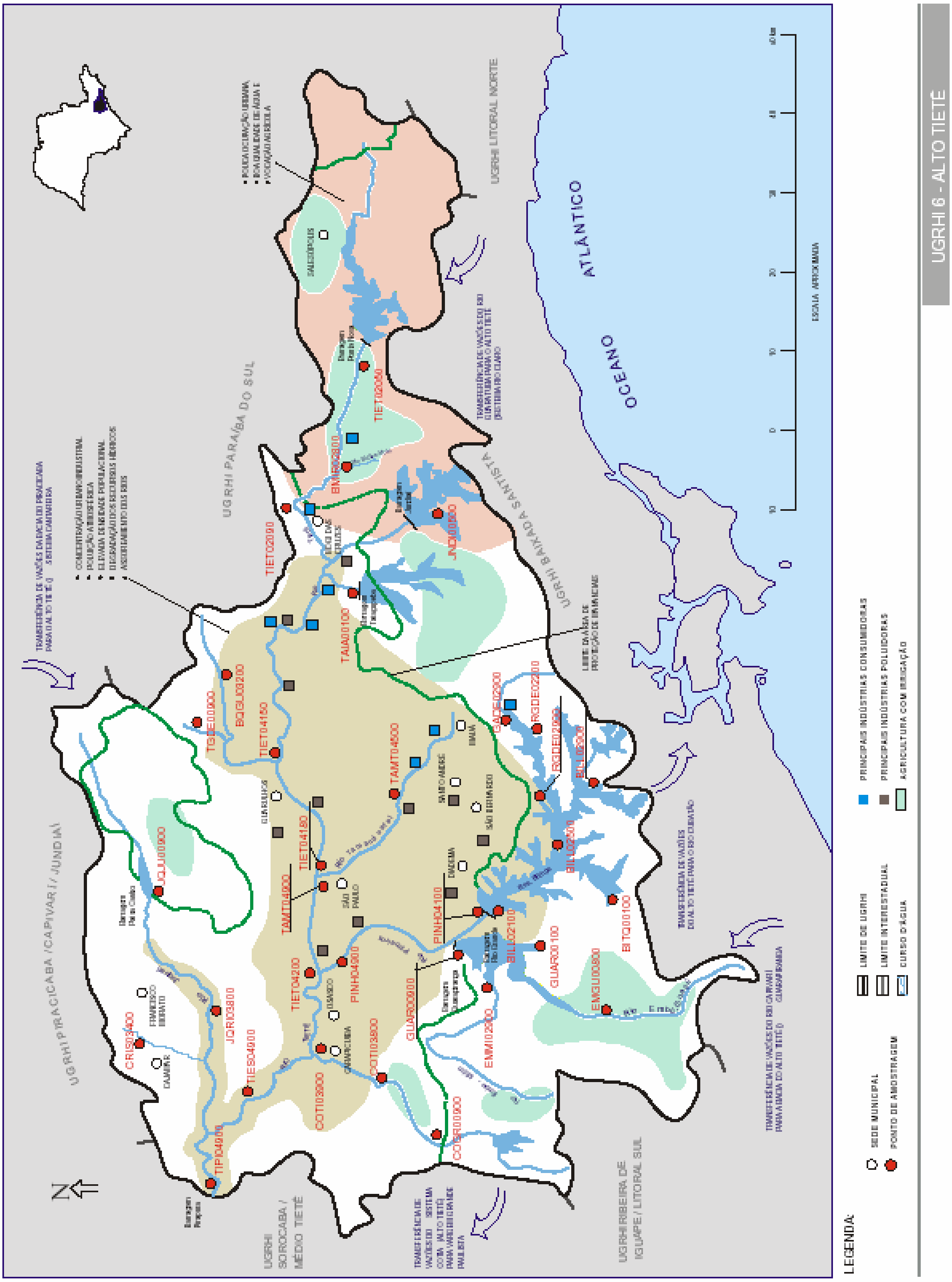
Previous Inflow	11							
	1	0.279	0.581	0.140	0.000	0.000	0.000	0.000
	2	0.225	0.132	0.146	0.095	0.216	0.187	0.000
	3	0.000	0.234	0.360	0.365	0.037	0.004	0.000
	4	0.064	0.118	0.272	0.277	0.223	0.047	0.000
	5	0.033	0.033	0.048	0.495	0.296	0.094	0.000
	6	0.000	0.000	0.059	0.098	0.256	0.469	0.119
	7	0.000	0.000	0.000	0.000	0.000	0.231	0.769

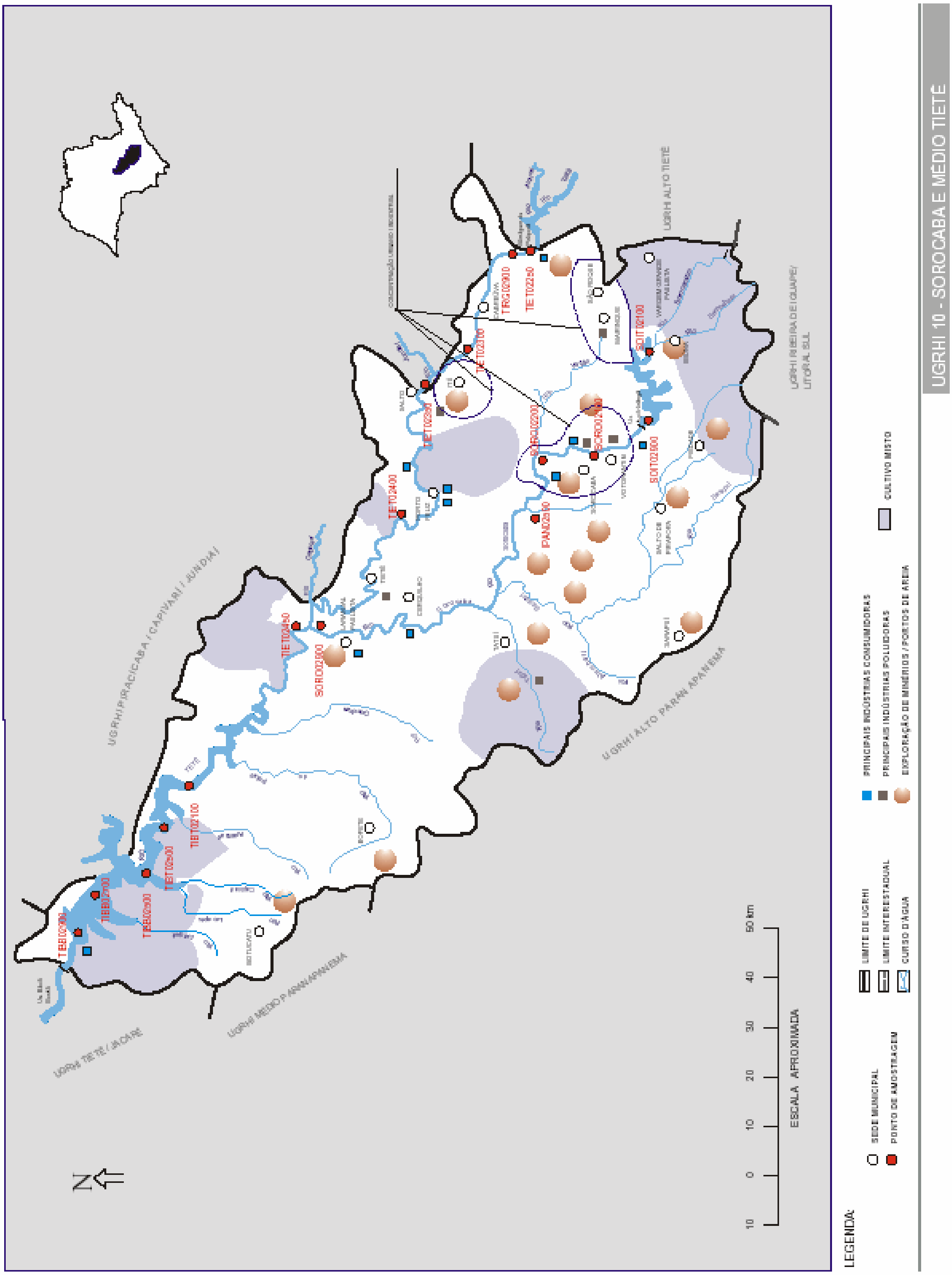
Previous Inflow	12							
	1	0.073	0.277	0.000	0.338	0.312	0.000	0.000
	2	0.140	0.219	0.029	0.297	0.066	0.249	0.000
	3	0.169	0.231	0.180	0.221	0.097	0.103	0.000
	4	0.000	0.194	0.289	0.232	0.219	0.066	0.000
	5	0.000	0.033	0.336	0.420	0.135	0.020	0.056
	6	0.000	0.000	0.109	0.241	0.252	0.169	0.229
	7	0.000	0.000	0.000	0.000	0.109	0.835	0.056

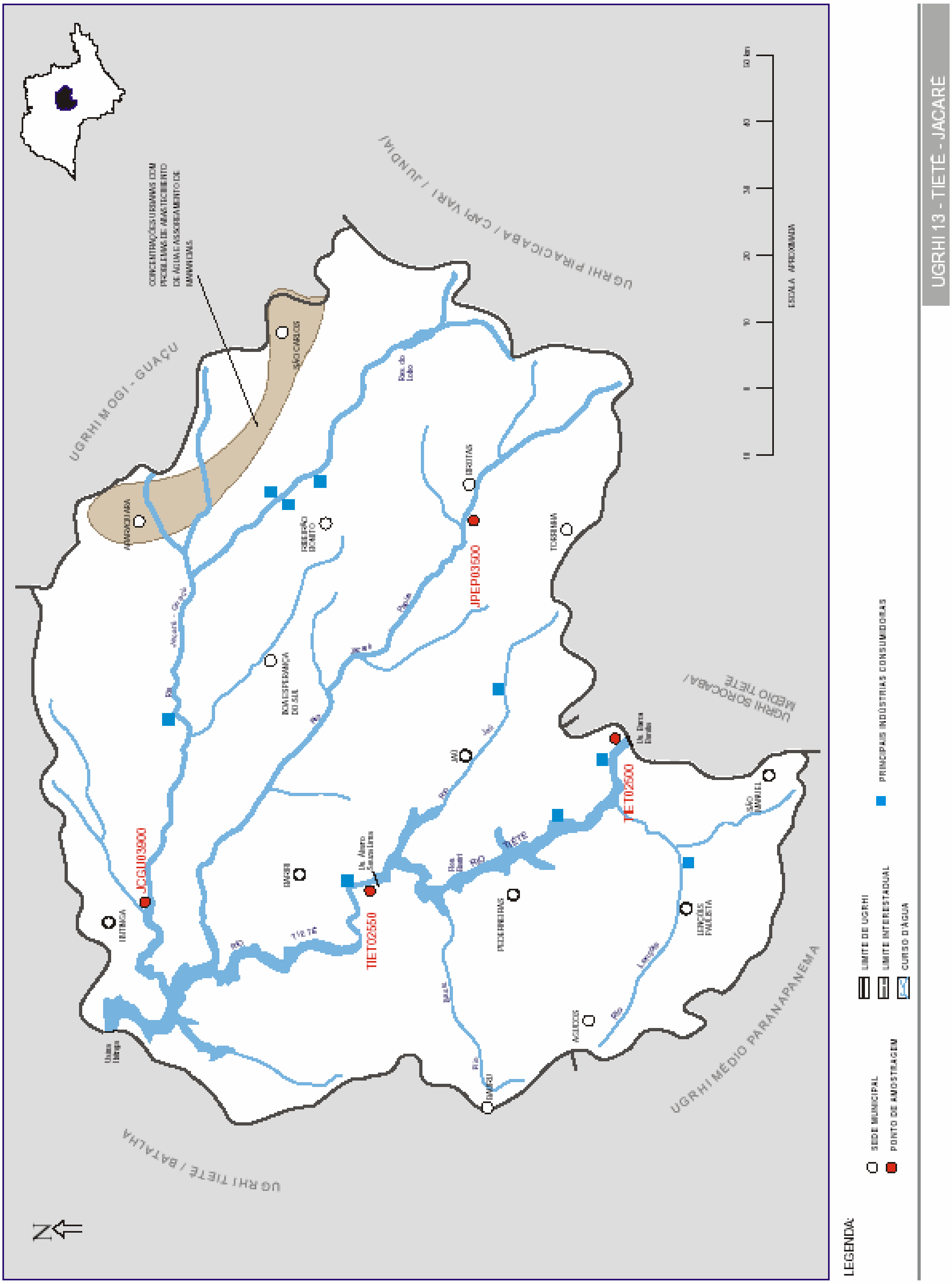
NOTE: Values in the first column represent last period inflow (I_{n-1}) and, first line, current month inflows ($I_{n,k}$).

APPENDIX IV: Attribute Maps of Tiete river Sub-basins (from CETESB)




UGRHI 6 - ALTO TIETÊ
OCEANO
ATLÂNTICO
UGRHI PARAÍBA DO SUL
UGRHI LITORAL NORTE
UGRHI BAIXADA SANTISTA
UGRHI PIRACICABA / CAPIVARI / JUNDIAÍ
UGRHI SOROCABA / MÉDIO TIETÊ
UGRHI RIBEIRA DE IGUAPE / LITORAL SUL
LEGENDA:
SEDE MUNICIPAL
PONTO DE AMOSTRAGEM
LIMITE DE UGRHI
LIMITE INTERESTADUAL
CURSO D'ÁGUA
PRINCIPAIS INDÚSTRIAS CONSUMIDORAS
PRINCIPAIS INDÚSTRIAS POLUIDORAS
AGRICULTURA COM IRRIGAÇÃO
ESCALA APROXIMADA


UGRHI 10 - SOROCABA E MÉDIO TIETÊ
UGRHI PIRACICABA / CAPIVARI / JUNDIAÍ
UGRHI ALTO TIETÊ
UGRHI RIBEIRA DE IGUAPE/ LITORAL SUL
UGRHI ALTO PARANAPANEMA
UGRHI MÉDIO PARANAPANEMA
UGRHI TIETÊ / JACARÉ
N
ESCALA APROXIMADA
10 0 10 20 30 40 50 km
LEGENDA:
SEDE MUNICIPAL
PONTO DE AMOSTRAGEM
LIMITE DE UGRHI
LIMITE INTERESTADUAL
CURSO D'ÁGUA
PRINCIPAIS INDÚSTRIAS CONSUMIDORAS
PRINCIPAIS INDÚSTRIAS POLUIDORAS
EXPLORAÇÃO DE MINÉRIOS / PORTOS DE AREIA
CULTIVO MISTO


UGRHI 13 - TIETÊ - JACARÉ
UGRHI MOGI - GUAÇU
UGRHI PIRACICABA / CAPIVARI / JUNDIAÍ
UGRHI MÉDIO TIETÊ / SOROCABA
UGRHI MÉDIO PARANAPANEMA
UGRHI TIETÊ / BATALHA
CONCENTRAÇÕES URBANAS COM PROBLEMAS DE ABASTECIMENTO DE ÁGUA E ASSOREAMENTO DE MANANCIAIS
SÃO CARLOS
ARARAQUARA
RIBEIRÃO BONITO
BROTAS
TORRINHA
JPEP03500
BOA ESPERANÇA DO SUL
JAÚ
JCGU03900
IBITINGA
Usina Ibitinga
Us. Álvaro Souza Lima
TIET02550
TIET02500
Us. Barra Bonita
PEDERNEIRAS
LENÇÓIS PAULISTA
AGUDOS
BAURU
SÃO MANUEL
ESCALA APROXIMADA
LEGENDA:
SEDE MUNICIPAL
PONTO DE AMOSTRAGEM
LIMITE DE UGRHI
LIMITE INTERESTADUAL
CURSO D'ÁGUA
PRINCIPAIS INDÚSTRIAS CONSUMIDORAS

APPENDIX V: Photos from Inspection to the Reservoir site (March, 2001)

Algae-bloom in the Barra Bonita reservoir near to the dam site (by Paulo Chaves)

Boat used for the water quality data collection (by Paulo Chaves)

Reservoir covered by aquatic plants near to the Dam structures (by Paulo Chaves)

Paulo Chaves and Prof. Jose Tundisi by the Reservoir (by Yosuke Yamashiki)

APPENDIX VI: Other publications by the author

PUBLICATIONS:

P. Chaves, T. Kojiri: “Stochastic reservoir operation considering water quality objectives using evolving neuro-fuzzy systems” – *under elaboration.*

P. Chaves, T. Kojiri: “Stochastic Fuzzy Neural Network: a case study of reservoir operation” – *under elaboration.*

P. Chaves, T. Kojiri: “Conceptual Fuzzy Neural Network for Water Quality Simulation” – *under elaboration.*

P. Chaves, T. Kojiri: “Quality of Water in Storage”. The Encyclopedia of Water. John Wiley and Sons, Inc. *Accepted and publication expected for 2005.*

P. Chaves, T. Tsukatani, T. Kojiri: “Operation of Storage Reservoir for Water Quality by using Optimization and Artificial Intelligence Techniques”, Mathematics and Computers in Simulation, Vol. 67 (4-5), pp. 419-432, 2004.

P. Chaves, T. Kojiri, Y. Yamashiki: “Optimization of Storage Reservoir considering Water Quantity and Quality”, Hydrological Processes Journal, Vol. 17 (15), pp. 2769-2793, October 2003.

P. Chaves, T. Kojiri: “Multi-objective Storage Reservoir Operation under Uncertainty”, Annuals of Disaster Prevention Research Institute, No. 46 B, pp. 899-918, Kyoto University, Japan, 2003.

P. Chaves: “Planning Operation of Storage Reservoir for Water Quantity and Quality”, Master Thesis, Civil Engineering Systems course, Water Resource Research Center, Graduate School of Engineering, Kyoto University, 2002.

CONFERENCE PAPERS & OTHERS PUBLICATIONS:

P. Chaves, T. Kojiri: “Reservoir Optimization through Stochastic Dynamic Programming with Fuzzy Conditional Probability”, Proceedings of the 2004 Annual Conference of the Japanese Society of Hydrology and Water Resources, pp. 32-33, Hokkaido, Japan, August, 19th to 21st, 2004. (Participation sponsored by the Water Resources Research Center, Kyoto University)

P. Chaves, T. Kojiri: “Introduction of Conceptual Fuzzy Neural Network into Hydrological Modeling”, Proceedings of the 6th International Conference on Hydro-Science and Engineering, ICHE-2004, pp. 477-489. Brisbane, Australia, May 31st to June 3rd, 2004.

P. Chaves, T. Tsukatani, T. Kojiri: “Operation of Storage Reservoir for Water Quality by using Optimization and Artificial Intelligence Techniques”, IMACS International Conference: Mathematical Modeling of Ecological Systems, Almaty, Kazakhstan, September 09th to 12th, 2003. (Participation sponsored by the International Institute of Economics, Kyoto University);

P. Chaves, T. Kojiri: “Storage Reservoir Planning considering Water Quantity and Quality”, Proceeding of the 13th IAHR-APD Congress, Vol. 2, pp. 1042-1047, Singapore, August 7th , 2002. (Participation sponsored by the Water Resources Research Center, Kyoto University)

P. Chaves, T. Kojiri: “Intergraded Water Resources Management: Formulation of Multi-objective Optimization on Water Quantity and Quality of Storage Reservoir”, Proceedings of the III International Workshop on regional approaches to reservoir development and management in the La Plata River Basin; Annex VI, pp. 71-82, Posadas, Argentina, 14th – 17th March 2001. (Participation sponsored by UNEP-IETC & ILEC)

P. Chaves, A. Teixeira: “Water Supply, Sewage and Drainage Systems for a Low-income Community in Rio de Janeiro, Brazil”, Undergraduate Final Project, Federal University of Rio de Janeiro, Brazil, 1999.

CONTACT INFORMATION:

E-mail: **phchaves@yahoo.com**