A METHODOLOGY FOR VALIDATION OF INTEGRATED SYSTEMS MODELS WITH AN APPLICATION TO COASTAL-ZONE MANAGEMENT IN SOUTH-WEST SULAWESI

Tien Giang Nguyen

Promotion committee:

Prof. dr.ir. H.J. Grootenboer	University of Twente, chairman/secretary
Prof. dr. P.G.E.F. Augustinus	University of Utrecht, promoter
Prof. dr. S.J.M.H. Hulscher	University of Twente, promoter
Dr. J.L. de Kok	University of Twente, assistant promoter
Prof. dr.ir. A.Y. Hoekstra	University of Twente
Prof. dr.ir. H.G. Wind	University of Twente
Prof. dr.ir. A.E. Mynett	UNESCO-IHE / WL | Delft Hydraulics
Prof. dr. S.M. de Jong	University of Utrecht
Dr. M.J. Titus	University of Utrecht

ISBN 90-365-2227-7

Printed by: PrintPartners Ipskamp, Enschede

A METHODOLOGY FOR VALIDATION OF INTEGRATED SYSTEMS MODELS WITH AN APPLICATION TO COASTAL-ZONE MANAGEMENT IN SOUTH-WEST SULAWESI

DISSERTATION

to obtain
the doctor's degree at the University of Twente,
on the authority of the rector magnificus,
prof.dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Friday August 26, 2005 at 15.00

by

Tien Giang Nguyen

born on April 12, 1976

in Hanoi

This dissertation has been approved by:

prof. dr. P.G.E.F. Augustinus	Promoter
prof. dr. S.J.M.H. Hulscher	Promoter
dr. J.L. de Kok	Assistant Promoter

To the memory of my father

Contents

Preface

Words are easily borrowed, but the emotional meaning from one's heart is difficult to describe. Therefore, I would like to ask for the forgiveness from those whose assistance cannot be appreciated in words and from those who, by any chance, I forgot to mention.

First of all I would like to thank prof. dr.ir. Herman Wind and his wife - Joke. The interviews prof. Wind held in Bangkok and the decision he made enabled me to be here, at Twente University, to fulfil my PhD research. *Herman and Joke, I will never forget the first meeting we had in Bangkok in April, 2001. Khap khun ma khap.*

The next person I want to thank is dr. Jean-Luc de Kok, my daily supervisor. His cleverness and intellectual skills have convinced me that I would be able to complete this thesis with his regular guidance. His tireless support during the four-year period of the research, in every aspect, made 'my wish' to come true. *Jean-Luc, I am very happy to be your first PhD student. The contents of the thesis belong to both you and me.*

Importantly, I would like to thank my two promoters, prof. dr. Suzanne Hulscher and prof. dr. Pieter Augustinus and my former promoter, prof. dr. Kees Vreugdenhil. Their outstanding knowledge and experience in both science and management resulted in this thesis. *I would like to thank you all for that. Specifically, thank you, Pieter, for your kindness, patience and useful comments during the preparation of the manuscript.*

I would like to thank the Netherlands Foundation for the Advancement of Tropical Research (WOTRO). The funding given to both the original Buginesia project and the resulting project, which is presented in this thesis, has been provided by this organisation.

It is also necessary to mention in particular a number of people from different institutions in the Netherlands and in Indonesia, who have been actively involved in the Buginesia project and contributed to this research. From Utrecht University, dr. Milan Titus, prof. dr. Steven De Jong, prof. dr. Piet Hoekstra; From ITC, dr. A. Sharifi and dr. Tjeerd Hobma; From UNESCO-IHE, prof. dr. ir. Arthur Mynett; From other Dutch institutions: dr. Lida Pet-Soede, dr. Maya Borel Best, dr. Wim van Densen and prof. dr. Leontine Visser; From the University Hasanuddin (UNHAS), prof. dr. Dadang Ahmad, prof. dr. Alfian Noor and mr. Mushta;. *I have learnt a lot from you all. Thank you very much for your fruitful cooperation.*

Social interaction plays an important role in one's research career. Therefore, I want to thank all colleagues inside and outside of the WEM group for making my working years here lively. Particularly, Huong Thi Thuy Phan; two pretty office-mates: Saskia and

Schretlen; people in the soccer team of the WEM group: Jan, Daniel, Martijn Booij, Maarten, Pieter Roos, Jebbe, Pieter Oel, Andrei, Freak, Judith, Daniella, Steven,…and their partners; Roos, Rolnan, Cornelie, Judith and Wendy from the Construction and Transport groups; Our secretaries: Anke, Joke and Ellen; Yan, Dong, Ping, Chang Wei and Jan from the 'Chinese community'. *Dank je wel and Xie xie*.

The work of preparing, distributing, and collecting the questionnaires from the end-users of RaMCo (Chapter 3) was done by Tessa Hofman. Arif Wismadi collected the socio-economic data for validating the model of land-use and land-cover changes (Chapter 4). Christian Loris collected the data and processed a part of them for validating the fisheries model (Chapter 5). Gay Howells checked the English of the final manuscript. *Thank the four of you for what you have done to make this thesis complete.*

My gratefulness goes to all of my Vietnamese friends living in the Netherlands, but particularly to Hien-Nhu, Hieu-Lam, Kim-V.Anh, Phuong-Ha, Tu-An, Duy-Chi, Thang-Mai, Trung-Thanh, Ha-Huong, Nhung, Hanh, Long, Kien, Hoa and Cuong. *Dear friends, to be your friends in Twente makes me feel like being at home.*

Hoang Tu and Jebbe van der Werf deserve the special thanks for what makes me ask them to be my 'paranimfen'. *Dank je wel*, *Jebbe, you are my closest Dutch friend*. *Tu, many thanks for the daily-life things we have been sharing*.

Almost the last person I want to thank here is my beloved girlfriend - Hue Chi. She is always with me when I need her most. The four-year period of doing research would have been much more difficult without her. *Darling, I love you so very much*.

Lastly but most importantly, I would like to express my deepest gratefulness to my family, including my father, Nguyen Dinh Thinh, my mother, Ly Thi Nguyet and my little sister, Nguyen Thanh Thuy. They are the people that support me the most. *Father! One paragraph of thankful words is absolutely far from enough for what you had done for me. The whole spirit of this thesis is devoted to you. In Heaven, you would be smiling….*

Nguyen Tien Giang

Enschede, July, 2005.

Chapter 1

Introduction

1.1. General introduction

Systems approach and integrated approach towards the planning and management of natural resources and environment are considered as promising approaches to achieve the sustainable development of a region, of a country, and of our common world. Consequently, an increasing number of Integrated Systems Models (ISMs) have been developed (e.g. Hoekstra, 1998; Turner, 2000; De Kok and Wind, 2002). However, the scarcity of field data for both model development and model validation, the lack of knowledge about the relevant internal and external factors of the real system and the model high aggregation levels (increase in scope but decrease in detail) create a number of critical questions such as: to what extent can such models contribute to our knowledge and ability to manage our environment? Are they useful and do they have an added value in comparison with conventional process models? Centred in these questions are the two questions: What is the validity of an ISM? How can this validity be determined and established? This thesis is aimed at addressing these two questions.

Rapid Assessment Model for Coastal-zone Management (RaMCo), which was developed by a Dutch-Indonesian multidisciplinary team (De Kok and Wind, 1999), serves as a case study to achieve the objective of the thesis. The theoretical justification for this choice is that RaMCo contains the typical characteristics of an Integrated Systems Model. The first characteristic is reflected in the RaMCo's ability to take into account the interactions of socio-economic developments, biophysical conditions and policy options. The second characteristic is the inclusion of the linkages between many processes pertaining to different scientific fields, such as marine pollution, land-use change, catchment hydrology, coastal hydrodynamics, fisheries and regional economic development. Practically, the model was chosen since its validation had not been carried out in the original project. In addition, the availability of the measured data (from 1996 until now) allows for the application of quantitative techniques which are suitable for the validation of ISMs. It is aware that, despite the typicality of RaMCo, other ISMs may differ in some aspects from the model considered. Therefore, the generality of the validation methodology established is discussed in the final chapter of the thesis.

The introductory chapter is organised as follows. Section 1.2 describes the concepts of systems approach, integrated approach and how they fit into the framework of the natural resources and environmental management. The role of ISMs as tools to facilitate this integrated management is explained. Difficulties involved with validation of these models are elaborated in Section 1.3. The research questions and sub-questions of the thesis are formulated in Section 1.4. A description of the case study is given in Section 1.5. The outline of the thesis is included in Section 1.6.

1.2. Background

1.2.1. Systems approach

Systems approach or systemic approach was born from the cross-fertilization of several disciplines: information theory (Shannon, 1948), cybernetics (Wiener, 1948), and general systems theory (Von Bertalanffy, 1968) more than half a century ago. As described by Rosnay (1979), it is not to be considered a "science," a "theory," or a "discipline," but *a new methodology that makes possible the collection and organization of accumulated knowledge in order to increase the efficiency of our actions.*

The systemic approach, as opposed to the analytical approach, includes the totality of the elements in the system under study, as well as their interaction and interdependence. It is based on the conception of systems. The systems approach got its well-known status after the two publications related to the depletion of world’s natural resources (Forrester, 1971; Meadows et al., 1972). To clarify the concept of systems approach, others approaches, with which it is often confused, are briefly mentioned.

- The systemic approach goes beyond the *cybernetics approach* (Wiener, 1948), whose main objective is the study of control in living organisms and machines.

- It must be distinguished from *General Systems Theory* (Von Bertalanffy, 1968), whose purpose is to describe in mathematical language the totality of systems found in nature.

- It is not the same as *systems analysis* (Miser and Quade, 1985), a method that represents only one tool of the systemic approach. The system analysis is elaborated later in Section 1.2.3.

- The systemic approach has nothing to do with a *systematic approach* that confronts a problem or sets up a series of actions in sequential manner, in a detailed way, forgetting no element and leaving nothing to chance.

The analytic approach seeks to reduce a system to its elementary elements in order to study them in detail and understand the types of interaction that exist between them. By modifying one variable at a time, it tries to infer general laws that will enable to predict the properties of a system under very different conditions. To make this prediction possible, the laws of the additivity of elementary properties must be invoked. This is the case in homogeneous systems, those composed of similar elements and having weak interactions among them. Here the laws of statistics readily apply, enabling to understand the behaviour of the disorganized complexity. The laws of the additivity of elementary properties do not apply in highly complex systems composed of a large diversity of elements linked together by strong interactions. These systems must be approached by new methods such as those that the systemic approach groups together. The purpose of the new methods is to consider a system in its totality, its complexity, and its own dynamics. Through simulation one can "animate" a system and observe in real time the effects of the different kinds of interactions among its elements. The study of this behaviour leads in time to the determination of rules that can modify the system or design other systems.

Systems Concepts

Various definitions of concepts of systems can be found in the literature (see Van Gigch, 1974; Rosnay, 1979; Kramer and De Smit, 1991). Following Kramer and De Smit (1991), a *system* is defined as a collection of entities together with the collection of relationships existing between these entities. An *entity* (element) is the component of the system. In principle any system can be decomposed into *subsystems,* a process which can be repeated as many times as the number of distinguishable hierarchic or *aggregation levels* the system comprises. The entities of the system at a lower hierarchic level and their interrelationships constitute the subsystems at that level. The choice of system entities simultaneously fixes the level of aggregation, and is not a trivial matter. In principle the level of aggregation depends on the purpose of the system model.

The *structure* of a system is also differently defined in the literature. A structure of a system, in view of systems modelling, can comprise: a spatial arrangement of elements, ordered levels (hierarchy) of subsystems or/and elements, and concentration and types of algebraic relationships between subsystems and/or elements. These three factors, together with the variety of elements (related to ordered levels), determine the *complexity* of a system. An extremely complex system model can be characterized by a rich variety of elements, a heterogeneous and irregular distribution of elements in space, many hierarchic levels, and nonlinear algebraic relationships between the elements. The complexity of a system is dependent on its nature and its boundaries.

The *boundaries* of a system separate the system from its environment. There are two types of boundary: physical and conceptual boundaries. The physical boundary determines the spatial scope of the system (e.g. a coastal zone) while the conceptual boundary differentiates exogenous from endogenous variables. *Exogenous* (i.e. external or independent) variables are those whose values arise independently of the endogenous (i.e. internal) variables. *A closed system* is a self-contained system without connections to exogenous variables. Oreskes et al. (1994), in arguing against the possibility of validating predictive models, indicate that an open system is a system which is not well defined (uncertain parameters, state variables, boundaries, etc.). Examples of open systems are: groundwater systems, social systems, as well as most of the natural systems.

Four types of variables characterize a model of a system (Kramer and De Smit, 1991): input variables, state variables, control variables, and output variables. The output variables of a system depend on the structure of the system (e.g. a transfer function) together with the input variables, control variables and state variables. Considering a system element with an input variable x(t), a state variable s(t), a control variable c(t) and an output variable y(t) as shown in Fig.1.1, the dynamic (time dependent) behaviour of this system element is governed by the following equations:

$$y(t) = f\left(x(t), c(t), s(t)\right) \tag{1.1}$$

$$\frac{\partial s(t)}{\partial t} = g\left(x(t), c(t), s(t)\right) \quad (1.2.)$$

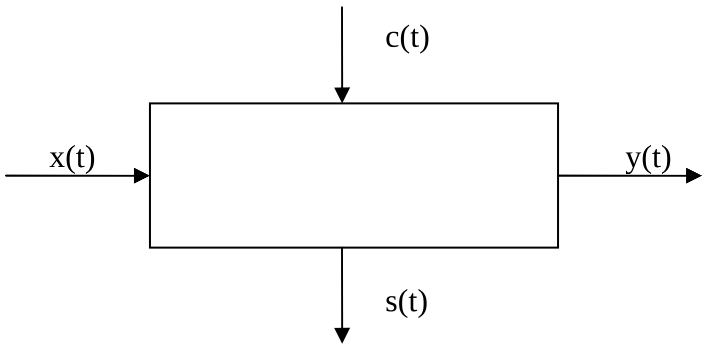

Fig. 1.1. General model of a system (Kramer and De Smit, 1991)

System dynamics
System Dynamics (SD) is a modelling approach which considers the structural system as a whole, focusing on the dynamic interactions between the components as well as on the behaviour of the complete system. SD was generalized from Industrial Dynamics (Forrester, 1961) and Urban Dynamics (Forrester, 1969), developed by Jay W. Forrester, at the Massachusetts Institute of Technology. This discipline is based on systems theory, control theory and the modern theory of nonlinear dynamics. There are some important concepts relevant to system dynamics: feedback, stocks and flows, mode and behaviour, time delays, and nonlinearity (Sterman, 2002) which require elaboration.

Positive and Negative Feedback

In a system where a transformation occurs, there are inputs and outputs. The inputs are the result of the environment's influence on the system, and the outputs are the influence of the system on the environment. Input and output are separated by duration of time, as in before and after, or past and present (Fig. 1.2).

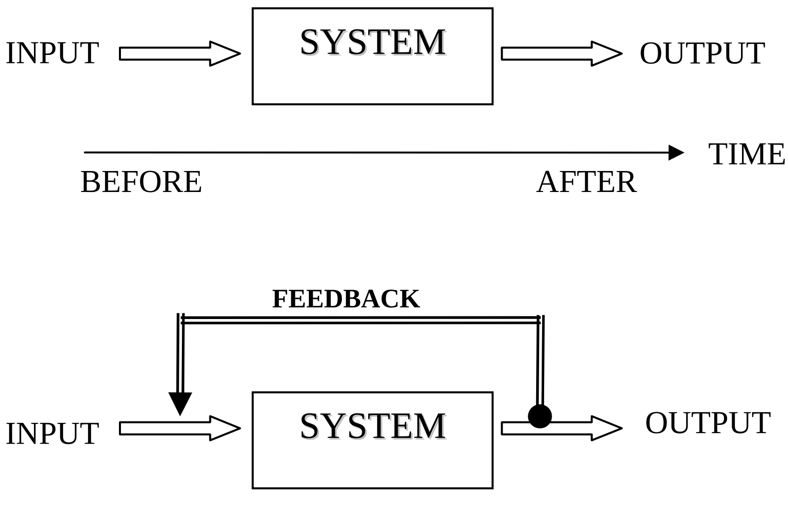

Fig. 1.2. System input-output and feedback (Rosnay, 1979)

In every feedback loop, as the name suggests, information about the result of a transformation or an action is sent back to the input of the system in the form of input data. If these new data facilitate and accelerate the transformation in the same direction as the preceding results, they are positive feedback; their effects are cumulative. If the new data produce a result in the opposite direction to previous results, they are negative feedback; their effects stabilize the system. In the first case there is exponential growth or decline; in the second case the equilibrium can be reached (Fig. 1.3).

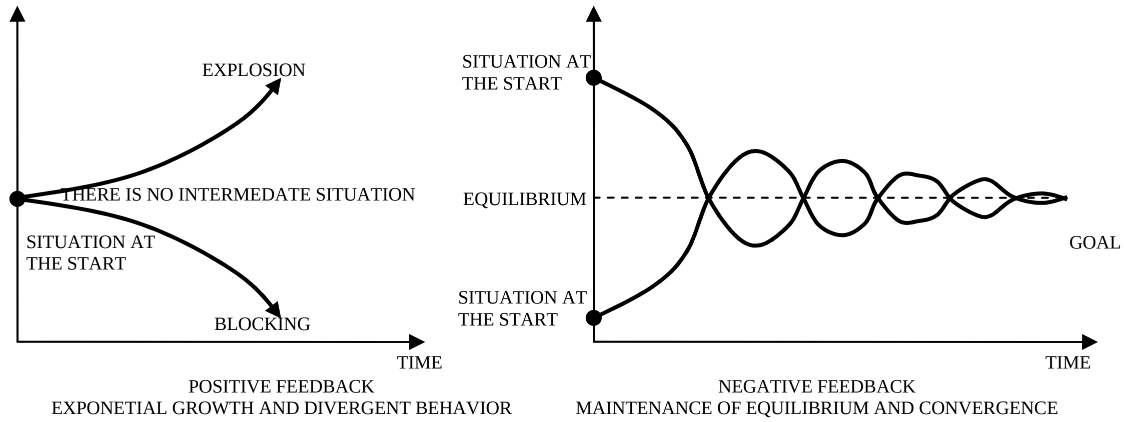

Fig. 1.3. Positive and negative feedback (Rosnay, 1979)

Positive feedback leads to divergent behaviour: indefinite expansion or explosion (a running away toward infinity) or total blocking of activities (a running away toward zero). Each plus involves another plus; it causes a snowball effect. Some examples are the population growth, industrial expansion, capital invested at compound interest, inflation, and proliferation of cancer cells. However, when minus leads to another minus, events come to a standstill. Typical examples are bankruptcy and economic depression.

Stocks and flows

The dynamic behaviour of every system, regardless of its complexity, depends ultimately on two kinds of variables: flow variables and state variables. The first are symbolized by the valves that control the flows, the second (showing what is contained in the reservoirs) by rectangles. The flow variables are expressed only in terms of two instants, or in relation to a given period, and thus are basically functions of time. The state (level) variables indicate the accumulation of a given quantity in the course of time; they express the result of integration. If time stops, the level remains constant (static level) while the flows disappear - for they are the results of actions. Hydraulic examples are the easiest to understand. The flow variable is represented by the flow rate, that is, the average quantity running off between two instants. The state variable is the quantity of water accumulated in the reservoir at a given time. If the flow of water is replaced by a flow of people (number of births per year), the state variable becomes the population size at a given moment.

Modes and behaviour of systems

The properties and the behaviour of a complex system are determined by its internal organization and its relations with its environment. To understand better these properties and to anticipate better its behaviour, it is necessary to act on the system by transforming it or by orienting its evolution.

Every system has two fundamental modes of existence and behaviour: *maintenance* and *change*. The first mode, based on negative feedback loops, is characterized by *stability*. G*rowth* (or *decline*) characterizes the second mode, based on positive feedback loops. The coexistence of the two modes at the heart of an open system, constantly subject to random disturbances from the system's environment, creates a series of common behaviour patterns. The principal patterns can be summarized in a series of simple graphs by taking a variable or any typical parameter of the system (size, output, total sales, and number of elements) as a function of time (Fig. 1.4).

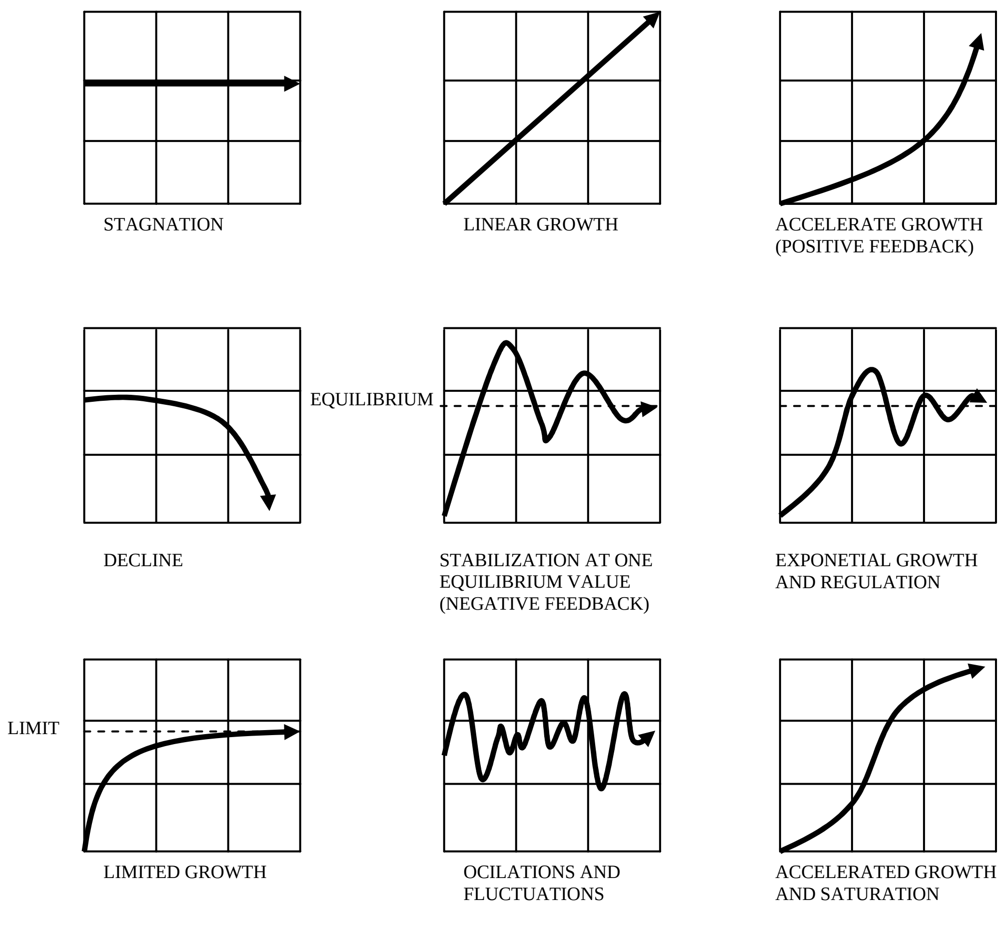

Fig. 1.4. System behaviour patterns (Rosnay, 1979)

1.2.2. Integrated approach and Integrated Assessment

The previous description of the systems approach indicates that the concept of integration only entered in the later stage of the evolvement of systems approach and is limited in integrating disciplines. The new requirements, for example involvement of stakeholders, interaction of different processes at different spatial and temporal scales, and sustainable development, promote a more advanced approach. This approach is referred to as *'integrated approach"*.

The term 'integrated' is often used interchangeably with the term 'holistic'. Schreider and Mostovaia (2001), however, formulate the differences between integrated (in the sense of holistic) approach and Integrated Assessment (in the sense of multidisciplinary). They consider an Integrated Assessment (IA) to be "integrated" in a holistic sense, if it can provide new qualitative knowledge, which cannot be obtained from each component of the IA. However, this separation becomes blurred when one considers a later definition of IA (Van Asselt, 2000):

***Integrated Assessment** is a structured process of dealing with complex issues, using knowledge from various scientific disciplines and/or stakeholders, in such a manner that integrated insights are made available to responsible decision-makers.*

Van Asselt also mentions that: *Integrated assessments should have an added value compared to insights derived from disciplinary research. An integrated approach ensures that key interactions, feedbacks and effects are not inadvertently omitted from the analysis.* It is clear that the integrated (in the sense of holistic) approach has been incorporated in the framework of IA. Therefore, instead of differentiating the integrated approach from IA, it is useful to clarify the meanings of 'integrated' and 'integration'.

As mentioned by Scrase and Sheate (2002), *definitions of assessment and integration unfortunately only add to the lack of precision and clarity surrounding the discourse*. Therefore, their uses in different contexts are investigated to extract the meanings that they have implied. Meijerink (1995) described the integrated approach to water management as a management method which requires an integration of three interrelated systems: natural (water system), socio-economic (water users) and administrative (water management). Janssen and Goldsworthy (1996) formulate 'integration' in the context of multidisciplinary research for natural resource management. Following Lockeretz (1991), they distinguish four forms of integration: additive, non-disciplinary, integrated, and synthetic. The disciplinary integration, which is involved with the respectful interactions among disciplinary scientists, forms the integrated research or interdisciplinary research. Rotmans and De Vries (1997) consider several aspects of integration. In studying closed systems, they describe the first aspect which involves two dimensions of integration: vertical and horizontal. The vertical integration is based on the causal chain. This integration closes the causal loop, linking the pressure (stimulus or input) to a state (state variable), a state to impact (objective variable or output), impact to response (control variable), and a response to a pressure. The horizontal integration addresses the cross-linkages and interactions between pressures, states, impacts and responses for the various subsystems distinguished in the integrated model. The second aspect of integration is that it should bridge what is usually referred to as the domains of natural and social sciences. Parker et al. (2002)

suggest that there are at least five different types of integration within the framework of Integrated Assessment Modelling (IAM). These are integrations of disciplines, of models, of scales, of issues, and of stakeholders. Scrase and Sheate (2002) give a more detailed and critical review on the uses and meanings of integration, integrated approach and integrated assessment. They found fourteen aspects subject to integration in different governance and assessment contexts, such as industry, regulation, planning and politics. These aspects are summarised in Table 1.1.

Table 1.1. Meanings of integration in environmental assessment and governance (After Scrase and Sheate, 2002).

Meaning	Main focus
1) Integrated information resources	Facts/data
2) Integration of environmental concerns into governance	Environmental values
3) Vertically integrated planning and management	Tiers of governance
4) Integration across environmental media	Water, land and air
5) Integrated environmental management (regions)	Ecosystems
6) Integrated environmental management (production)	Engineering systems
7) Integration of business concerns into governance	Capitalist values
8) Triplet of environment – economy – society	Development values
9) Integration across policy domains	Functions of governance
10) Integrated environmental-economic modelling	Computer models
11) Integration of stakeholders into governance	Participation
12) Integration among assessment tools	Methodologies/procedures
13) Integration of equity concerns into governance	Equity/socialist values
14) Integration of assessment into governance	Decision/policy context

The concept of (environmental) governance is defined *as a body of values and norms that guide or regulate state-civil society relationship in the use, control and management of the natural environment.* They also argue that *integration* is a matter of value judgments concerning assessment design in specific historical and social contexts. It implies that integrated approach can be understood as a 'new paradigm' in Thomas Kuhn's (Kuhn, 1970) point of view. In view of the above investigation, ***integration*** *is tentatively interpreted as an act or a process of joining or combining something with something else*. ***The integrated approach*** *is a way of perceiving and solving problems by integrating information, scientific disciplines, tools, interests and other aspects in a systemic way in order to increases the efficiency of our actions*.

1.2.3. Integrated management and policy analysis

Integrated management

Rapid changes of objectives and methodological approaches towards the management of natural resources and environment can be observed in the late twentieth century. The concept of sustainable development introduced in the Brundtland report by the World Commission on Environment and Development (WCED, 1987) accelerated this development process. Sustainable development is defined as: '…*the development that meets the needs of the present without compromising the ability of future generations to meet their own needs*…'. Traditional approaches to natural resource management, which involved single objective (e.g. quantity), sector (e.g. agriculture), discipline (e.g. hydrology) and resource (e.g. water resource), have been being replaced by new approaches which involve multiple objectives, inter-sectors, multidiscipline and multiple resources (Van Ast, 1999; Nakamura, 2003). The research subject is also extended from a single subject like a river or an estuary to a complete water system such as a river basin or a coastal area. These changes result in integrated coastal-zone management, integrated river basin management and/or ecosystem-based river basin management (Nakamura, 2003). Embedded in these approaches are the concepts of participatory management and adaptive management (Miser and Quade, 1988; Clark, 2002; Bennett et al., 2004). These concepts were derived to take into account the multiple perspectives of different agents and to overcome the inherent large uncertainty in model and data. In the World Coast Conference, held in 1993, the following definition of integrated coastal zone management was given (WCC, 1993): '*Integrated coastal zone management involves the comprehensive assessment, setting of objectives, planning and management of coastal systems and resources, taking into account traditional, cultural and historical perspectives and conflicting interests and uses; it is a continuous and evolutionary process for achieving sustainable development*'.

In general, managing natural resources and environment comprises the following four stages (WCC, 1993):

1. problem definition;
2. <u>policy formulation</u>;
3. policy implementation;
4. monitoring & evaluation.

A key step in the <u>policy formulation</u>, which aims at identifying, analyzing and evaluating management strategies, is that of *policy analysis*.

Policy analysis and rapid assessment models

According to Miser and Quade (1985), systems analysis is interchangeably termed as policy analysis in the US and operational analysis in the UK. This is the multidisciplinary problem-solving activity that has evolved to deal with the complex problems that arise in public and private enterprises and organizations. Systems analysis can be described as the invention-and-design (or engineering) art of applying scientific methods and knowledge to complex problems arising in public and private enterprises and organizations and involving their interactions with society and environment (Miser and Quade, 1985). It is not a method or technique, nor is it a fixed set of techniques; rather it is an approach, a way of looking at a problem and bringing scientific knowledge and thought to bear on it. The central purpose of systems analysis is to help public and private decision and policy makers to understand the problem better, so to better manage the policy issues that they face. The successful application of system analysis may help to overcome one or more of the following difficulties: inadequate knowledge and data, many disciplines involved, inadequate existing approaches, unclear goals and shifting objectives, pluralistic responsibilities, resistance to change in social systems, and complexity. System analysis is concerned with theorizing, choosing and acting. Hence, its character is threefold: descriptive (science), prescriptive (advisory) and persuasive (argumentative-interactive). Five steps are suggested in the framework of policy analysis (Miser and Quade, 1985):

1. formulating the problem;
2. identifying, designing, and screening the possible alternatives;
3. forecasting future contexts or states of the world;
4. building and using models for predicting the results; and
5. comparing and ranking the alternatives

Policy analysis, in their view, is primarily concerned with deciding what to do; that is, what is preferred. Policy analysis should not be confused with implementation planning, which is concerned with deciding how to do something. The implementation planning can be referred to as a comprehensive analysis (assessment), while policy analysis corresponds to rapid assessment. Similar frameworks for structured problem-solving strategies are found in (Mintzberg et al., 1976), (Ackoff, 1981) and (Checkland, 1981).

The above framework indicates the importance of using models as tools to assist the policy analysis. These models can be referred to as policy analysis models (Miser and Quade, 1985) or rapid assessment models (De Kok and Wind, 2002). Since they must evaluate many possible policies in terms of many possible impacts, policy analysis models should strive for flexibility, inexpensive operation, and relatively fast response. Moreover, they should allow policies to be described at a relatively gross and conceptual level. Implementation planning models, in contrast, can, and generally do, operate at a considerably more detailed and concrete level, since they will be used to evaluate only a few alternatives.

Combining the conceptual guidelines provided by Miser and Quade (1985) and Randers (1980), six steps can be distinguished in the policy analysis using integrated systems modelling to support management (De Kok and Wind, 1999):

1. the model inception phase

2. the qualitative systems design
3. the quantitative systems design
4. the model implementation
5. the model validation
6. the analysis of policy alternatives

During the inception phase, the problems are defined, alternative solutions to solve these problems are generated and the problem context is described. Qualitative systems design involves the designing of the system structure. During this phase the elements, processes, subsystems which are relevant to problems are selected. The system diagram which links these elements is also established in this phase. Once all the relevant elements and the structure have been identified, the quantitative systems design takes place by collecting the theoretical concepts and data required to describe the systems relationships. This leads to a set of equations and parameters' values. The process of establishing the model parameters' values is called model calibration. The next step, the model implementation, is the formal procedure which results in a computational framework of analysis (a quantitative model). During this phase, modellers are required to verify the quantitative model to ensure that all the elements and relationships are mathematically described correctly. When a quantitative model of the system is available, tests can be carried out to improve confidence in the usefulness of the model. This is the model validation phase. The model calibration, verification and validation will be elaborated throughout the next chapters of the thesis. The policy analysis using integrated systems modelling ends with the activities of comparison and ranking of alternatives, which were mentioned earlier.

1.3. The problem of validating Integrated Systems Models

The systems approach and integrated approach have been promoted for decades. Consequently, there have been an increasing number of studies adopting the systems approach and the integrated approach, especially in the fields of modelling climate change (Dowlatabadi, 1995; Hulme and Raper, 1995; Janssen and de Vries, 1998) and natural resources and environmental management (Stephens and Hess, 1999; Turner, 2000; De Kok and Wind, 2002). These studies are often involved with the design and application of a number of Integrated Systems Models (ISMs). These models are designed to support scenario analysis, but none of them were completely validated in a systematic manner. There are various reasons that can obstruct an effective validation of ISMs. One of them is attributed to the philosophical debate about justification of scientific theories (Kleindorfer et al., 1998). This controversial debate results in a confusing divergence of terminologies and methodologies with respect to the model validation. A few examples related to this philosophical debate are described below.

The spread of positivism as a dominant philosophical school during the second half of the 19th century and first half of the 20th century has had a strong effect on the issue of verification or validation of scientific theory and scientific models. According to positivists, scientific theories are both derived and verified in the light of inductive logic. This means that a theory or hypothesis can be generalized from singular statements (observations or experiments); and the established theory can be verified by conducting observations (experiments) and comparing these with the consequences of a theory.

In opposition to the positivistic school, Popper (1959) argues that scientific theories are established on the base of deductive logic. This means that singular statements are deduced from a universal statement (a theory). The origins of universal statements are not subject to scientific methods. According to Popper, a theory can only be falsified (invalidated) on the base of new empirical evidence, but can not be verified by them. When new evidence favours the consequences of a theory, a theory is said to be corroborated in the light of this evidence. Concerning the validation of scientific theories, Popper also suggested that:

"There are always two competing hypotheses, the two differ in some aspects; and it makes use of the difference to refute (at least) one of them"

Kuhn (1970), in arguing against positivism, put the evolvement of scientific theories into historical context. He argues that scientific theories are derived from a Gestalt, a set of exemplars, or what he calls a paradigm. With regard to the verification of scientific theories Kuhn states:

"One of the future discussions of verification is comparing theories. Noting that no theory can ever be exposed to all possible relevant tests, they ask not whether a theory has been verified but rather about its probability in the light of the existing evidence actually exist. To answer that question, one important school is driven to compare the ability of different theories to explain the evidence at hand"

Furthermore, attention to the issue of model validation in natural sciences was called back in the last decade by some strong scepticists (Konikow and Bredehoeft, 1992; Oreskes et al., 1994). For example, Oreskes et al. (1994) argue that the verification or validation of numerical models of natural systems is impossible. This is because the natural systems are never closed and model results are always non-unique. The openness of natural systems is caused by unknown input parameters and subjective assumptions embedded in observation and measurement of both independent and dependent variables. The problem of non-uniqueness of parameter sets (equifinality) allows for two models to be simultaneously justified by the same available data. A subset of this problem is that two or more errors in auxiliary hypotheses may cancel each other out. They concluded that the primary value of models is heuristic (i.e. models are representations, useful for guiding further study but not susceptible to proof).

In addition to the difficulties related to the validation of natural system models that are set forth above, the validation of ISMs faces several other challenges. The first one is the complexity of an ISM. All ISMs try to address complex situations so that all ISMs developed for exploring such situations are necessarily complex (Parker et al., 2002). The consequences of model complexity on model validation are significant. It can trigger the 'equifinality' problem mentioned before. The dense concentration of interconnections and feedback mechanisms between processes create the need to validate the ISM as a whole, since the validity of each sub-model does not warrant the validity of the whole systems model. Furthermore, the complexity of an ISM amplifies the uncertainty of the final outcome through the chain of causal relationships (see Cocks et al., 1998). Second, the integration of human behaviour into the model creates another challenge. Human behaviour is highly unpredictable and difficult to model quantitatively. It implies that the historical data on processes, which are related human activities, are poor in predictive description of the system future states. This triggers the philosophical problem that successful replication of historical data does not warrant the validity of an ISM. Third, the increase in the scope of the integrated model, both spatially and conceptually, requires an increasing amount of data which are never obtained or rarely measured (see Beck and Chen, 2000). Last, the oversimplification of the complex system (high aggregation level) makes the problem of system openness worse. It is necessary to simplify a real system into a tractable and manageable numerical form. In doing so, the chance of having a more open system is increased.

In summary, the following five factors mostly hamper the validation of an Integrated Systems Model (some may be interrelated):

- Lack of conventional definitions of model validity, model validation and model validity criteria (philosophical problem)
- Complexity of Integrated Systems Models (methodological problem)
- Human involvement (psychological problem)
- Scarcity and absence of field data (data problem)
- High level of aggregation (system openness problem)

Uncertainty does not appear in the list, not because it is unimportant but because uncertainty is embedded in every aspect mentioned above. According to Walker et al. (2003), uncertainty is *any deviation from the unachievable ideal of completely deterministic knowledge of the real system.*

1.4. Research aim and Research questions

The difficulties, which are related to the validity and validation of ISMs, form the central motivation for our research, which aims at *establishing an appropriate validation methodology for ISMs*.

To achieve this objective, the following research questions are addressed:

1. How can validity and validation of Integrated Systems Models (ISMs) be defined?

2. How can validation of an ISM be done?

Since a model is only an abstract simplification of a real system, which is designed for some prescribed purposes, the validity of any model should be judged in view of these purposes. Therefore, the first main research question can be split up into the two following research sub-questions:

1.1. What are the purposes of an Integrated Systems Model?

1.2. What are the appropriate definitions of validity, validation and validity criteria of an Integrated Systems Model with respect to these purposes?

In view of the systems concepts (i.e. elements, structure and behaviour) and the validation difficulties set forth above, the second main research question can be addressed by answering the following sub-questions:

2.1. How can the validity of the elements and the structure of an ISM be established?

2.2. How can the validity of the future behaviour described by an ISM be established?

2.3. How can the validity of the model behaviour be established if the observed data for validation are only available to a limited extent?

The answers to the research questions mentioned above will lead us to a methodology for the validation of Integrated Systems Models for natural resources and environmental management.

1.5. Case study description

1.5.1. RaMCo

In 1994, The Netherlands Foundation for the Advancement of Tropical Research (WOTRO) launched a multidisciplinary research program. The four-year project aimed at developing a scientific framework of analysis for sustainable coastal-zone management. The coastal zone of Southwest Sulawesi, Indonesia, served as the study area. In the project scientists from various scientific disciplines (i.e. marine ecology, hydrology, fisheries, coastal-oceanography, cultural anthropology, human geography, and systems science) cooperated to develop a methodology to support the coastal zone management (De Kok and Wind, 1999). A Rapid Assessment Model for Coastal Zone Management (RaMCo) (Uljee et al., 1996; De Kok and Wind, 2002) was one of the main outcomes of this project.

RaMCo is an Integrated Systems Model, which models the interactions of socio-economic developments, biophysical conditions and policy options. It allows for the analysis and comparison of different management alternatives under various socio-economic and physical conditions for different qualitative and quantitative scenarios and policy options ("what-if" analysis, Fig.1.5). The model encompasses a number of sub-models, namely, marine fisheries, catchment hydrology, land-use and land-cover changes, marine hydrodynamics, and marine ecology. Previously, each sub-model of RaMCo had been calibrated separately, using the available field data from Southwest Sulawesi (Indonesia), expert knowledge and data obtained from the literature. However, the validation of RaMCo as a whole did not take place during the project. The availability of RaMCo provides an excellent case study (see Section 1.1.) to achieve the aim of this thesis.

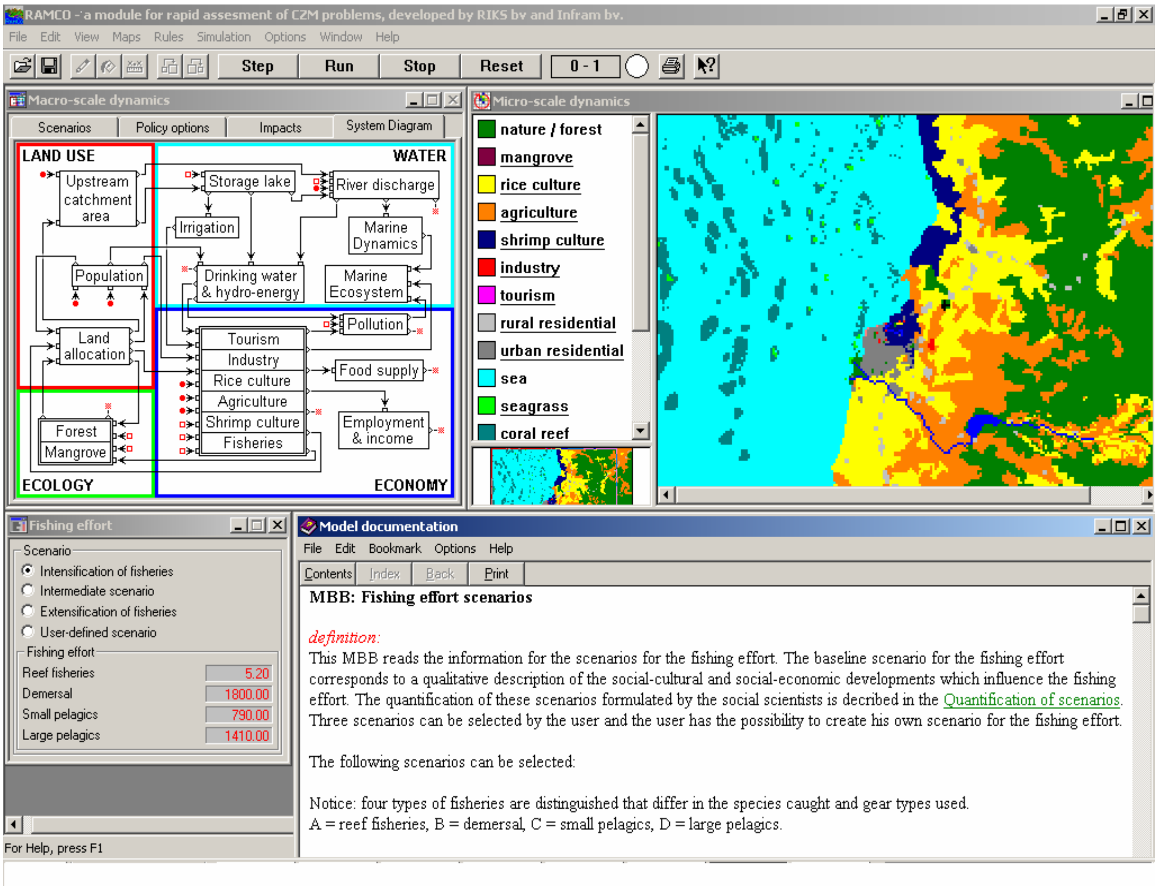

Fig.1.5. The user-interface of RaMCo

1.5.2. Study area

Geography and Administration

The study area for RaMCo occupies a total area of about 8000 km^2 (80km x100km), of which about one half is on the mainland. The off shore part covers the Spermonde Archipelago. The whole study area lies in the South-West part of the South Sulawesi Province, which is one of the four provinces located on the island Sulawesi (Indonesia). It consists of four rural districts (kabupaten): Maros and Gowa in the East, Pangkep in the North, Takalar in the South and the capital of South Sulawesi (Makassar) in the West. The only district which does not border the coast is the Gowa district (Fig. 1.6).

Topology and Geology

Topologically and geologically, the mainland of the study area can be separated into two regions: the lowlands in the Western part and highlands in the Eastern part.

The Western part, from the coast up to some 20 km landwards, is a relatively flat area with the elevation (AMSL) ranging between 0 and 100 m. The slopes in this part are gentle, ranging from 0 % to 8 %. The City of Makassar is located in this flat area. From the coastline going landwards, the geology of this part is determined by quaternary marine and fluvial deposits, and tertiary volcanic sedimentary rocks. The marine deposits are mainly limited to the embouchures and lower courses of the local rivers. They consist of clay, sand and shells. The fluvial deposits were formed by meandering rivers like the Jeneberang river. They occur as natural levees, back swamps, crevasses. Locally, outcrops of limestone of Tertiary age (Miocene) occur (e.g. in the vicinity of Maros).

The highland part is around 40 km in width and ranges in elevation from 100 m to about 3000 m in the very East of the Southwest Sulawesi. It is dominated by the volcano complex of the Lompo Batang Mountain (2876 m (AMSL). The slopes generally vary in a range of 5% to 47%. In this part, a geological survey was carried out for the Jeneberang catchment (Suriamihardja et al., 2001). Two types of rocks are found in this area: volcanic rocks (e.g. andesites, basalts) and sedimentary rocks, mainly of volcanic origin (e.g. tuffs, breccias and conglomerates).

Hydrometeorology

The study area is situated near the equator and has a monsoon tropical climate pattern. There are two distinct seasons: a rainy (wet) season, which contributes around 75% of the total annual rainfall. The wet season begins in November and ends in April; the dry season starts in May and lasts until October. The wettest month is December and the driest month is August or September. The average annual rainfall amount measured in the Jeneberang catchment is around 3000 mm. Spatially distributed, the rainfall increases from North-West to South-East with the increase in elevation. In the West, near the sea, the annual rainfall is around 2000 mm. The annual mean temperature is about 30 ^{o}C. The average monthly humidity is about 85 % in the rainy season and 75 % in dry season (JICA, 1994).

The study area has three main rivers: the Maros river in the North, the Tallo and the Jeneberang rivers in the middle (Fig. 1.6). The Jeneberang river is the most important one with respect to its scope as well as to the roles it plays in the socio-economic and ecological development of the study area. The Jeneberang river flows through the Gowa district and empties into Makassar Strait at the South side of Makassar City, forming a delta. Main tributaries of this river include: the Kunisi River, the Malino River, the Jenerakikang River and the Jenelata River. The minimum and maximum river discharges of the Jeneberang river measured during the period 1983 to 1993 were 2.7 m^3/s and 2,037 m^3/s, respectively (Suriamihardja et al., 2001). The river sediment mainly consisted of washload (75 %) supplied by sheet and rill erosion on the valley slopes (CTI, 1994). The estimate of the average annual sediment yield at the outlet of the river during that period was 1.83 million tonnes. Together with the sediment load, the Jeneberang carries nutrients and freshwater towards the sea, resulting in a higher nutrient level and lower salinity near the shore, compared to the rest of the Shelf. The increase in suspended sediment concentration in Jeneberang river (due to land-use change) and its effect on the lifetime of a reservoir are described in Chapter 4.

Oceanography

The offshore part of the study area covers the Spermonde Archipelago, which is an island group in the Makassar Strait west of Sulawesi. The coastal waters cover about

4000 km^2 with coral reefs, coral islands and sandy shallows, organized in four zones more or less parallel to the coast, and deeper water up to a maximum depth of 60 m. The dominant current direction in the Makassar Strait and over the Shelf is southward. The maximum tidal amplitude is 1.2 m. Sea surface water temperature is 28.5 ^oC and decreases to about 26 ^oC at 20 m depth. The salinity is about 33 ‰, except for the surface layer near the mouth of the Jeneberang river, where it can be as low as 20 ‰ during periods of high river discharge.

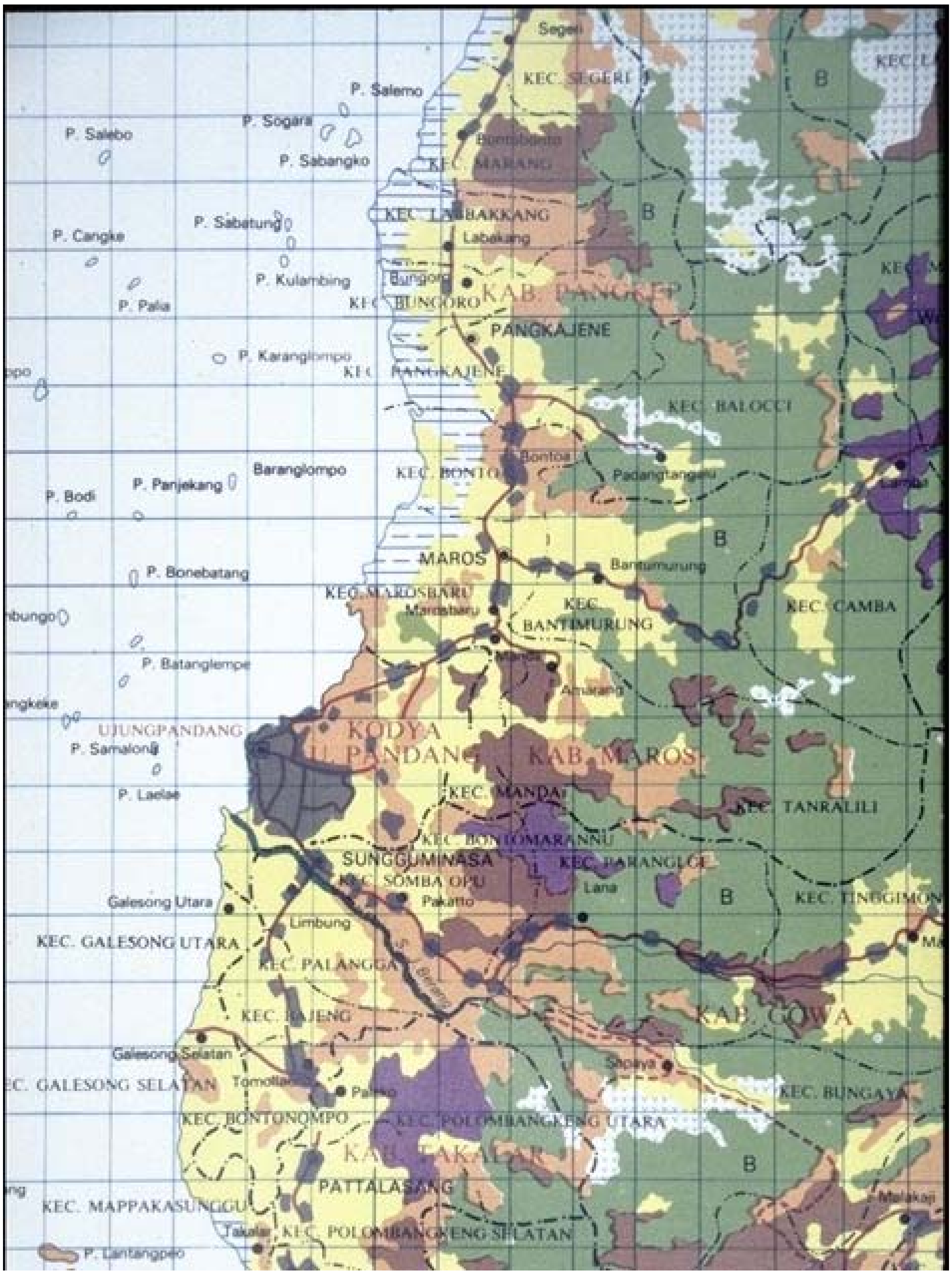

Fig.1.6. Map of study area for RaMCo

Socio-economic characteristics

The total population size of the study area consisted of more than two millions in 1994, half of which were living in Makassar. The high migration rate plus the high natural birth rate make Makassar the most populated city of the study area.

A clear stratification of resources can be observed in the region. Fisheries and reef exploitations are the main sources of income on the islands of the archipelago. Fish and other marine animals are caught around reefs and in the open sea. Along the coast, brackish-water ponds (tambak) are used to cultivate fish, prawns and seaweed. Irrigated rice fields dominate the lower part of the river basins meanwhile rain-fed rice fields are located in the higher area. In the hilly and mountainous area, horticulture such as maize, potatoes and cassava are cultivated. Forest gardens that house the industrial tree such as coffee and cacao can be found. Though agriculture is still of major importance both for income and employment, the significance of non-agricultural activities such as construction and industry is growing. Major projects, which are ongoing or planned to develop the urban area, include the Makassar harbour, the nearby Hasanuddin airport and regional tourism. Large-scale industrial development will be concentrated in the 700 ha KIMA industrial site, situated in the north of Makassar. The regional GDP development in the period of 1991 to 2001 is depicted in Fig.1.7. It is noted that, due to the Asian economic crisis in 1997, the inflation of the Indonesian Rupiah is remarkably high.

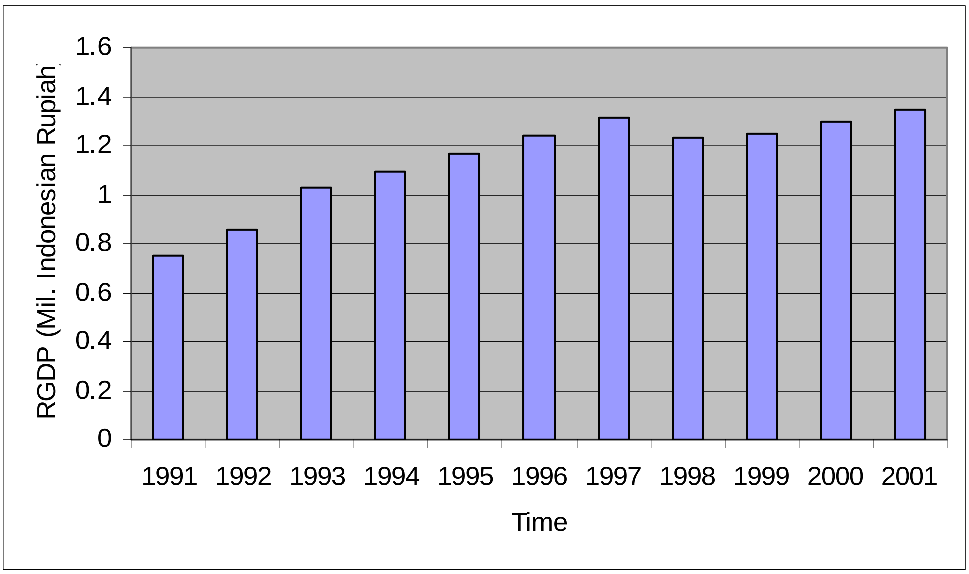

Fig.1.7. Regional GDP development of South Sulawesi (sources: Indikator Ekonomi, Prov Sulawesi Selatan; unit: Mil. Indonesian Rupiah)

1.6. Outline of the thesis

In Chapter 2, the validity and validation of ISMs are defined. A conceptual framework and the detailed steps designed for validation of ISMs are presented. This framework and the procedure reflect the philosophical position taken in this thesis, which lies somewhere between objectivism (in the sense that there is an ultimate truth) and relativism (one model is as good as another), between rationalism and empiricism. Based on this position, we treat an ISM as a tool which is designed for specified purposes. The model validation is considered to be a process, which should take these purposes into account. The first main research question is addressed in this chapter.

In Chapter 3, a validation procedure, which can identify the strength and weakness of the model components and its structure by using the available data from literature and local expert opinions, is described. The approach is based on the Morris sensitivity analysis, a simple expert elicitation technique and the Monte Carlo uncertainty analysis to facilitate three validation tests, namely Parameter-Verification, Behaviour-Anomaly, and Policy Sensitivity tests. Two management variables: the living coral reef area and the total Biological Oxygen Demand (BOD) discharged to the coastal seawater are selected for the purpose of demonstration. This procedure aims at establishing the validity of the model structure and its relevant components, keeping in mind the model's purpose as a tool for discussion between experts and the end-users. This chapter addresses the research question 2.1 of the thesis.

Chapter 4 is devoted to the description of a new approach towards validation of ISMs using future qualitative scenarios. Within this approach, expert knowledge is elicited in the form of future qualitative scenarios and translated into quantitative projections using fuzzy set theory. Trend line comparison of the behaviour projections made by the model and by experts can reveal the structural faults of the model. This new approach is derived to establish the validity of an ISM with respect to its purpose as a communication tool between system experts (i.e. scientific experts and resource managers). This chapter addresses the research question 2.2.

In contrast to Chapters 3 and 4, where the procedure and the new approach are aimed to test the systems model as a whole, Chapter 5 is devoted to the development of a procedure to separately test a process-based model embedded in ISMs. It is based on the fact that for a small model it is easier to collect empirical (i.e. observed) data needed for the quantitative validation. Within this method, residual analysis is proposed to examine the pattern replication ability of the model. The Mitchell (1997) test is used to test the predictive accuracy, and the extreme behaviour test is adopted to test the plausibility of the model. This addresses the research question 2.3 of the thesis.

The final chapter gives an overall discussion and conclusion on the methodology for validation of Integrated Systems Models such as RaMCo. The limitations as well as innovative points of the established methodology are discussed. Recommendations for the future research on the validation of ISMs finalize the thesis.

Chapter 2

Methodology

2.1. Introduction

Finding proper definitions for the validity and validation of a model is still an issue that creates a lot of arguments among scientists and practitioners. Although the literature on model validation is abundant, this issue is still controversial (Kleijnen, 1995; Rykiel, 1996; Oreskes, 1998). The term validity has sometimes been interpreted as the absolute truth (see Rykiel, 1996 for a detailed discussion). However, increasing evidences accumulated from scientific research and the literature show that this is a wrong interpretation of the validity of an open system model (Oreskes et al., 1994; Sterman, 2002; Refsgaard and Henriksen, 2004). It is widely accepted that models are tools designed for specified purposes, rather than truth generators. Therefore, the validity of an ISM can be considered to be equivalent to the user's confidence in the model's usefulness (Forrester and Senge, 1980). Validation is defined by them as the process of establishing confidence in the soundness and usefulness of a model.

As a result of the diversity of definitions of validity and validity criteria, methodologies developed for the model validation are also scattered. Oftentimes, point-by-point comparisons between simulated and real data are considered to be the only legitimate tests for model validation (Reckhow et al., 1990). These tests are usually used to evaluate the model behaviour to conclude on the model's validity. However, these tests are argued to be unable to demonstrate the logical validity of the model's scientific contents (Oreskes et al., 1994), to have poor diagnostic power (Kirchner et al., 1996) and even to be inappropriate for the validation of system dynamics models (Forrester and Senge, 1980). A review of methodologies for the validation of process models and decision support systems is given by Finlay and Wilson (1997). However, those methodologies give insufficient guidelines for solving particular problems related to the validation of Integrated Systems Models such as the scarcity of field data, the qualitative nature of the social sciences and the uncertain (future) context of the system studied (e.g. uncertain parameters, inputs and boundaries).

The objective of this chapter is to provide a brief review on model validation and to define validity, validation and validity criteria for Integrated Systems Models. Based on these definitions, a methodological framework and a detailed procedure are developed to validate Integrated Systems Models such as RaMCo.

Part of this chapter has been published as:
Nguyen, N.T. and De Kok, J.L., 2003. Application of sensitivity and uncertainty analyses for validation of an integrated systems model for coastal zone management. Proceedings of the International Congress on Modelling and Simulation, MODSIM, 2003 (ed. Post, D.A.). Townville, Australia. 542-547.

2.2. Literature review

This section presents a review of the representative frameworks, approaches and techniques for model validation which can be found in scientific literature dating back to the 1980s. The models to be validated, which are included in this review, consist of simulation models in operational research (Shannon, 1981; Sargent, 1984, 1991; Balci, 1995; Kleijnen, 1995; Fraedrich and Goldberg, 2000), models in earth sciences (Flavelle, 1992; Ewen and Parkin, 1996; Beck and Chen, 2000), agricultural models (Mitchell, 1997; Scholten and ten Cate, 1999), ecological models (Van Tongeren, 1995; Kirchner et al., 1996; Rykiel, 1996; Loehle, 1997), system dynamics models (Forrester and Senge, 1980; Barlas, 1994; 1999) and integrated models (Finlay and Wilson, 1997; Beck, 2002; Parker et al., 2002; Poch et al., 2004; Refsgaard et al., 2005). The controversial debate on terminologies for model validation (Oreskes et al., 1994; Oreskes, 1998; Rykiel, 1996; Beck and Chen, 2000) points to the ambiguity and overlap between the terms: model testing, model selection, model validation or invalidation, model corroboration, model credibility assessment, model evaluation and model quality insurance. To counter the ambiguity of the terminology, a clear definition of validity and validation of ISMs is proposed in Section 2.3.

The most common framework for model validation, which is widely accepted in the modelling community, can be attributed to Sargent's work (1984; 1991). Sargent considered model validation as substantiation that a computerised model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model. In this framework, the validity of a simulation model consists of three dimensions: conceptual validity, operational validity and data validity. To determine the conceptual validity of a model, two supplementary approaches are often used. The first approach is to use mathematical and statistical analyses (e.g. correlation coefficient, Chi-square test) to test the theories and assumptions (e.g. linearity, independence) underlying the model. The second approach is to have an expert or experts evaluate the conceptual model in terms of both the model logic and its details. This approach is often referred to as *peer review*, and is aimed at determining whether the appropriate details, aggregation level, logic, mathematical and causal relationships have been used for the model's intended purpose. Two common techniques used for the second approach are face validation and traces (Sargent, 1984; 1991). It is worth noting that the input-output behaviour of the model is not considered in conceptual validation although both expert opinion and observed data can be used. Operational validity, in Sargent's term, is primarily concerned with determining that the model's output behaviour has the accuracy required for the model's intended purpose over the domain of its intended application. Three conventional approaches for operational validation based on the comparison of model output and observed data are graphical comparison, hypothesis testing and confidence intervals (Sargent, 1984). In addition, two other comparison approaches, using goodness-of-fit statistics (e.g. root mean square) and residual analysis between model output and observed data, are mentioned by Flavelle (1992). These common approaches based on the comparison between model output and observed data are often referred to as *history-matching* (Beck, 2002), and will be discussed in more detail in Chapter 5. More techniques developed for operational validation, which range from qualitative, subjective, informal tests (e.g. face validity of model behaviour) to quantitative, objective and formal tests (e.g. statistical tests), are described in (Sargent 1984; Balci, 1995; Kleijnen, 1995;

Rykiel, 1996; Mitchell, 1997; Scholten and ten Cate, 1999; Fraedrich and Goldberg, 2000). It is important to emphasise that the relevance of the available validation approaches and techniques depends on the availability of field data and the level of understanding of the system studied (or scientific maturity of the underlying disciplines), as recognised by Kleijnen (1995), Rykiel (1996) and Refsgaard et al. (2005). Furthermore, the requirement of validity of a model under a set of experimental conditions under which the model is intended to be used is emphasised and studied by several authors (e.g. Ewen and Parkin, 1996; Kirchner et al., 1996). Ewen and Parkin (1996) proposed a 'blind' testing approach to the validation of the catchment model to predict the impact of changes in land-use and climate, given the limitations of existing approaches, such as the simple split-sample testing, differential split-sample testing, proxy-catchment testing and differential proxy-catchment testing. This 'blind' testing approach, however, does not consider the interactive natural-human systems which is more complex and qualitative in nature.

Another conceptual framework for the validation of system dynamics models has been suggested by Forrester and Senge (1980). Within this framework, validation is defined as the process of establishing confidence in the soundness and usefulness of the model. According to these authors, model validity is equivalent to the user's confidence in the usefulness of a model. The confidence of the model users is gradually built up after each successful validation test. Validation tests are divided into three major groups: tests of model structure, tests of model behaviour and tests of policy implication. Particular validation tests have been proposed, corresponding to each group. The important characteristics of this conceptual framework are: the focus of validation on the structure of the model system, the vital roles of the experts' knowledge/experience and qualitative, informal tests (e.g. extreme condition test and pattern test) in the validation process. These characteristics are reflected by the extensive use of terms such as soundness, plausibility and confidence. Barlas (1994, 1999) separates validation tests into two main groups: direct structure testing and indirect structure (or structure-oriented behaviour) testing. Perceiving that pattern prediction (period, frequencies, trends, phase lags, amplitude) rather than point prediction is the task of system dynamics models, he has developed formal statistics and methods which can be used to compare the simulated behaviour patterns with either observed time series or anticipated behaviour patterns. In line with this philosophical perspective on model validation, Shannon (1981) proposed a similar conceptual framework for the validation of simulation models in operational research. The differences in Shannon's framework are the integration of verification and validation, and an extensive inclusion of the formal, quantitative, statistical approaches to model validation. A closely related framework for the validation of ecosystem models is proposed by Loehle (1997), in which a new version of the hypothesis testing approach is considered to be essential for the validation of ecological models.

As the complexity of integrated models used in decision making increases, the usefulness of quantitative validation approaches based on the comparison between model output and observed data decreases. This is due to the scarcity and uncertainty of field data for the model calibration and validation. The model validation using peer review is also challenged by the conflict of interests of the peers and the limited number of capable peers, due to the multidisciplinary nature of the integrated models (Beck, 2002; Parker et al., 2002). These foster a shift of model validation perspective from scientific theory testing to evaluating the appropriateness of the model as a tool designed for a specified task. In accordance with this view, the two supplementary approaches, which have just begun to develop, are: i) judging the trustworthiness of the model according to the quality of its design in performing a given task, and ii) using the information (experience) obtained from the interactions and dialogues between the modellers and a variety of system experts (resource managers, scientific experts) and stakeholders. An example of the former approach is given by Beck and Chen (2000), in which the model quality is judged, based on the properties of internal attributes - the number of key and redundant parameters. Although the need for the latter approach to model validation is recognised (Beck and Chen, 2000; Parker et al., 2002; Poch et al., 2004; Refsgaard et al., 2005) appropriate tools and methods have not been developed yet.

In summary, although the literature on model validation is abundant most of the available techniques, methods, and approaches focus on quantitative tests for operational validation (or historical matching), given that the observed data are available. The conceptual validity or structural validity, which is equally important for integrated models, has been a neglected issue. There is a lack of consideration of the uncertain future conditions, under which the model is intended to be used in model validation frameworks. In addition, there is little attention to the qualitative nature of social science, which is often required to be incorporated in integrated systems models to support the decision making process.

2.3. Concept definition

Purposes of Integrated Systems Models

Since a model is only an abstract and simplification of a real system, which is designed for some prescribed purposes, the validity of any model should be judged with respect to these purposes. The literature and our own experiences provide the following main functions of an ISM (De Kok and Wind, 2002; Parker et al., 2002):

1. Database and library function: an ISM provides quick access to the storage of field data (in the form of tables, graphs and maps), theoretical concepts (in the form of equations, structural diagrams) and scientific references.

2. Educational function: an ISM can be used to develop the skill of inquiring, understanding and looking at a problem from an integrated systems perspective, a perspective that perceives a real and complex world with many types of interactions, for example, between social, economic and biophysical subsystems.

3. Research prioritising function: by working with an ISM on a particular problem, one can determine which areas of research are important to the problem at hand but lack measurements and/or theoretical background. Research efforts and budget can then be prioritised accordingly.

4. Scenario building function: an IMS can act as a tool for scenario building and for discovering our ignorance.

5. Communication and discussion function: an ISM can be used as a platform which facilitates discussions among system experts and between system experts and stakeholders. These discussions are aimed to arrive at a common view of the problems and common ways to solve them.

6. Decision support function: an ISM is used as a tool to describe the impact of measures and scenarios on the achievement of policy objectives (i.e. policy analysis).

Validation of an ISM is always important, but essential with respect to the last four purposes.

Validity, validation and validity criteria

In view of the purposes of ISMs and the concepts of systems approach, the validity of an integrated systems model pertains to four aspects: the *soundness and completeness of the model structure, plausibility and correctness of the model behaviour.* S*oundness* of the structure is understood to be based on valid reasoning thus be free from logical flaws. C*ompleteness* of the structure means that the model should include all elements relevant to defined problems and their causal relationships which concern the stakeholders. *Plausibility* of the model behaviour means that behaviour should not contradict general scientific laws and established knowledge. Behaviour *correctness* is understood as the extent to which computed behaviour and measured behaviour are in agreement. This extent should be within the allowable permit (validity criterion), which again depends on the purpose of a model and the requirements of the model users. These four aspects lead us to the following definition of the validity of an ISM:
'*The validity of an Integrated Systems Model is the soundness and completeness of the model structure together with the plausibility and correctness of the model behaviour.'*

Before refining the definition of the validation of Integrated Systems Models, a few remarks are given to clarify this definition:

- An Integrated Systems Model like RaMCo should not be understood as a quantitatively predictive model, which is mentioned by Oreskes (1998). Therefore, the term "validation" can be used.

- Validation can take place after the model-building phase, but it is not the end of the model life cycle. In other words, a model is always in need of adjustment when new data and new knowledge are available, and validation facilitates that adjustment process. The main purpose of model validation is not seeking the yes or no answer but establishing the validity of a model.

- Calibration is the process of specifying the values of model parameters with which model behavior and real system behavior are in good agreement.

- Verification is the process of substantiating that the computer program and its implementation are correct, i.e., debugging the computer program (Sargent, 1991).

In view of the model purposes and in line with our definition of model validity, we define validation of an Integrated Systems Model as: "*the process of establishing the* *soundness and completeness of model structure together with the plausibility and correctness of the model behaviour*".

The process of establishing the validity of the model structure and model behaviour addresses all three questions concerned with validation as stated by Shannon (1975; 1981). In other words, validation is carried out to address the three following questions, which are the modified ones from Shannon (1981) and Parker et al. (2002):

i) Are the structure of the model, its underlying assumptions and parameters contradictory to their counterparts observed in reality and/or to those obtained from expert knowledge?

ii) Is the behaviour of the model system in agreement with the observed and/or hypothesized behaviour of the real system?

iii) Does the model fulfil its designated tasks or serve its intended purpose?

Consequently, one main purpose of validation is to show transparently both the strong and weak points of the model to its potential users. The potential users could be the decision-makers (i.e. resource managers), analysts (i.e. people acting as intermediates between scientists and decision-makers), or the model builders themselves (Uljee et al., 1996). Another component of model validation is to find suggestions for improving the model structure and its elements so that the validity criteria are met. This leads us to requirements for the definitions of performance criterion and validity criterion.
A performance criterion defines what aspect of the model we want to examine and what references are used for this examination. For example, a certain performance criterion was drafted as "the ability of the model to match historical field data". The aspect of the model examined here is "the ability of the model to (re)produce a plausible input-output relationship" and reference for this examination is obtained from "observed data". A performance criterion determines what test(s) should be performed for the validation.

A validity criterion defines how good a model is, given the performance criterion. This criterion can be either qualitative or quantitative, which depend on the purpose of the model. For instance, Mitchell (1997) proposed as a validity criterion for a predictive model as "ninety five per cent of the total residual points should lie within the acceptable bound".

2.4. Conceptual framework of analysis

It is necessary to distinguish three systems (Fig. 2.1) that will frequently be mentioned later on. The *real system* includes existing components, interactions, causal linkages between these components and the resulting behaviour of the system in reality. However, in most cases we do not have enough knowledge about the real system. The *model system* is the abstract system built by the modellers to simulate the real system, which can help managers in decision-making processes. The *hypothesized system* is the counterpart of the real system, which is constructed from the hypotheses for the purpose of model validation. The hypothesized system is created by and from the available knowledge of experts and/or the experiences of the stakeholders with the real system through the process of observation and reasoning. With the above classification, we can carry out two categories of tests, namely, empirical tests and rational tests with and without field data (Fig. 2.1). Rational tests can also be used to validate a model when the data for validation are available only to a limited extent.

We define empirical tests as those tests that are based on the direct comparison between the model outcomes and the field data. Empirical tests are conducted to examine the ability of a model to match the historical data (hindcasting), the future data (forecasting), and other qualitative behaviours (e.g. frequency, mode) of the real system. In case no data are available, the hypothesized system and the model system are used to conduct a series of rational tests, such as: parameter-verification, structure-verification, and extreme policy tests (Forrester and Senge, 1980). These tests are referred to as rational tests, since they can be carried out, based on the availability of expert knowledge and through reasoning processes. Rational tests are increasingly important for the situation where the real data of the complex system are lacking and subject to considerable uncertainty.

There should be a clear distinction between two terms: objective variable and stimulus. Objective variables are either output variables or state variables that decision-makers desire to change. They can also be referred to as Management Objective Variables (MOVs). Examples of objective variables in RaMCo are the living coral reef area (an output variable, in Chapter 4) and sediment yield at the outlet of a basin (a state variable, in Chapter 5). Stimuli (drivers) are input variables which, in combination with control variables and state variables (Chapter 1), drive the objective variables.

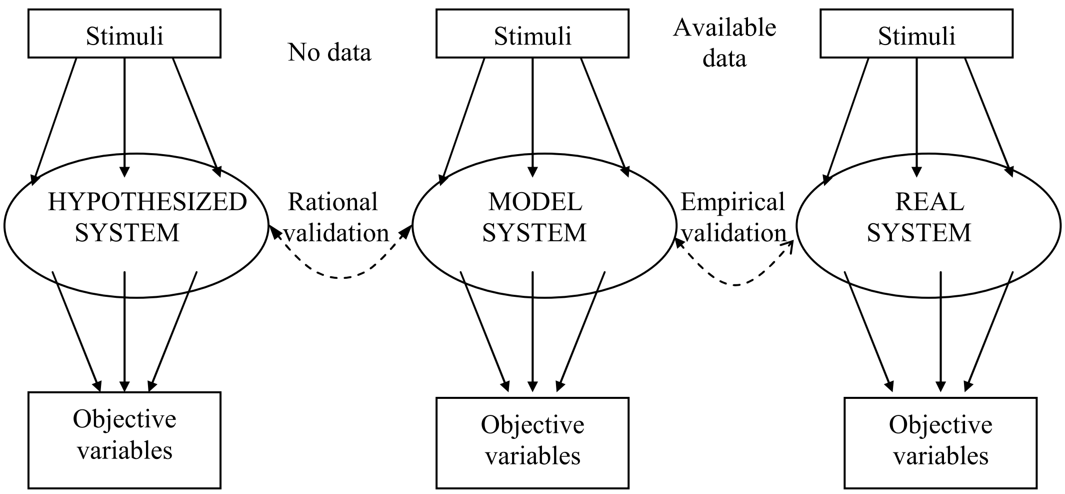

Figure 2.1. Conceptual framework of analysis for validation of RaMCo.

In figure 2.1, we have the three systems as mentioned. With the same stimuli as the inputs of each system, we have different values of objective variables as the systems' outputs. The differences are caused by the lack of knowledge of the real system and/or other problems (e.g. errors in field data measurements, computational errors). The model builders always want the model behaviour to be as close to the behaviour of the other two systems as possible. If validation data are not available, one has to assume (for practical reasons) that the hypothesized system made up by experts is a better presentation of the real system, as compared with the model system created by modellers. To obtain a higher degree of confidence, one can calibrate or validate expert knowledge as in the case of data validation (Sargent, 1991). Examples of expert knowledge calibration techniques are group meetings and the Delphi technique (Shannon, 1975), Analytical Hierarchy Process (Zio, 1996), and Adaptive Conjoint Analysis (Van der Fels-Klerx et al., 2000).

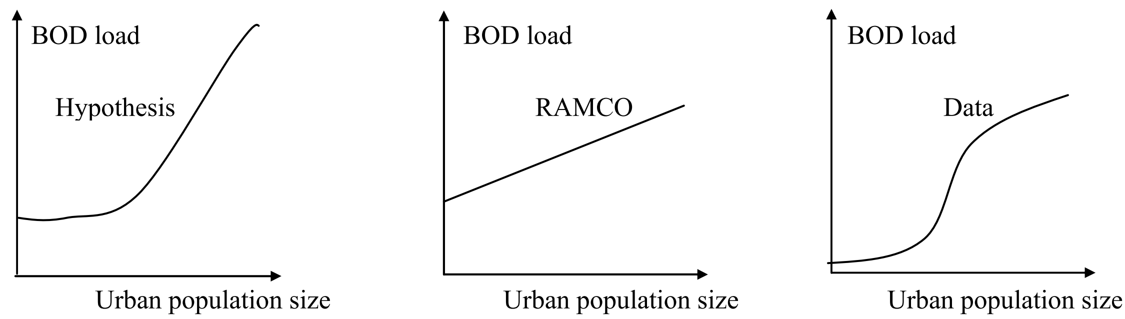

Figure 2.2. A hypothetical example demonstrating the validation framework. From left to right: Biological Oxygen Demand (BOD) load generated by the hypothesized system, the model system, and the real system as a result of increase in population size.

In figure 2.2, there are three graphical representations of Biological Oxygen Demand (BOD) load as the objective variable of a hypothesized system, a model system and a real system. The stimulus presented on the horizontal axis is the increase in the urban population of Makassar (formerly known as Ujung Pandang). The solid straight line and curves present the trend and amplitude of the BOD load as a result of changes in the urban population size. For this example, hypothesized behaviour is omitted because we have enough data to conduct the validation tests. A symptom generation test (Forrester and Senge, 1980) is applied in order to conclude that the model is able to generate the symptom of difficulty (the increase in BOD load corresponding to the increase in urban population with the same magnitude) that motivated the construction of the model. However, the simulated behaviour pattern (linear) is different from the observed pattern (accelerated growth). After that, tests proposed by Mitchell (1997), Scholten and Van der Tol (1994) and Scholten et al. (1998) can be conducted to give the quantitative measure of agreement between model outcome and observation.

2.5. Procedure for validation

One of the reasons that make validation of an Integrated Systems Model difficult is its complexity. In order to overcome this problem, we propose a procedure for the validation of ISMs, which consists of sixteen systematic steps (Fig. 2.3).

In Fig. 2.3, Phase 1 was designed to serve three purposes. For the first purpose, during the model design stage, several components and subsystems with unknown contributions to the system outputs were included. Therefore, it is necessary to know, after the modelling has been completed, which components are relevant to the system outputs, based on the model (sensitivity). The information obtained from this phase can be compared with expert opinions (hypotheses) and/or field data obtained in Phase 2 (step 5) to assess the soundness and completeness of the model's structure. The second purpose is to reduce the workload of collecting field data for the validation. Only the data on a number of model inputs, parameters, and state variables, which are specified as important in Phase 1 and from expert opinions (in step 5, Phase 2), need to be collected. The third objective is to reduce the workload for testing the model. Since all the tests focus on the relevant subsystems and clusters that are specified as important by the model system and experts/stakeholders, work can be saved.

Starting with the results obtained from Phase 1, collecting field data and expert knowledge about the system studied is conducted in Phase 2. More attention is paid to inputs and parameters that have strong influence on interested outputs and those involved with large uncertainty. Evaluation of field data can be carried out at this point. This evaluation tells us whether the field data are of good quality and the data set is large enough for empirical tests (which rarely happens). Otherwise, one has to rely on rational tests, based purely on hypotheses or the combination of expert knowledge, literature and the available field data.

Having data obtained in Phase 2 and bearing in mind what performance criteria are to be used, suitable tests for sub-models and clusters chosen in Phase 1 and Phase 2 are specified in Phase 3. The validity criterion for each test will be taken from literature or decision-makers, since it should be based both on the nature of the test and the purpose of the model. To deal with the uncertainty of inputs, parameters in a model, propagation of uncertainty will be carried out in a specific test. The computer model should be adapted to facilitate this analysis. The last step of Phase 3 is to represent the results of tests in easily understandable forms for those who are not familiar with mathematics and have no deep knowledge of the model.

Phase 4 is involved with finalizing the results obtained from preceding phases. The first two steps (steps 13 and 14) would require expert-group meetings to draw conclusions on model quality, model usefulness and to suggest the solutions to improve the weak points of a model. This phase ends with a detailed report, describing the whole process of validation and recommendations.

In the discussion, the term "purpose of the model" has been repeated to emphasize that the purpose of the model decides the framework of validation as well as the details of most steps. RaMCo was designed to link measures, scenarios, and Management Objective Variables (MOV) in order to support the decision-making process. This means that point-prediction is generally not the target of RaMCo since it is not a predictive model in the strict sense.

2.6. Conclusion

The common purposes of Integrated Systems Models have been reviewed. Based on these purposes, the literature on model validation and the concepts of systems modelling approach, the validity of an ISM is claimed to comprise four aspects: the soundness and completeness of the model structure together with the plausibility and correctness of the model behaviour. The correctness of the model behaviour is elaborated in Chapter 5, but briefly mentioned here as the agreement between the trends and magnitudes of the behaviours produced by the model system and the real system (i.e. field data).

It is concluded that a point-by-point goodness of fit between model behaviour and real data is neither a sufficient nor an appropriate condition for the validity of an ISM. The conceptual validity or structural validity, which is equally important for integrated models, has been a neglected issue. There is a lack of consideration of the uncertain future conditions, under which the model is intended to be used in model validation frameworks. In addition, there is little attention to the qualitative nature of social science, which is often required to be incorporated in integrated systems models to support the decision making process.

To achieve a model which meets the four criteria for validity a methodological framework for validation has been established. The realisation of this framework into systematic steps is also outlined. Within this methodological framework, expert knowledge and local stakeholders' experiences play an important role in the process of establishing the validity of an Integrated Systems Model. The use of expert and stakeholders' opinions will be demonstrated in Chapters 3 and 4 of the thesis.

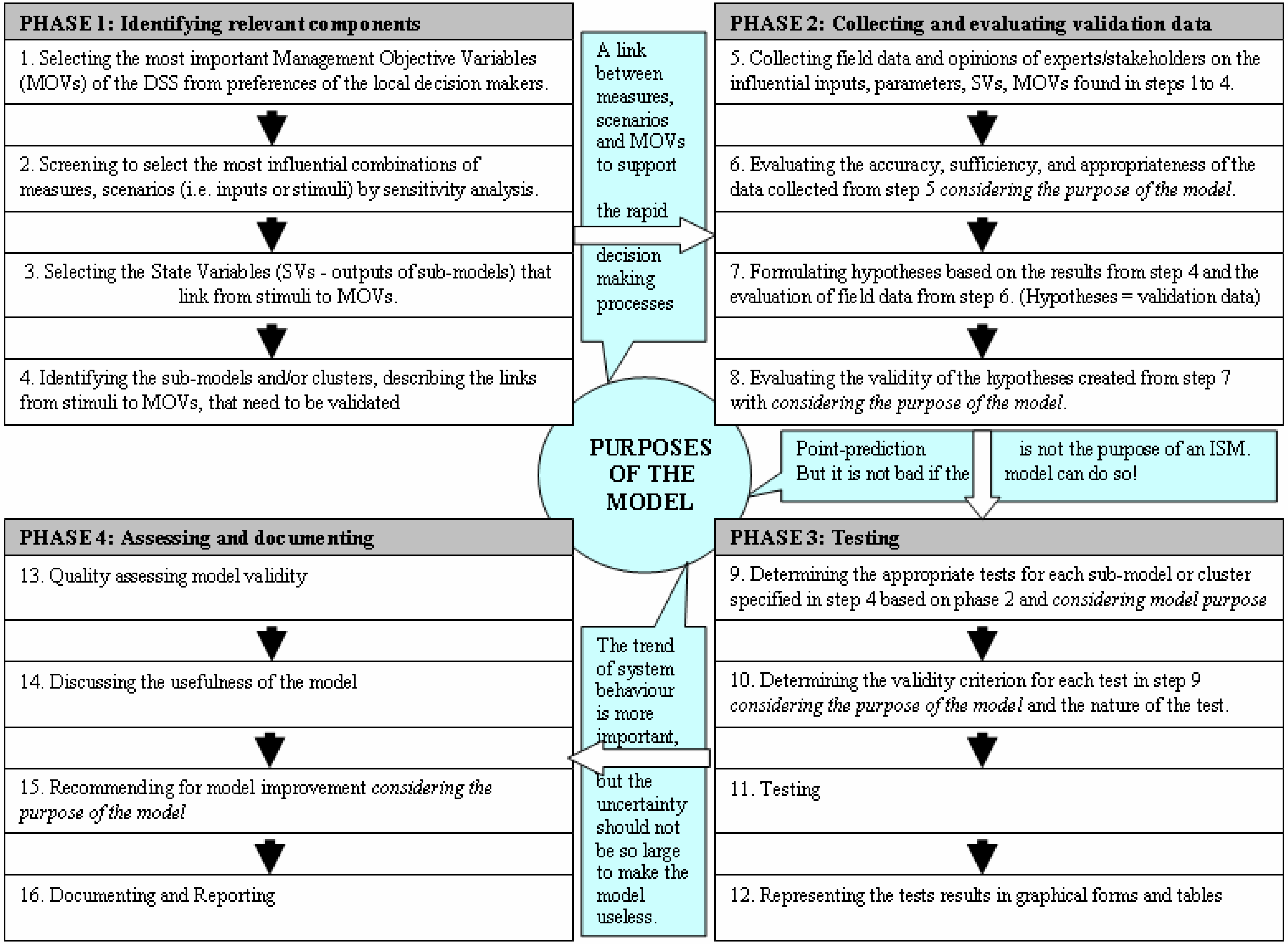

Figure 2.3. Flowchart for validation of ISMs

Chapter 3

Validation of an integrated systems model for coastal zone management using sensitivity and uncertainty analyses

Abstract: RaMCo (Rapid Assessment Model for Coastal Zone Management) is a decision support system, which encompasses a number of process models related to marine fisheries, hydrology, land-use/land-cover changes, coastal hydrodynamics, marine ecology, and the linkages between them. The complexity of the model and the scarcity of field data make empirical validation of the integrated system difficult. This calls for validation procedures which can identify the strength and weakness of the model with the available data from literature and experts' opinions. In this chapter, such a procedure is described. The approach uses the Morris sensitivity analysis, a simple expert elicitation technique and Monte Carlo uncertainty analysis to facilitate three validation tests, namely, Parameter-Verification, Behaviour-Anomaly, and Policy Sensitivity tests. The usefulness of the procedure is demonstrated for two case examples, namely pollution of waste water discharge and the coral living area.

3.1. Introduction

There are various reasons that make the validation of an integrated system model (ISM) more difficult than the validation of a conventional process model. The most important problems are: the inherent complexity of an ISM, scarcity of field data, the lack of knowledge about internal and external factors as well as the linkages between component processes of the real system (Jansen and de Vries, 1999; Beck and Chen, 2000). Although suggestions made for validating such models are available from the literature (Forrester and Senge, 1980; Finlay and Wilson, 1997; Saltelli and Scott, 1997; Parker et al. 2002; Jakeman and Letcher, 2003), appropriate validation procedures for ISMs have not been fully developed yet.

Sensitivity and Uncertainty Analyses (SUA) are considered to be essential for model validation (Saltelli and Scott, 1997; Scholten and Cate, 1999; Refgaard and Henriksen, 2004). Depending on the questions the validation need to answer, different types and techniques of SUA have been applied (Kleijnen, 1995; Tarantola et al., 2000; Beck and Chen, 2000).

In this chapter, a validation procedure using sensitivity and uncertainty analyses is presented and applied to validate RaMCo. The Morris method (Morris, 1991) is used to

The main part of this chapter has been submitted as:
Nguyen, T.G and DE Kok, J.L., 2005. Validation of an integrated systems model for coastal zone management using sensitivity and uncertainty analysis. Journal of Environmental Modelling and Software. (In revision).

determine the parameters, inputs and measures that have important effects on the outputs of the model. The opinions of the end-users (local scientists and local stakeholders) on the key influential factors affecting the corresponding outputs are elicited. Monte Carlo uncertainty analysis is applied to propagate the uncertainty of the model inputs and parameters to the uncertainty of the output variables. The results obtained are used to conduct three validation tests suggested by Forrester and Senge (1980). They are Parameter-Verification, Behaviour-Anomaly and Policy-Sensitivity tests. These tests are conducted to reveal the weaknesses of the parameters and structure employed by RaMCo. The pollution of waste water discharged into the coastal sea, as expressed in the total Biological Oxygen Demand (BOD), and the living coral area serve as case examples.

3.2. Methodology

3.2.1. Basics for the method

The purpose(s) of a model should guide the process of its validation. There has been an increasing consensus among researchers and modellers that a model's purpose is the key factor determining the selection of the validation tests and the corresponding validity criteria (Forrester and Senge, 1980; Rykiel, 1996; Parker et al., 2002). RaMCo is intended to be used as a platform which facilitates discussions between scientific experts and scientific experts, and between scientific experts and stakeholders. These discussions are aimed to arrive at a common view on the problems and the ways to solve them. Therefore, the terms "scientific experts", "stakeholders", "common view" and "common solutions" require more elaboration in the context of validation.

Stakeholders play an important role in the validation process of an ISM (Jakeman and Letcher, 2003). Since the main purpose of an ISM is to define a "common view" and find "common solutions" for a set of problems perceived by scientific experts and stakeholders, the role of stakeholders should be considered during the validation of an ISM. The stakeholders could include both decision-makers and people affected by the decisions made. Acting as a policy model, an ISM can be considered useful when it is able to simulate the problems and their underlying causes that the stakeholders experience in the real system. Furthermore, an ISM should be able to distinguish the consequences of various policy options so that the decisions can be made with a certain level of confidence.

The validation of an ISM is a process of testing the model to unravel the errors in and the incompleteness of the model so that suggestions for improvements can be made. The validity of a model cannot be assessed on the basis of a single test, but a series of successful tests should be carried out to increase the user's confidence in the usefulness of a model. Forrester and Senge (1980) designed seventeen tests for the validation of system dynamics models, some of which are closely related. These tests are categorized in three main groups: tests of model structure, tests of model behaviour and tests of policy implications.

3.2.2. The testing procedure

The approach presented in this chapter uses SUA as tools to facilitate three of the validation tests proposed by Forrester and Senge (1980). These tests include: Parameter-Verification, Behaviour-Anomaly and a Policy-Sensitivity tests, which are described in detail in subsection 3.2.6. These three tests are selected because of the absence of field data for the model validation, the extent to which the computer program can be modified and the availability of local experts/stakeholders' opinions. The testing procedure can be described as the following.

First, the Morris method (Morris, 1991) is applied to determine the parameters, inputs and measures (together these are called factors) which have important effects on the objective variables. The first round of the Morris analysis adopts the set of model factors, the ranges of which were set by the modellers. The important processes containing the factors, which have dominant effects on the objective variables, are interpreted. Next, the elicitation of the expert opinions of the local stakeholders and the local (scientific) experts is conducted. This results in factors and processes that the local experts consider to have dominant effects on the objective variables concerned. The factors, embedded in the processes, pointed out by the first round of analyses and by the local experts, are subject to the Parameter-Verification test. This test involves the examining of the important factors to determine if they correspond conceptually and numerically to knowledge of the real system. After finishing the Parameter-Verification test, the second round of the Morris analysis is carried out with the set of refined factors. Here, Anomaly-Behaviour enters into the testing process. Within this test, the orders of importance of the factors determined by the model and by the local experts are compared. If there are differences between the two, the causes of the differences are analyzed. Convincing evidence (from measured data and literature) must be collected to explain the causes of the differences. If no evidence can be found, the model requires improvement, simply because the stakeholders do not trust the usefulness of the model. Once a general agreement about the model structure is reached, Policy-Sensitivity tests are conducted. In this test, the refined uncertainty ranges of selected factors are propagated to the objective variables using Monte Carlo uncertainty analysis (Morgan and Henrion, 1990) under two conditions that the recommended policies are in effect and not in effect. The confidence of end-users is gained if the model enables a distinction between the consequences of various policy options, given the uncertainty in model inputs and parameters. The following subsections describe the Morris method, the method for elicitation of local expert opinions, the Monte Carlo uncertainty analysis, and the details of the three tests used for the above testing procedure.

3.2.3. The sensitivity analysis

Different types (local versus global) and a variety of techniques (e.g. regression analysis versus differential analysis) are available for Sensitivity Analysis (SA). Some investigations of these techniques were described by Iman and Helton (1988) and Campolongo and Saltelli (1997). The selection of the SA method to solve a particular problem is often based on the model complexity and the nature of the questions the analysis needs to answer. Following the guidelines given by Morgan and Henrion (1990), the present study adopts the Morris method (Morris, 1991) for the analysis.
Morris (1991) made two significant contributions to sensitivity analysis.

First, he proposed the concept of elementary effect, $d_i(X)$, attributable to each input x_i. An elementary effect can be understood as the change in an output y induced by a relative change in an input x_i (e.g. an increment of 10 kg BOD/day of the total BOD load to the coastal sea is induced by a 33 % decrease in the total water treatment plant capacity). The elementary effect is given by:

$$d_i(X) = \frac{y(x_1, x_2, ..., x_i + \Delta, ..., x_k) - y(X)}{\Delta} \quad (3.1)$$

In Eq. (3.1), X is a vector containing k inputs or factors $(x_1,...x_i,...x_k)$. A factor x_i can randomly take a value in an equal interval set $\{x_i^1, x_i^2, ..., x_i^p\}$, each with equal probability. In this set of real numbers, x_i^1 and x_i^p are the minimum and maximum values of the uncertainty range of factor x_i, respectively. The symbol p denotes the number of levels chosen for each factor. For the sake of technical convenience, each element of vector X is assigned a rational number (Morris, 1991) or a natural integer number (Campolongo and Satelli, 1997) in the Morris design. Therefore, after the design, transformation of these factors to real numbers is necessary for model computations. The symbol Δ denotes a predetermined increment of a factor x_i, whose value is chosen in such a way that $x_i + \Delta$ remains within the uncertainty range of x_i. The frequency distribution F_i, constructed by randomly selecting r elementary effects of each factor x_i, gives an indication of the degree and nature of the influence of that factor on the specified output. For instance, a combination of a relatively small mean μ_i with a small standard deviation σ_i indicates a negligible effect of the input x_i on the output. A large mean μ_i and a large standard deviation σ_i indicate a strong nonlinear effect or strong interaction with other inputs. A large mean μ_i and small standard deviation σ_i indicate a strong linear and additive effect.

Second, Morris designed a highly economical numerical experiment to extract k samples of elementary effect; each with a size r (k is the number of analyzed factors and r is the number of elementary effects constructing one F_i). The total number of model runs is in the order of k (rather than k^2). Readers interested in the Morris design are referred to Morris (1991) and Campolongo and Saltelli (1997) for the details.

The purpose of the Morris method (Morris, 1991) is to determine the model factors that have important effects on a specific output variable by measuring their uncertainty contributions. The order of importance of these factors results from the following four sources of uncertainty: i) Model structure uncertainty (the way modellers conceptualize the real system, e.g. the aggregation level); ii) Inherent variability of factors observed in the real system, e.g. the price of shrimp; iii) The deterministic changes of decision variables, e.g. capacities of water treatment plants, and iv) Uncertainty introduced by the analysts (the analysts' knowledge about model parameters and inputs, e.g. estimates of factors' ranges). The "true" order of importance, according to the model, of a factor should be determined only from the first three sources of uncertainty and variation. The last source of uncertainty should be minimized, in order to correctly determine the order of importance for each factor with the Morris analysis. This is the reason for using the preliminary results of the Morris analysis and expert opinions to carry out the Parameter-Verification test and for using the results from the second round of the Morris analysis to conduct the Behaviour-Anomaly test.

3.2.4. The elicitation of expert opinions

Elicitation of expert opinions has been proposed for both uses as a heuristic tool (i.e. it is used in the exploratory context) and as a scientific tool (i.e. it is used in a justification context) (Cooke, 1991). The procedures guiding expert elicitation vary from case to case, depending on the purpose of the elicitation (Ayyub, 2001). This section describes the procedure followed to obtain opinions from local scientific experts and local stakeholders about the factors that have important effects on the organic pollution of the coastal sea, and on the area of living coral. The ultimate purpose is to find the differences between the model behaviour and the system behaviour anticipated by local experts. With the obtained results, a series of validation tests can be conducted, focusing on the causes of the differences. This subsection describes the main steps in the elicitation process: selecting experts and stakeholders, eliciting and combining expert opinions.

Selection of respondents for the elicitation: The definitions and criteria to select experts for elicitation may vary, depending on the nature of the answers elicitors want to obtain. For example, Cornelissen et al. (2003) define an expert as a person whose knowledge in a specific domain (e.g. the welfare of laying hens) is obtained gradually through a period of learning and experience. They distinguish stakeholders from experts by differentiating the roles the two groups play in the different phases of the systems evaluation framework. These phases include the following activities: defining public concern, determining multiple issues, defining measurable indicators, and interpreting information on measured indicators to derive conclusions. In view of the purpose of the elicitation stated previously, we define experts as knowledgeable people who participate in the processes of operation and management of the real system directly (decision-makers and experienced staff), and indirectly (local scientists).

Elicitation: The elicitation was conducted using a questionnaire technique. The elicitation started with an expert training section, including a presentation of RaMCo during workshops, an explanation of the purpose of the questionnaires and a clarification the terminology used in the questionnaires. The questionnaires were delivered to the participants during workshops and collected during the week after. This gave the experts sufficient time to think about the questions and answers thoroughly. In the questionnaire, participants were asked to add the missing factors/processes to the given set of factors/processes that could have important effects on the model objective variables. They were asked directly to rank the order of importance of these factors (see appendix A for an example). In comparison with other elicitation methods, such as Adaptive Conjoint Analysis (Van der Fels-Klerx et al, 2000) and the Analytical Hierarchy Process technique (Zio, 1996), this simple method assumes that the experts are unbiased and internally consistent (i.e. calibration is considered unnecessary). In view of the purpose of the questionnaire, the availability of experts, and their willingness, this method is considered appropriate for the current case study.

Aggregation: To aggregate expert opinions, the mathematical approach (in contrast to the behavioural approach) is adopted (Zio and Apostolakis, 1997). For the stakeholder group, the simple average method is used. For the group of local scientists, in addition to the simple average method, an attempt was made to associate a weight to each expert's answer, depending on 1) Knowledgeable Fields (*KF*), 2) Professional Title

(*PT*), 3) Years of Experience (*YE*), 4) Source of Knowledge (*SK*), and 5) Level of Interest (*LI*). These factors are selected from a larger set of aspects proposed to have direct contributions to the overall ranking of experts' judgments by Cornelissen et al. (2003) and Zio (1996). This weight association aims to examine whether the result obtained from simple average method is substantially altered when weights of experts are included. Equations 3.2 and 3.3 are used to calculate the final ranking for each factor/process:

$$x = \frac{1}{S}\sum_{i=1}^{n} w_i x_i \qquad (3.2)$$

where $$S = \sum_{i=1}^{n} w_i$$

$$w_i = \frac{1}{8} KF_i \left(PT_i + YE_i + SK_i + LI_i\right) \qquad (3.3)$$

In Eq. 3.2, w_i is the weight assigned to the expert *i*, which represents the degree of confidence that the analyst associates with the answers of the expert *i*, to a certain set of questions related to an objective variable; x_i is the ranking of a factor/process with regard to its relatively influential importance, given by the expert *i*; *x* is the aggregated ranking of that from all experts. In Eq. 3.2, KF_i reflects the fields of expertise, expert *i* has knowledge about, which have values of zero or one; PT_i, YE_i, SK_i, LI_i represent professional title, years of experience, source of knowledge and level of interest of expert *i* on a certain set of questions, respectively. Their values are in the range between zero and two. The result of Eq. 3.3 is the weight assigned to the expert *i*, which has minimum values of zero when the expert *i* does not have knowledge about a certain objective variable and a value equal to one when the expert *i* has the highest quality on every aspect previously defined (Appendix B).

3.2.5. The uncertainty propagation

As mentioned in the previous sections, the Monte Carlo uncertainty analysis is used to propagate the uncertainty in model parameters and inputs to the uncertainty of the output variables. The quantities subjected to the uncertainty propagation in policy models may include decision variables, empirical parameters, defined constants, value parameters, and others (Morgan and Henrion, 1990). Morgan and Henrion argue that it is generally inappropriate to represent the uncertainty of decision variables and value parameters by probability distributions. However, it is useful to conduct a parametric sensitivity analysis on these quantities to examine the effect on the output of deterministic changes to the uncertain quantity. This parametric sensitivity analysis can be conducted by means of the Morris analysis, as in this chapter. Examples of the decision variables in RaMCo are the number of fish blasts, the total capacity of urban wastewater treatment plants, and the total capacity of treatment plants for industrial wastewater. Examples of empirical parameters in RaMCo are the price of shrimps and the BOD concentrations in urban wastewater. For the technical details of the Monte Carlo uncertainty propagation readers are referred to (Morgan and Henrion, 1990).

3.2.6. The validation tests

This subsection describes three of seventeen tests for the validation of system dynamics models proposed by Forrester and Senge (1980).

Parameter-Verification test

Parameter verification means comparing model parameters to knowledge of the real system to determine if parameters correspond conceptually and numerically to real life. The failure of a model to mimic the behaviour of a real system could result from the wrong estimations of the values and the uncertainty ranges of the model parameters (numerical correspondence). Besides, the parameters should match elements of system structure (conceptual correspondence). For a simple model, it is often easy to fit the model output with the measured data by varying the parameters' values. This is called calibration. However, for ISMs, the difficulty in obtaining data on parameters, inputs and outputs makes this kind of calibration almost impossible. Moreover, due to the requirement of a sound structure of an ISM, the plausibility of the parameters and inputs of the model should be taken as one of the criteria to conclude on the model structure plausibility and the model usefulness. For that reason, Forrester and Senge (1980) suggest it as a validation test. In this study, the Parameter-Verification test is extended to include both model parameters and inputs.

Behaviour- Anomaly test

The behaviour anomaly test aims to determine whether or not the model behaviour sharply conflicts with the behaviour of the real system. Once the behavioural anomaly is traced back to the elements of the model structure responsible for the behaviour, one often finds obvious flaws in the model assumptions.

Policy-Sensitivity test

The policy sensitivity test aims to determine if the policy recommendations are affected by the uncertainties in parameter values or not. If the same policies would be recommended, regardless of parameter values within a plausible range, the risk of using the model will be less than if two plausible sets of parameters lead to opposite policy recommendations. In this chapter, we put this test in a similar context while retaining its meaning and purpose. The usefulness of a policy model is increased if it enables a distinction between the consequences of various policy options, given the uncertainty in model inputs and parameters.

3.3 Results

3.3.1. Sensitivity analysis

In a sensitivity analysis of an ISM, the selection of the quantities of interest depends on the aim of the analysis. The purpose of the current analysis is to determine the order of importance of the factors/processes provided by the model and to compare them with the experts' experiences. Therefore, the total BOD load to the coastal sea and the living coral area after five years of simulation (the year 2000) are selected to be the quantities of interest.

In the first round of the Morris analysis, all model factors are grouped and the representative factors for each group are traced back and selected qualitatively on the basis of the quantities of interest. This results in a reduction in the number of factors to be analyzed, from 309 to 137 (k=137). Next, the quantitative ranges of the selected parameters and inputs are obtained from the default set of the factors' ranges defined by the modellers. Since RaMCo does not only include inputs and parameters but also 'measures' and 'scenarios', an adaptation is needed to allow for the Morris method. To compare the importance of the measures with other parameters and inputs, all the measures are assumed to be implemented simultaneously. The magnitude of a decision variable (affected by a measure) is treated in the same manner as it is done with inputs and parameters. One example is the decision to build a water treatment plant to reduce the concentration of BOD in the wastewater discharged to the sea. Next, the Morris design is applied with the number of levels for each factor equal to four (p=4), the increment of x_i to compute elementary effects $d_i(x)$, $\Delta = 1$ (Campolongo and Saltelli, 1997). The selected size of each sample $r = 9$. A total number of model evaluations N=1142 (N= r (k + 1)) is performed. Finally, the two indicators representing the importance of each factor uncertainty, the mean μ and the standard deviation σ, are computed and plotted against each other (Fig. 3.1).

It can be seen from Fig. 3.1 that there are only three important processes that have a significant contribution to the total BOD load. In descending order of importance, these are: brackish-pond culture (factors 68, 86, 87,124, 13, 14 and 15), urban domestic wastewater (factors 120, 113 and 55) and industrial wastewater (factor 5).

In Fig. 3.2, the results obtained from the second round of the Morris analysis show some interesting points. Contrary to the results of the Morris analyses applied to natural system models (Campolongo and Saltelli, 1997; Comenges and Campolongo, 2000), the rankings provided by μ and σ are different (see Table 3.1). This can be attributed to the highly complex combination of both linear and nonlinear relationships between the output and the input variables. However, the rankings measured by μ and by the Euclidean distance from the origin in the (μ, σ) plane are well in agreement (Table 3.1). This shows that the mean μ is a good indicator to measure the overall influence of a factor on a certain output as suggested by Morris (1991).

Fig. 3.2 does not show distinct clusters of factors belonging to each process that have significant effects on the objective variable, as shown in Fig. 3.1. This is because there are no dominant processes that have much higher effects than others. From Fig. 3.2 and Table 3.1, it can be noticed that the most important process influencing the total BOD load is the domestic wastewater discharge. For comparison between the effect of the urban wastewater discharge and wastewater discharge of shrimp-pond culture, the sum of the mean μ from all factors belonging to each process is computed. The shrimp-pond culture contributes a value of 12.2 to the variability of the total BOD, while industrial wastewater contributes a value of 11.0. The small difference between the two values does not allow a strict judgment concerning the difference in the order of importance of the two processes.

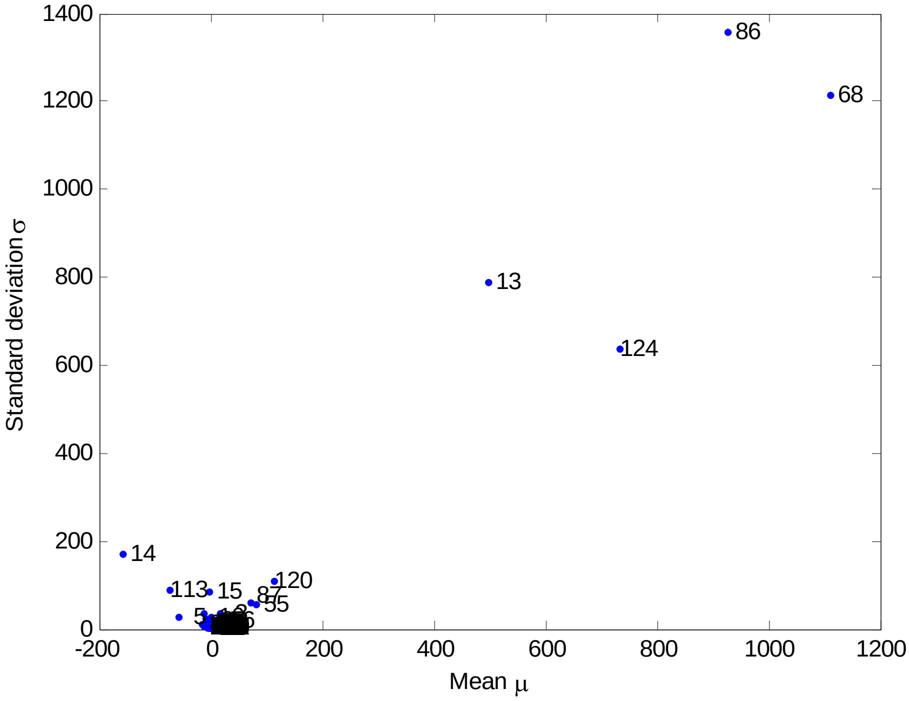

Fig. 3.1. Means and standard deviations of the distributions of elementary effects of 137 factors on the total BOD load resulting from the first round of analysis.

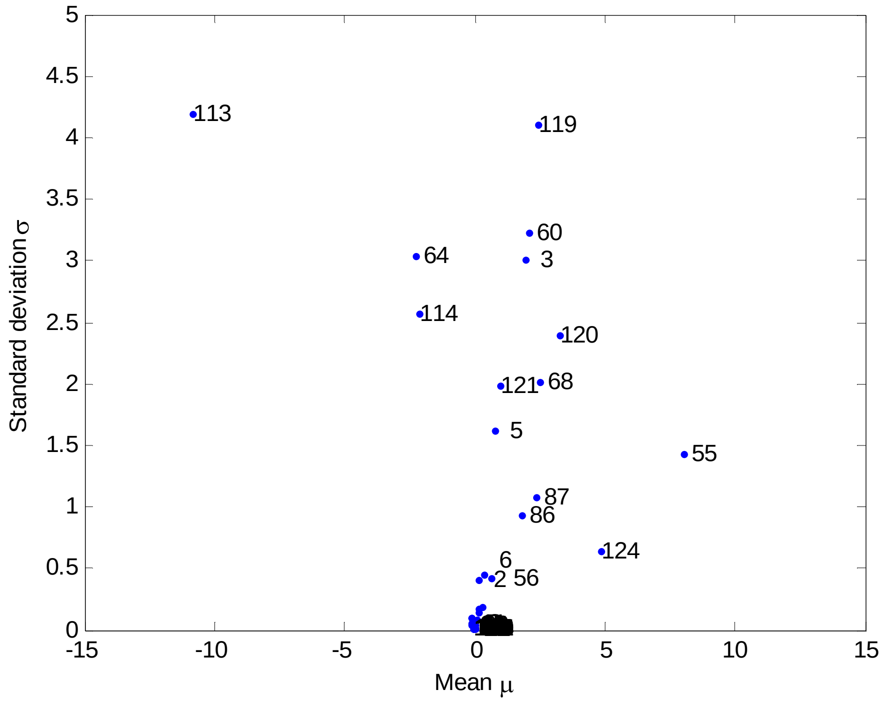

Fig. 3.2. Means and standard deviations of the distributions of elementary effects of 137 factors on the total BOD load resulting from the second round of analysis.

Table 3.1. Results of Morris analysis on the relative important effects of 137 factors on the total BOD load. The 20 factors are listed in descending order of the importance, resulting from the second round of analysis.

Factor	$\lvert\mu\rvert$	σ	$\sqrt{\mu^2+\sigma^2}$	Short description
113	10.81	4.19	11.59	Total purification capacity of domestic wastewater treatment plants (10^6 $m^3 day^{-1}$)
55	8.05	1.42	8.18	Percentage of urban connected households (%)
124	4.85	0.64	4.89	BOD generated by 1 kg of shrimp (kg BOD kg^{-1} shrimp)
120	3.26	2.39	4.04	BOD concentration of domestic wastewater before purification (mg l^{-1})
68	2.56	2.01	3.25	Spatial growth rate of shrimp pond area $(10^6 IDR)^{-1}$
119	2.47	4.10	4.78	Production of wastewater per industrial production value [$10^6 m^3$ $(10^6 IDR)^{-1}$]
87	2.40	1.07	2.63	Yield of the extensive shrimp culture (ton ha^{-1})
64	2.26	3.04	3.78	Time for investment of industry to take effect (month)
114	2.14	2.57	3.34	Total purification capacity of industrial water treatment plants (10^6 m^3 day^{-1})
60	2.08	3.23	3.84	Slope coefficient of the relationship between investment and production of industry (-)
3	1.97	3.00	3.59	Urban Income (10^6 IDR cp^{-1} $year^{-1}$)
86	1.82	0.93	2.05	Yield of the intensive shrimp culture (ton ha^{-1})
121	1.03	1.99	2.24	BOD concentration of industrial wastewater before purification (mg l^{-1})
5	0.82	1.62	1.81	Yearly investment on the industry (10^6 IDR $year^{-1}$)
56	0.63	0.42	0.76	Water demand for unconnected households (m^3 cp^{-1} day^{-1})
6	0.38	0.44	0.58	Yearly investment on shrimp intensification (10^6 IDR $year^{-1}$)
122	0.30	0.19	0.35	BOD concentration of domestic wastewater after purification (mg l^{-1})
123	0.19	0.17	0.25	BOD concentration of industrial wastewater after purification (mg l^{-1})
13	0.17	0.13	0.22	Relative growth rate of shrimp price (-)
2	0.15	0.40	0.43	Immigration scenario selection (-)

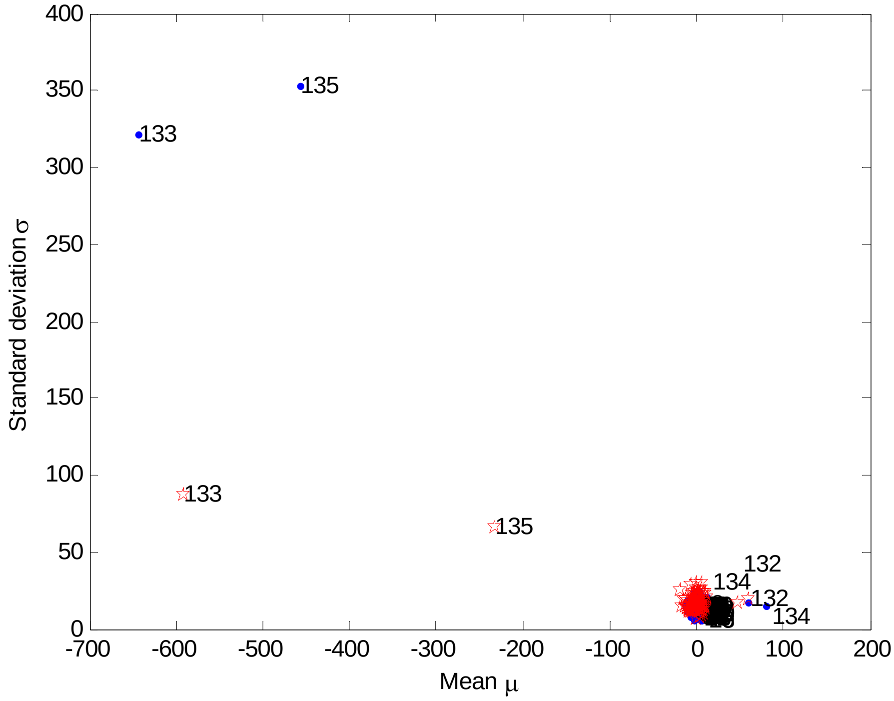

Fig. 3.3. Means and standard deviations of the distributions of elementary effects of 137 factors on the living coral area at the first (dot) and the second (star) rounds of analysis.

Fig. 3.3 shows the four important factors that have an effect on the total area of living coral from the first (dot) and second (star) rounds of the Morris analysis. Factors 133 (damaged surface area of coral reef per fish blast) and 135 (the number of fish blasts per year per ha) demonstrate that the most important process, influencing the living coral area, is blast fishing. Factor 132 (natural growth rate of coral reef) and factor 134 (recovery rate of damaged coral) play a relatively small role compared to blast fishing. The other factors, such as the effect of suspended sediment, are so small that they are outstripped by the effect of a stochastic module to generate the spatial distribution of fish blasts over the coastal sea area.

3.3.2. Elicitation of expert opinions

As mentioned in Section 3.2, the respondents subject to elicitation include stakeholders and scientific experts. The stakeholder elicitation was based on the availability of an advanced course of environmental studies in South Sulawesi, focusing on an integrated approach, held at the University of Makassar (UNHAS). The group of participants consisted of 27 staff members, working in various provincial and district departments (Bappeda in Indonesian). They are the people who work on the relevant issues of the real system daily. Their educational backgrounds were different, but the majority had Engineering and Master degrees in Agriculture, Aquaculture, Water Resources, Meteorology, Infrastructure and Marine Biology. The scientist elicitation was based on the scientific experts coming from the various faculties of UNHAS and a few people from Provincial Departments (two persons) and a Ministry (two persons) with higher educational backgrounds. With the two groups mentioned, it is also possible to study the differences in understanding and perception of the environmental problems between the two groups.

Tables 3.2 and 3.3 show the results of expert opinion aggregation of the two groups. Depending on the objective variables, the number of respondents answering a specific set of questions varies. There are 18 and 15 respondents answering the issue of coral reef degradation and marine pollution, respectively for the first group. The corresponding numbers for the second groups are 7 and 8, respectively.

Table 3.2. Results of the analysis of the important factors/processes affecting the organic pollution, elicited from local stakeholders and scientific experts (SEs)

Factor	Stakeholders			SEs (simple average)			SEs (weighted average)	
	Ave.	Std.	Rank	Ave.	Std.	Rank	W. Ave.	Rank
Domestic	1.50	0.94	1	1.50	0.55	1	1.45	1
Industry	1.73	1.22	2	1.50	0.89	2	1.60	2
Shrimp	2.00	1.03	3	2.38	0.71	3	2.50	3

Table 3.3. Results of the analysis of the important factors/processes affecting the living coral area, elicited from local stakeholders and scientific experts (SEs)

Factor	Stakeholders			SEs (simple average)			SEs (weighted average)	
	Ave.	Std.	Rank	Ave.	Std.	Rank	W. Ave.	Rank
Sus. Sed.	2.74	0.73	5	2.29	0.95	3	2.29	3
Blast	2.00	1.41	1	1.29	0.49	1	1.35	1
Cyanide	2.17	1.47	2	2.00	1.15	2	1.97	2
Nat. gro.	2.22	1.26	3	2.57	0.98	4	2.73	4
Recover	2.61	1.42	4	3.00	1.15	6	3.13	6
Mining	2.95	1.35	6	2.71	0.95	5	2.85	5

In Table 3.3, a low average (Ave.) value indicates a high ranking of a factor and a low standard deviation (Std.) value indicates a high consensus among the respondents on the ranking of a factor. This means that there is a high consensus among the scientific experts on the most important effect of blast fishing on the living coral reef area. The result obtained from the stakeholder group also indicates the most important effect of blast fishing, but with a higher variation (Std. = 1.41). Cyanide fishing is indicated to be the second most important factor by both groups. The order of importance of the remaining four factors differs slightly between the two groups. However, there is a general agreement between the two groups about the relatively low effect of coral reef mining for construction on living coral area.

For the organic sources that contribute to the organic pollution of coastal waters, the same average values of domestic and industrial sources provided by scientific experts indicate an equal order of importance of the two (Table 3.2). However, for domestic wastewater, a higher consensus was obtained. When using the weighted average method for the combination of expert opinions, the results show the distinction between the two. The ranking, in descending order, is domestic wastewater, industrial wastewater, and shrimp-pond culture wastewater. This ranking is the same as the ranking indicated by the stakeholders.

The results in Tables 3.2 and 3.3 show that, in most cases, the standard deviations given by the scientific experts are smaller than those given by the stakeholders. This indicates a higher degree of consensus among the scientific experts than among the stakeholders. Besides, the differences in the average ranking values of two successive factors/processes given by the scientific experts are larger than the corresponding values given by the stakeholders (except for domestic wastewater and industrial wastewater in Table 3.2). This could indicate that the scientific experts have more confidence to differentiate the order of importance of the factors/processes than the stakeholders.

The association of the weights to individual expert's answers results in the same ranks of factors/processes influencing the two objective variables, compared with the ones obtained by the simple average method (Tables 3.2 and 3.3). This is an indication that the simple average method is appropriate for this study. It is noted that the weights (w_i) computed by Eq 3.3 are based on a subjective assumption of equal weights of the four aspects (PT, YE, SK, LI). Different sets of these weights can be assigned to study the sensitivity of these aspects to the final results. However, this is beyond the scope of this chapter.

3.3.3. Uncertainty analysis

In Fig. 3.4, the uncertainty propagations of the input factors to the living coral reef area are conducted for two cases. The first case is the extended current situation (no measure), where the ban on blast fishing is not in effect due to a number of socio-economic and political reasons. The second case is the situation where the ban on blast fishing is assumed to be enforced (with measure). This situation is inspired by a study on blast fishing in Komodo National Park (Pet-Soede et al., 1999) where about 90 % of fish blasts were reduced after a patrolling programme had been implemented. The uncertainty bounds are subject to a 95% confidence level, with a sample size of 1000 simulation runs. A similar approach is followed for the total BOD discharge into the coastal seawater and the results are shown in Fig. 3.5. This depicts the extended current situation and the situation where urban wastewater treatment plants are installed, both under the assumption that 90% of the urban households are connected to the water supply network.

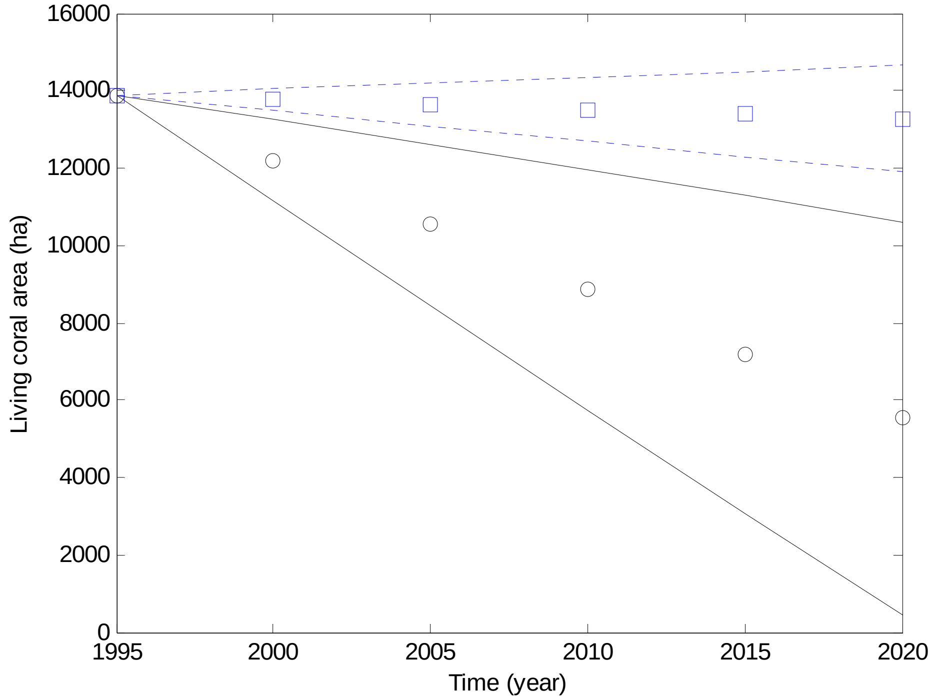

Fig. 3.4. Results of the uncertainty analyses for the living coral area for two cases: a) full enforcement of a ban on blast fishing (dotted lines: 95 % confidence bounds, and □: mean) and b) without this measure (solid lines: 95 % confidence bounds, and o: mean).

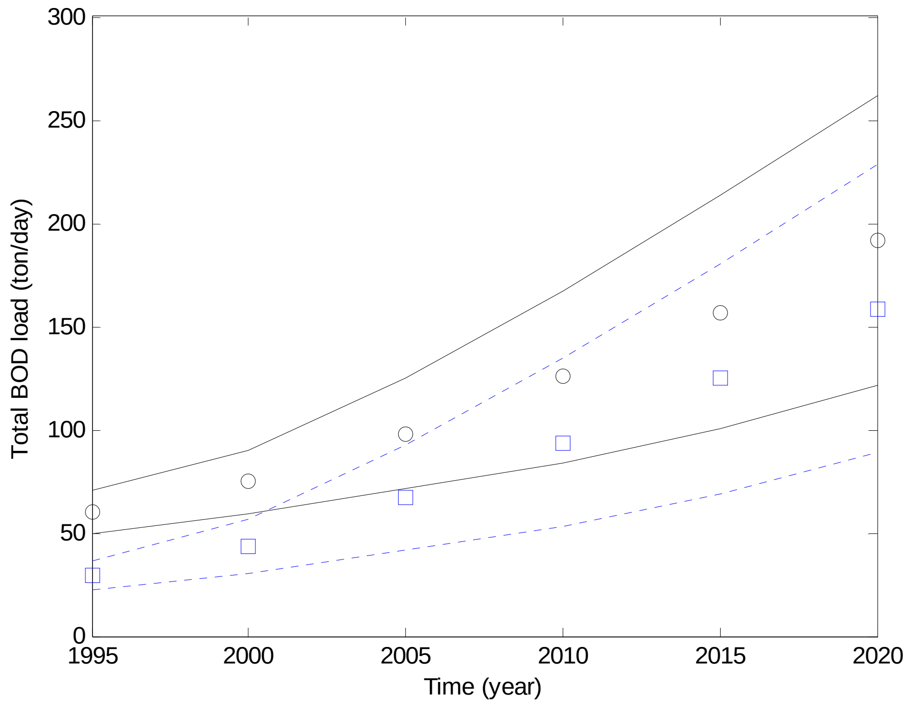

Fig. 3.5. Results of the uncertainty analyses for the total BOD load to the coast for two cases: a) with the implementation of wastewater treatment plants of 145000 m^3/day (dotted lines: 95 % confidence bounds, and □: mean), and b) without this measure (solid lines: 95 % confidence bounds, and o: mean).

3.3.4. Parameter-Verification test

The first round of the Morris method results in the most important factors that have an effect on the total BOD load and the living coral reef area, as shown in Figs. 3.1 and 3.3 (dot), respectively. The corresponding orders of importance of these factors result from the model as well as the analyst's errors, as explained previously. To reduce the analyst's error in estimating the ranges of parameters and inputs, a comparison of the results of the first round and the opinions of the local stakeholders and experts was used as a clue for the investigation. For the total BOD load, all parameters and inputs which belong to the three important processes, suggested by the local stakeholders/experts were subject to a careful examination. A number of refinements to the uncertainty range of these parameters and inputs were made. For example, the literature study (Fung-Smith and Briggs, 1996; Otte, 1997) revealed an overestimation of factor 124 (amount of BOD generated by 1 kg of shrimps). In contrast, the industrial investment (factor 5) was overlooked by assigning it a small range. Similarly for the living coral reef area, factor 133 (damaged surface area of coral reef per fish blast) was overestimated, while factor 135 (number of fish blasts per ha per year) was underestimated (Pet-Soede et al., 1999). The natural growth rate (factor 132) and the recovery rate of damaged coral (factor 134) are also adjusted according to Saila et al. (1993) and Fox et al. (2003). After refining all the ranges of the important factors discovered in the first round of the Morris analysis and the local stakeholders/experts' opinions, the second round was carried out. The results are shown in Fig. 2, for BOD load and Fig. 3.3 (star) for the total area of living coral reefs. Table 3.2 shows that the percentage of urban households connected to the water network is a strong determinant of the total BOD load. This percentage is treated as a constant parameter in RaMCo. It should be converted to an input variable which is driven by socio-economic and policy options in RaMCo.

3.3.5. Behaviour-Anomaly test

As shown in Fig. 3.2, the uncertainty of the BOD load is reduced remarkably due to the refinement of the uncertainty range of input variables and parameters. The order of importance of the relevant processes has changed, in comparison to the results from the first round of the Morris analysis. The first impression is that there is an agreement between the model and the stakeholders/experts (Table 3.2) about the most important source of organic pollution, originating from domestic wastewater. However, there is a disagreement about the order of importance of industrial wastewater and shrimp-pond culture. There are three possible reasons to explain this difference. First, the shrimp-ponds are located along the coast, whereas domestic and industrial wastewater discharges come from the city of Makassar. This may distort the experience of the experts with regard to the order of magnitude. Second, the assumption about the proportional relationship between shrimp yield and BOD load may not be valid. Empirical data and research on this relationship are lacking in the scientific literature, so it requires further investigation. Third, the variability of the BOD concentration of industrial wastewater is very large and strongly dependent on the types of industry prevailing in the study area. The analysis of BOD concentration of industrial wastewater was based on a previous investigation of industrial sectors carried out by JICA (1994) which, according to its authors, should be interpreted carefully. Therefore, more research on this topic should be conducted. Obvious flaws in the model cannot be found in this case, but suggestions for further investigation are justifiable.

For the important factors influencing the area of living coral reefs, there is agreement that blast fishing is the most influential process. A comparable result is obtained on the natural growth rate and the recovery rate of damaged coral (Fig. 3.3 and Table 3.3). However, a shortcoming of RaMCo is that it does not include the process of fishing using cyanide poison, which is evaluated as being more important than the natural growth rate and the recovery rate by both stakeholders and experts. The effect of suspended sediment on the living coral is ranked differently by stakeholders and experts (Table 3.3). The results of the model agree more with the stakeholders' assessments. Nevertheless, the results call for an in-depth investigation of the effect of the suspended sediment on the living coral reef for the study area. The most important conclusion drawn from this test is the incompleteness of RaMCo because of the exclusion of the effect of cyanide fishing.

3.3.6. Policy-Sensitivity test

As depicted by Fig. 3.4, the difference between the extended current situation and the situation with an enforcement of the ban on blast fishing is clear. There is no overlap between the two confidence bounds. It makes the decision-makers more confident in using the model, since the uncertainty in the model parameters still allows for distinguishing measures.

For the BOD load depicted in Fig. 3.5, there is a large overlap between the two situations where construction of urban wastewater treatment plants is present and not present. The difference between the two predicted average time series of BOD load is small compared with the overlapping of the confidence bounds after the year 2005. This suggests that this measure should not be implemented separately. This calls for combining measures, such as the installation of industrial wastewater treatment plants and water treatment structures for the shrimp-pond culture area. In this case, this test does not increase the confidence of the decision-makers.

3.4. Discussion and conclusions

Based on the results the following recommendations for the model improvement are made to increase the confidence of end-users in the usefulness of RaMCo for managing the coastal zone of South-west Sulawesi. First, the assumption that the BOD load generated by shrimp-pond culture is proportional to the shrimp production seems to have weak empirical support. This should be changed. Second, poison fishing should be included in the model to make the model more complete. Third, the percentage of urban households connected to the water network is a strong determinant of the total BOD load. This constant parameter should be converted to an input variable which is driven by socio-economic and policy options in RaMCo. Besides, the results suggest further work on a detailed survey of the prevailing industrial sectors and an in-depth investigation of the suspended sediment transport process influencing the living coral cover in the study area.

The examples clearly demonstrate that the Morris (1991) method can be a valuable tool in the validation of an Integrated Systems Model. First, it helps to pinpoint the parameters, inputs and measures that need careful investigation in the process of model validation. Second, it allows the end-users of the model to judge qualitatively the validities of the hypotheses embedded in a model. Third, it helps to find the backbone of a model on which the validation should be based. Besides, the Morris method has the following advantages over other conventional SA methods: computational efficiency, the problem of unit dependence is overcome, the degree of uncertainty in each input is taken into account, the non-monotonic problem is resolved, and the behaviour of the model over the entire response surface is taken into consideration. The two disadvantages, as pointed out by Campolongo and Saltelli (1997), are: it gives qualitative answers and it is unable to separate interactive effects from nonlinear effects. However, these do not counterbalance the advantages.

The current method of expert elicitation does not take into account two aspects of the expert opinion, namely, bias and inconsistency. However, it is simple, informative, and time and cost effective. Given its purpose as an exploratory tool, it is acceptable for this type of application. Alternative methods such as Delphi and Adaptive Conjoint Analysis may further improve the credibility of the results.

The proposed approach for the validation of Integrated Systems Models is a combination of sensitivity and uncertainty analyses with three of the validation tests for system dynamics models proposed by Forrester and Senge (1980). Taking into account the increasing difficulties in collecting data for empirical validation of ISMs, the current approach is one of the possible ways to get out of the “impasse” mentioned by Beck and Chen (2000). Our argument for the current approach is that one main purpose of ISM validation is to show transparently both the strengths and weaknesses of a model to its users. To the model builders, validation can reveal flaws in the model, from which they may see a need to improve or rebuild it. To the analysts, validation can provide the necessary information to facilitate the process of calibration for other applications, and analysis of the results before transferring them to the decision-makers. Finally, to the decision makers, validation informs them of the degree of confidence in using the model results to support their decision-making processes. This argument is in line with the current view that the validation of integrated assessment models is a process, not a final product of integrated assessment (Parker et al., 2002); and one important component of it is adaptive feedback between stakeholders and researchers (Jakeman and Letcher, 2003). Another approach to the validation of Integrated Systems Models, which uses the expert’s knowledge in the form of qualitative scenarios, has been proposed by the authors (Nguyen et al., 2005).

3.5. Appendices

Appendix A: Example of the questionnaire

In order to make the RaMCo a useful tool in practice, we would like to have your valuable contributions to the process of model validation by thoroughly filling this questionnaire.

No.	*Question*	*Answer*
A	What is your name?	
B	What is your title? (e.g. Prof., Dr., Deputy head of the department)	
C	Where do you work? (e.g. Department of Forestry, UNHAS University)	

D: What is/are your field(s) of expertise?	Marine ecology	Land use management	Marine water quality	Marine fisheries	Other (please specify):
E: How long have you been working in these field(s)?					

Coral reefs

In this section, you are asked for **the relative importance order** of factors and processes that have effects on coral reefs. Please answer these questions by marking them in the appropriate places.

No.	*Question*	*Answer*	
33	Do you have knowledge of the coral reef?	YES	Please go on to question 34
		NO	Please go on to question 47
34	Where did you obtain your knowledge to answer these questions? (Multiple answers possible)	Information gathered in practice	Information gathered through research
35	Are you interested in coral reefs?	Very Interested	
		Interested	
		Moderate	
		Little	
		Not at all	

No.	***Factor/ process***	***1: super important***	***2: very important***	***3: important***	***4: not so important***	***5: not important***	***6: I have no idea***
36	The impact of suspended sediment on coral reefs						
37	The fisheries using dynamite						
38	Cyanide fishing						
39	The expansion of the coral reef area						
40	The recovery rate of damaged coral						
41	The use of coral for the supply construction						

There may also be some factors/processes we overlooked. Please add them to the list and explain how important these factor/processes are, by giving them a ranking too!

No.	**Factor/process**	**1**	**2**	**3**	**4**	**5**	**6**
42							
43							
44							
45							
46							

Appendix B. Weighting factors for aggregation of expert opinions

Table.1. Weighting factor for Professional Title (PT)

Stakeholders/ Policy Makers	***Research Experts***	***Weighting Factor***
Head of an institution	Professor	2.0
Head of a department	Doctor	1.5
Staff member	Master of Science/Engineer	1.0

Table.2. Weighting factor for Source of Knowledge (SK)

Source of Knowledge	***Weighting Factor***
Information gathered from practice	1.0
Information gathered from research	1.0
Information gathered from both practice and research	2.0

Table.3. Weighting for Years of Experience (YE)

Time active in field of expertise	***Weighting Factor***
0 – 5 years	0
5 – 10 years	0.5
10 – 15 years	1.0
15 – 20 years	1.5
More than 20 years	2.0

Table.4. Weighting factor for Level of Interest (LI)

Level of Interest	***Weighting factor***
Very interested	2
Interested	1.5
Moderately	1.0
Slightly interested	0.5
Not at all interested	0.0

Chapter 4

A new approach to testing an integrated water systems model using qualitative scenarios

Abstract: Integrated systems models have been developed over decades, aiming to support the decision makers in the planning and managing of natural resources. The inherent model complexity, lack of knowledge about the linkages among model components, scarcity of field data, and uncertainty involved with internal and external factors of the real system call their practical usefulness into doubt. Validation tests designed for such models are just immaturely developed, and are argued to have some characteristics that differ from ones used for validating other types of models. A new approach for testing integrated water systems models is proposed, and applied to test the RaMCo model. Expert knowledge is elicited in the form of qualitative scenarios and translated into quantitative projections using fuzzy set theory. Trend line comparison of the projections made by the RaMCO model and the qualitative projections based on expert knowledge revealed an insufficient number of land-use types adopted by the RaMCo model. This insufficiency makes the model inadequate to describe the consequences of the changes in socio-economic factors and policy options on the erosion from the catchment and the sediment yields at the inlet of a storage lake.

4.1. Introduction

As every model is an abstraction of a real system, model developers and model users have to struggle with the question of how to validate a model. This methodological problem is argued to root in the controversial debate on justification, verification of scientific theories, and of models in a philosophical perspective (Barlas, 1994; Kleindorfer et al., 1998). The usefulness of endeavour to prove the validity of any predictive model of a natural system (open system) has been questioned (Konikow and Bredehoeft, 1992; Oreske et. al, 1994). Several authors have suggested that model validity should always be considered within the model's applicability domain or model context (Rykiel, 1996; Refsgaard and Henriksen, 2004). In addition, the purposes of a model are essential in the selection of appropriate validation tests (Nguyen and De Kok, 2003). Depending on different classification criteria, model validation tests can be categorized as qualitative or quantitative, formal or informal, static or dynamic, theoretical or operational, and so on. Traditional statistical methods are proved to have a limited capacity in testing integrated dynamics models (Forrester and Senge, 1980). One of the reasons is that both system dynamics models and Integrated Systems Models do not strive for prediction of future values – that is, not for "point-prediction". These models should predict certain aspects of behaviour in the future. Examples included

The main part of this chapter has been submitted as:
Nguyen, T.G and De Kok, J.L., Titus, M.J., 2005. A new approach to testing an integrated water systems model using qualitative scenarios. Journal of Environmental Modelling and Software. (In revision).

pattern-prediction and event-prediction. Another reason is that statistical tests hardly say anything about the structural errors within a model. The problem of equifinality (Refsgaard and Henriksen, 2004) – structural errors and errors in parameter estimation compensating for each other - is often encountered. This is even more of a problem in the case of integrated models in which many submodels are linked together to predict management variables.

Integrated systems models (ISM) and integrated water systems (IWS) models have been developed over decades, aiming to support the decision-makers in the planning and managing of natural resources. Without effective validation, the design of an IWS model remains an art rather than a science. Validation of IWS models is useful for their theoretical improvement. Moreover, the validation is necessary prior to any practical implementation of these models. Inherent model complexity, scarcity of field data, and uncertainty over internal and external factors of the real system make the validation of an IWS model a difficult task. Furthermore, the poor predictive ability of the historical data to describe future situations in the complex system involved with social and economic factors hinders the effectiveness of available validation techniques. On the other hand, due to their characteristics, validation tests for IWS models can go beyond the tool kit of available validation tests for conventional process models (Forrester and Senge, 1980; Beck and Chen, 2000). Therefore, the validation of IWS models is likely to depend less on conventional and classical tests, and more on integrated validation tests that are yet to be developed (Parker et al., 2002). In this paper, a new approach for testing IWS models is developed and applied to validate the RaMCo model. The approach is designed to test the capability of the model to describe dynamic behaviours of system output variables under a variety of possible socio-economic scenarios and policy options. The sediment yield at the inlet of the Bili-Bili dam, one of several state objective variables in the model, is selected as a case example.

This chapter is organized as follows. Section 4.2 is devoted to an overview for the new approach and the description of procedural steps to be taken during the testing process. A case study is then introduced in Section 4.3, in which the conceptual model, the mathematical equations used in RaMCo to model land-use change dynamics, the link to soil loss computation, and the sediment yield at the inlet of the storage lake are explained. Section 4.4 describes the process of formulating qualitative experts' scenarios. Translating these into quantitative projections of objective variables using fuzzy set theory is demonstrated in Section 4.5. The comparison of experts' projections and RaMCo projections in terms of trend lines are presented in Section 4.6. The chapter is concluded with a discussion on the usefulness of the validation approach and recommendations for further improvement of the RaMCo model.

4.2. Validation methodology

4.2.1. Overview of the new approach

The design of our new approach was motivated by the three reasons that limit the relevance of the conventional approaches to the validation of ISM: i) the limited predictive ability of historical data to describe the future behaviour of interactive natural-human systems, ii) the qualitative nature of the social science and iii) the scarcity of field data for model validation. The proposed approach acknowledges that we cannot develop any model which is a true representation of the real system. Validation tests should be designed to unravel the incompleteness of or errors in a model in the view of the system experts. The ultimate objective of IWS validation, according to Forrester and Senge (1980), is to obtain a better model, which has sound theoretical content (model structure) and can fulfil its intended purpose(s). One aspect of model validation is to determine whether a model is ill or well designed for its purpose (Beck and Chen, 2000). The validity of a model cannot be achieved by conducting a single test, but a series of successful tests could increase the users' confidence in the model's usefulness. The new approach presented in this chapter is developed to determine whether a model is ill or well designed, with regard to the purpose of an IWS model as a tool capable of reflecting the system experts' consensus about the dynamic behaviours of system output variables, under a set of possible socio-economic scenarios and policy options.

The underlying principle of the new approach is that system experts are asked to make an artificial closed system (hypothesised system) with the system's components, prescribed system inputs (drivers), driving mechanism, and the qualitative response of system's outputs in the form of qualitative scenarios. Fuzzy logic is applied to produce quantitative projections of the output variables from qualitative descriptions of the hypothesised system. The creation of the hypothesised system provides a platform on which "experiments" can be conducted to obtain the system's outputs under the feasible sets of system's inputs. In each experiment, the socio-economic factors and policy options are input by the experts, reflecting one possible future description of the real system. The comparisons of the trend lines between the two systems' outputs under different scenarios are made to arrive at the plausibility of the model structure and the validity of the assumptions. Thus, an obvious difference between the outputs produced by the two systems, in terms of trend lines, can reveal the structural faults of the model system. Otherwise, the model is said to pass the current test. The procedural steps to build an experts' hypothesised system, to use qualitative scenarios and a fuzzy rule-based method to make quantitative projections of system behaviours are presented in the following subsection.

4.2.2. The detailed procedure

There are three phases to be taken during the testing process of an IWS model using qualitative scenarios: 1) formulating experts' qualitative scenarios; 2) translating the qualitative scenarios; 3) conducting simulations by the IWS model and comparing the outputs produced by the two systems in term of trend lines.

Formulating experts' qualitative scenarios

In the context of this chapter, a scenario is defined as '*a description of a future situation and the course of events which allows one to move forward from the original situation to the future situation*' (Godet and Rounbelat, 1996). Qualitative scenarios describe possible futures in the form of words or symbols while quantitative scenarios describe futures in numerical form (Alcamo, 2001). The common understanding is that a scenario is not a prediction of the future, but an alternative image of how the future might unfold. The purpose of scenarios is manifold. Some of them are: illustrating how alternative policy pathways can achieve an environmental target, identifying the robustness of policies under different future conditions, providing the non-technical audience a picture of future alternative states of the environment in an easily-understandable form (narrative description), and providing an effective format on which information in both qualitative and quantitative forms can be assimilated and represented. In this paper, scenarios are proposed as testing experiments to test the capability of an ISM to describe the consequences of possible socio-economic conditions and policy options on the management variables.

A good scenario should be relevant, consistent (coherent), probable and transparent. In principle, only a few substantially different scenarios are needed. Although different authors (Von Reibnitz, 1988; Van der Heijden, 1996; Alcamo, 2001) developed somewhat different procedures and terminologies for the scenario building, these procedures share the same iterative form and have the following steps in common:

1) Establishing a scenario building team and defining the goals of scenarios
2) Analysing data and studying literature
3) Specifying driving forces and driving mechanism (structuring scenarios)
4) Developing the storylines (scenarios in narrative form)
5) Testing the internal consistency of scenarios

In applying scenarios for testing IWS models, the composition of the scenario building team (step 1) and testing the consistency of scenarios (step 5) are particularly important, which require more elaboration.

The participatory approach to scenario building is widely acknowledged, which requires a wide spectrum of knowledge and opinions from multidisciplinary team members (Schwab et al., 2003; Van der Heijden, 1996). In developing scenarios used in international environmental assessment, Alcamo (2001) recommends having two building teams: a scenario team and a scenario panel. The former, which consists of the sponsors of the scenario building exercise and experts, should include around three to six members. The latter, which consists of stakeholders, policymakers and additional experts, should include around fifteen to twenty-five members. For the purpose of testing ISMs, we propose to distinguish two groups in the scenario building team. The first group includes model developers (they are also interdisciplinary scientists), experts (scientists who may have different views about the model system) and additional analysts (scientists who are not involved in the model building). The second group consists of multidisciplinary experts, resource managers and stakeholders. The second group can play a role both as the fact-contributor and scenario evaluator in the scenario building for the testing of ISMs. Preferably, the stakeholders and resource managers should participate at the beginning of the scenario building process (steps 1 to 3).

In the iterative scenario building process, the consistency of the scenarios need to be tested. Van der Heijden (1996) and Alcamo (2001) recommend two similar approaches to establishing the consistency of scenarios, which include two supplementary tests: scenario-quantification testing and actor-testing. Quantification testing comprises quantifying the scenarios and examining the quantitative projections of the system indicators (management variables). Actor-testing diagnoses the inconsistencies by confronting the internal logic of the qualitative scenarios with the intuitive human ability to guess at the logic of the various actors (stakeholders, resource managers and additional experts). We propose to use physical, biological constraints (e.g. the total available area of a watershed) to check the quantitative projections (e.g. the projections of the areas of different land-use types) for quantification testing. In actor-testing, both the narrative descriptions of the scenarios and the quantitative projections of the system indicators should be communicated to the second group (stakeholders, resource managers and additional experts) by means of report papers, workshops and the internet.

Translating qualitative scenarios

For the translation of qualitative scenarios, the application of fuzzy set theory is proposed. Fuzzy set theory was originally developed by Zadeh (1973), based on the concepts of classical set theory. The essential motivation, as he claimed, for the development of fuzzy set theory is the inadequacy and inappropriateness of conventional quantitative techniques for the analysis of mechanistic systems (e.g. physical systems governed by the laws of mechanics) to analyse humanistic systems. The design of a fuzzy system comprises five steps (Mathworks, 2005), which can be reduced to four main steps (De Kok et al., 2000):

1) Translation of the independent and dependent variables from numerical into the fuzzy domain (fuzzification)
2) Formulation of the conditional inference rules
3) Application of these rules to determine the fuzzy outputs
4) Translation of the fuzzy outputs back into the numerical domain (defuzzification)

In order to test the internal consistency of scenarios, scenario quantification-testing needs to be conducted. Therefore, the process of scenario translation is extended to include step 5 (testing the internal consistency of scenarios). These five steps are demonstrated by the application described in Section 4.5.

Conducting simulations by the ISM and comparing the results

After translating the qualitative scenarios to get the quantitative projections of the output variable, simulations made by the IWS model are conducted. Comparison of the output behaviours produced by the two systems in terms of trend lines is carried out. This phase will be demonstrated in Section 4.6.

It is recommended that the interactive communication within the first group should be carried out in all three phases (qualitative scenario building, scenario translating and comparing results). In doing so, any possible disagreements between model developers and experts can be brought out for discussion at every step. Thus, the biasness or inconsistency of the expert(s) can be minimised.

4.3. The RaMCo Model

The study area for RaMCo occupies a total area of about 8000 km^2 (80km x100km), of which more than half is on the mainland (De Kok and Wind, 2002). The offshore part covers the Spermonde archipelago where multi-ecosystems such as coral reef, mangrove and seagrass can be found. On the mainland, the city of Makassar has a fast-growing population of 1.09 million (1995), which is expected to double in twenty years. In the upland rural area, the forest area is rapidly declining, due to the increase in cultivated land. The expansions of urban areas and the conversion of uncultivated to cultivated land are imposing a strong demand on the effective management of water and other ecological systems in the coastal area.

To meet the rapidly increasing demand for water supplies for domestic uses, industry, irrigation, shrimp culture and the requirements for flood defence of the city of Makassar, the construction of a multi-purpose storage lake started in 1992. The dam was closed for water storage in November 1997 (Suriamihardja et al., 2001). The watershed of the Bili-Bili dam covers the total area of 384 km^2, which represents the upper part of the Jeneberang river catchment. The dam was designed to have its effective storage capacity of 346 million m^3 and dead storage capacity of 29 million m^3 (CTI, 1994). Its expected lifetime of 50 years was determined by computing the total soil loss due to erosion of the watershed surface. The computation was carried out using the Universal Soil Loss Equation (USLE) in combination with the land cover map surveyed in 1992. No future dynamic development of land-use in the watershed area was taken into consideration. Analyses of recently measured sediment transport rates at the inlet of the Bili-Bili dam and land-use maps show an obvious decrease in the storage capacity of the dam, due to increasing sediment input (CTI, 1994; Suriamihardja et al., 2001). This calls for a proper land-use management strategy to minimise the sediment eroded from the watershed surface that runs into the reservoir.

RaMCo quantitatively describes the future dynamic land-use and land-cover changes under the combined influence of socio-economic factors. Then, the resulting soil losses from the watershed surface and the resulting sediment yields at the inlet of the Bili-Bili dam are computed. The following are conceptual and mathematical descriptions of this chain-model.

4.3.1. Land-use/land-cover change model

Land-use types

During the design stage, a problem-based approach was followed to select relevant land-use-types (De Kok et al., 2001). In RaMCo, a distinction was made between static land-use types (land-use features) and active land-use types (land-use functions). Land-use features such as beach, harbour and airport are expected to be relatively stable in their size and location over the time frame considered. Land-use functions such as industry, tourism, brackish pond culture, rice culture and others are expected to change both in space and over time under the influence of various internal and external driving factors (drivers). In this paper, attention is paid to the two land-use types: nature and mixed agriculture. The model treats the “nature” land-use type as the uncultivated land which is a combination of natural forest, production forest, shrubs and grasses. Mixed agriculture represents food crop culture (other than rice culture) such as maize, cassava and cash crops such as coffee and cacao. These types of land-use predominate in the Bili-Bili catchment and are expected to change rapidly, affecting the amount of sediment transported into the reservoir. In addition to the two defined categories, three other land-use types exist in RaMCo: rural residential, rice culture and inland water.

Drivers of land-use changes: temporal dynamics versus spatial dynamics

The drivers of land-use changes in the RaMCo model can be separated into three categories: i) socio-economic drivers, such as price, cost, yield, technology development and demography; ii) management measures, such as reservoir building and reforestation; and iii) biophysical attributes, such as soil types and road networks. The first two groups of drivers, in combination with the availability of irrigated water and suitable land, determine the rate of land-use change (temporal dynamics), while the final group determines places where the changes take place (spatial dynamics). The rate of change in area for each land-use type is computed by a so-called macro-scale model, which is discussed in more detail below. In the micro-scale model, the spatial allocations of these changes are determined by adopting the constrained cellular automata (CCA) technique. A full description of this technique is outside the scope of this paper. Those who are interested in the details of the CCA approach and the model structure are referred to White and Engelen (1997) and De Kok et al. (2001).

Macro-scale model

As mentioned above, the macro-scale model computes the rates of change, i.e. land demand for different land-use types. Since this chapter focuses on land-use change and the resulting soil loss in the Bili-Bili watershed area, only three land-use types are discerned in the following section, namely mixed agriculture, rice culture, and nature. Inland water and rural residential land-use types are included in RaMCo but excluded in this discussion because of the small portions of land they occupy in the basin and their relative stability in size and locations.

For agricultural land-use, following the assumption that the land demand is proportional to the net revenue per unit area, the rate of change in land-demand can be computed as (De Kok et al., 2001):

$$\Delta A(t) = \alpha\big(p(t)y(t) - c(t)\big)A(t)\left[1 - \frac{Z(t)}{Z_{tot}}\right] \tag{4.1}$$

where $\Delta A(t)$ and $A(t)$ are the rate of change and area of mixed agriculture at time t, $p(t)$ and $c(t)$ are price and production cost per unit area, and $y(t)$ is the yield which can accommodate technological changes. The growth coefficient α was calibrated using statistical data on the above defined variables. The variable $Z(t)$ is the sum of geographical suitability for agriculture over all cells occupied by agriculture at time t, and Z_{tot} is obtained by extending the sum over all cells on the map. The use of these variables ensures that expansion ceases if the maximum suitable area is approached.
For rice culture, Eq.(1) is still applicable but rice yields are obtained in a different way to account for the irrigation function of the storage lake:

$$y_{rice}(t) = f(V)\eta(t)y_{irr} + \big(1 - f(V)\eta(t)\big)y_{nirr} \tag{4.2}$$

In Eq.(2), y_{irr} and y_{nirr} are the maximum yields of rice culture with and without irrigation respectively. The dimensionless function $f(V)$ has a value ranging from 0 to 1, and reflects the irrigation priority using the actual and maximum volumes of the storage lake. The variable $\eta(t)$ denotes the spatial fraction of rice fields which can be irrigated.

The land demand of "nature" land-use type is computed by:

$$\Delta A_n(t) = \alpha A_n(t)\left[1 - \frac{Z_n(t)}{Z_{n,tot}}\right] + \delta_n(t) \tag{4.3}$$

where α is the natural expansion rate of nature (forest), and $\delta_n(t)$ accounts for the area of reforestation at time t, a management variable.

According to these equations, each sector can expand until the maximum suitable area is reached. This allows for a situation where more or less land is allocated to all the sectors taken together than the total available land. Thus, an allocation mechanism has been introduced. If the total computed land demand is less than the available land, the allocated land equals the demands for these sectors. The remainder is assigned to nature (forest). In case total computed land demand for all sectors exceed the available area, the allocated land for each sector is normalised as follows:

$$A_i(t) = \frac{A_{available}}{\sum \overline{A_i(t)}} \overline{A_i(t)} \tag{4.4}$$

where $A_i(t)$ and $\overline{A_i(t)}$ are allocated land and computed land demand for land-use type i respectively.

4.3.2. Soil loss computation

To couple the process of land-use changes to predict the sediment yields at the outlet of Bili-Bili watershed area, the Universal Soil Loss Equation (USLE) in spatially distributed form is used. The original USLE (Wischmeier and Smith, 1965) has the following equations:

$$A = R.K.L.S.C.P \tag{4.5}$$

where A is the computed soil loss per unit area, expressed in metric tons/ha; R is rainfall factor, in MJ-mm/ha-h and MJ-cm/ha-h if rainfall intensities are measured in mm/h and cm/h respectively; K is the soil erodibility factor, in metric tons-h/MJ-cm; C is a cover management factor (-); P is a support practice factor (-); L is the slope length factor, in m, and S is the slope steepness factor. The product of L and S is computed by:

$$LS = \left(\frac{\lambda}{22.13}\right)^m \left(0.0065 s^2 + 0.045 s + 0.065\right) \tag{4.6}$$

in which λ is the field slope length, in m, and *m* is the power factor whose value of 0.5 is quite acceptable for the basin with a slope percentage of 5% or more (Wischmeier and Smith, 1978); *s* is the slope percentage.

The RaMCo model allows the use of spatial databases to facilitate the computation of soil erosion from individual (400mx400m) mesh cell. Maps containing factors on the right hand side of Eq. (4.5) are referred to as factor maps. These factors maps were derived from spatial databases such as topographic maps, geological maps, land-cover maps, and isohyetal maps (CTI, 1994). Equations 4.5 and 4.6 are used to compute soil loss from every cell in the map.

4.3.3. Sediment yield

To predict sediment yields at the outlet of the watershed the Gross Erosion-Sediment Delivery Method (SCS, 1971) is used in combination with the USLE. The gross erosion (E), expressed in metric tons, can be interpreted as the sum of all the water erosion taking place such as sheet and rill erosion, gully erosion, streambank and streambed erosion as well as erosion from construction and mining sites (SCS, 1971). According to the previous study on sediment in the Jeneberang river (CTI, 1994), the sediment consists mainly of washload caused by sheet and rill erosion. Moreover, sand pockets and Sabo dams were designed to trap coarser sediment resulting from other types of erosion. Thus the neglecting of other erosion types is acceptable with respect to our purpose of estimating the sediment yield at the inlet of the Bili-Bili Dam site. The sediment yield (*Sy*), the amount of soil routed to the outlet of the catchment in metric tons per ha, can be computed by multiplying the gross erosion (E) by the sediment delivery ratio:

$$S_y = E.SDR \tag{4.7}$$

where *SDR* is the sediment delivery ratio, which depends on various factors such as channel density, slope, length, land-use, and area of the catchment. Methods have been proposed in the past to estimate the SDR (SCS, 1971). This research adopts the values established in Morgan's (1980) table (CTI, 1994), which is widely used in Indonesia. In order to identify the areas that are susceptible to erosion for the development of soil conservation strategies, the whole basin was subdivided into eight sub-basins. Equation 7 is applied to each sub-catchment, and the sediment yields are added together to obtain the total sediment yield running into the reservoir.

4.4. Formulation of scenarios for testing

The iterative processes of qualitative scenario formulation have commonly five steps (Section 4.2.). In step 1 (establishing a scenario building team) of this exercise, two groups were distinguished. The first group consists of a model developer, an expert and an analyst. The second group consists of around twenty local scientists and potential end-users of RaMCo. Due to practical reasons (e.g. distance, finance), the second group only participated in step 5 of the current exercise. In step 2, extensive data collection and historical study were carried out for the study area as well as for other regions (e.g. Yoyakarta and Sumatra) in Indonesia. In this section, steps 3 and 4 (structuring scenarios and developing qualitative scenario) are described. Since step 5 (testing consistency of scenario) is involved with scenarios quantification, it is described in the end of Section 4.5.

4.4.1. Structuring scenarios

As mentioned in the Section 4.3, in the Bili-Bili catchment, five land-use types were distinguished by modellers, which include nature (forest), agriculture, rice culture, rural residential land, and inland water. This categorization may or may not be sufficient to give a satisfactory description of the real system, given the specified purpose of the model. According to the expert, the separation of nature into forest and shrub and grassland, and the separation of agriculture into dry upland farming and mixed forest garden are necessary to describe the effect of management measures on land-use changes and the resulting dynamic change in soil erosion from the catchment surface. Thus, the new hypothesised land-use system consists of five active types: forest, shrub and grassland, dry upland farming, mixed forest garden, paddy field and two relatively static types: inland water and rural residential.

The drivers and driving mechanism of the land-use system are shortly described in figure 1, which is the result from extensive discussions with in the first group.

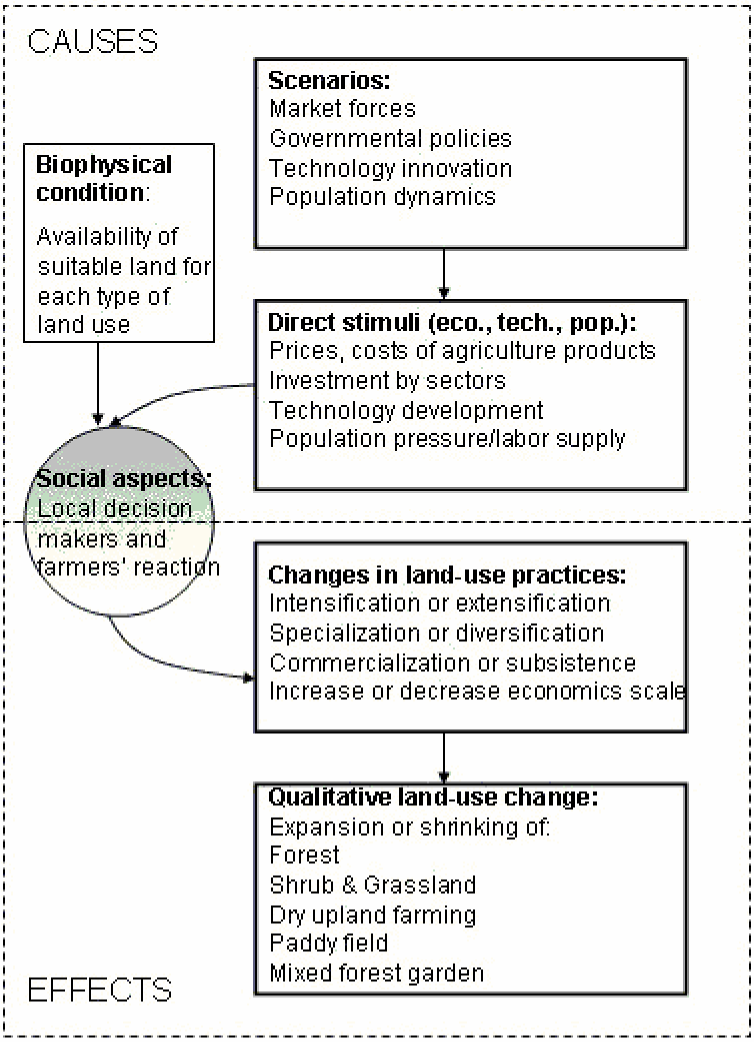

Figure 1. Reasoning process underlying the scenario-based qualitative projection of the rates of land-use changes.

4.4.2. Developing qualitative scenarios

Based on the purpose of the scenarios and the insights gained from field research, three qualitative scenarios were formulated for the dynamic land-use system in the Jeneberang catchment. Scenario A reflects an extrapolation of the socio-economic, policy conditions and their effects on the land-use system under the Suharto presidency period (1967-1998). Scenario B represents the post-Suharto period (present situation), in which the forest is more open for logging and is invaded by subsistence farming due to the maximum economic growth objective and the lack of law enforcement from the government. In scenario C, a sustainable development option is projected in which an economic goal is achieved while the environmental issues are kept to a minimum through policy measures such as law, cheap credits and land-conversion programmes.

Scenario A: guided market economy

The guided market economy as developed during the New Order, has been based on strong government interferences and a bureaucratic approach, causing much abuse of power and funds and often leading to misinvestments. On the other hand it should be acknowledged that government programmes focusing on the boosting of food production, infrastructure, public services (health and education) and industrialisation have had positive impacts in terms of employment creation and income improvements. Environmental conditions (pollution, deforestation and erosion) however, usually have been neglected, as have most issues of regional and social equity. This scenario is assumed to cause the following shifts and changes in land use practices:

Forest: a gradual retreat of primeval and secondary forest fringes due to the progressive invasion by marginalised upland farmers in search for timber, firewood and land to cultivate food and cash crops

Shrubs and grasses: Expanding in the higher uplands because of the abandonment of exhausted and unproductive dry farming fields left in fallow. Retreating in the lower uplands through their conversion in mixed forest garden.

Dry upland farming (*Tegalan*): Expanding *tegalan*-fields in the higher uplands because of land hunger of small peasants and the stimulation of dry food crop cultivation by government programmes.

Mixed forest gardens: Some expansion may occur by planting of lucrative tree crops like cocoa or clove. Most of this expansion will be realised on wasteland areas (shrub and grassland) or marginal *tegalan* fields at lower altitudes (< 1000 m.).

Paddy fields (sawah): Lack of irrigable land in the Jeneberang Valley and the long dry season are limiting the expansion opportunities for wet rice cultivation beyond the valley bottoms and lower slopes.

Scenario B: maximum growth

The maximum growth scenario is based on the principles of free trade, a facilitating government policy and the attraction of foreign and domestic corporate capital. Through the use of capital and technology, intensive modes of production and increasing economies of scale this will lead to higher levels of productivity and decreasing product prices. In agriculture, this implies that only the bigger farmers are able to draw advantage from this type of development (as only these farmers have enough land, capital and knowledge), whereas the smaller peasants have to revert to subsistence agriculture or labour intensive types of commercial farming with few inputs and low productivity levels.

Forests: These are increasingly affected by the expansion of subsistence farming and commercial farming in dry upland areas due to processes of marginalisation among landless and small farmers, and the expansion of cash crop cultivation.

Shrubs and grasses: This type of waste land probably will not change very much in total area for the same reasons as in scenario A.

Dry upland farming: While there is continuing encroachment of dry upland farming into the forest fringes of the higher uplands, there also is an increase in the productivity of *tegalan* agriculture on existing fields. Total *tegalan* area however, will only expand slightly due to the intensification of *tegalan* agriculture and the advancement of agro-forestry systems in the lower areas.

Mixed forest gardens: A similar expansion of agro-forestry systems on the lower slopes and foothills of the Jeneberang Valley would be expected due to the drive for increasing perennial cash-crop production for the export market (i.e. coffee, cacao and clove).

Paddy fields: Few changes can be expected in terms of areal expansion, but productivity of wet rice fields is assumed to rise considerably due to capital investments by richer farmers in high-yielding variety, fertilisers and so on.

Scenario C: sustainable development

This sustainable development scenario is based on a selective operation of the market economy in combination with an active role of the government in securing principles and conditions of sustainability. With respect to agricultural land use this policy requires that farmers are both stimulated and controlled by environmental laws, extension programmes, cheap credits and (initial) subsidies on appropriate inputs. Furthermore, the government should actively support rural economic diversification by improving the rural infrastructure, public services and human resource development, in order to reduce dependency on agriculture and pressures on local natural resources.

Forests: These will show a recovery, both in area and quality due to more strict regulations and controls on the use of existing forest areas (protected forest and production forest) and the reforestation of waste land areas (shrub and grassland).

Shrub and grasslands: This wasteland area gradually will be reduced in size and improved by regreening projects. Reduction may also be achieved by converting the waste land areas into agro-forestry systems.

Dry upland farming: *Tegalan* agriculture of annual food crops will become more productive and sustainable through improved cultivation methods, including the integration of animal husbandry, crop diversification and terracing.

Mixed forest gardens: Programs for promoting the sustainable cultivation of perennial cash crops in mixed forest gardens will expand agro-forestry systems in the foothill areas of the valley (i.e. both in the marginal *tegalan* areas and the wasteland areas).

Paddy fields: The irrigated paddy fields in this scenario will not expand very much for the same reasons as in the previous scenarios. Productivity probably will not increase as much as in scenario B, due to the limited use of chemical inputs.

4.5. Translation of qualitative scenarios

It is worth noting that depending on the analyst‘s view on how he interprets fuzziness, the details of step 1 (fuzzyfication) and 2 (formulation of inference rules) may be substantially different from the present exercise. In the absence of both statistical field data and a number of experts having a good knowledge of the field, the approach adopted by the authors here is presented as if fuzziness is subjective, context-dependent (in accordance with the original idea of Zadeh, 1973) and stems from an individual expert. However, guidelines for the design of these steps based on different views (i.e. fuzziness is objective and stems from group) are available in the literature, and are given when necessary. The following subsections give the detailed descriptions of the five steps (mentioned in the end of Section 4.2.2) applied for this example.

4.5.1. Fuzzification

The fuzzification, which can be described by the process of establishment of membership functions, requires several steps, consisting of the establishment of ranges in the numerical domains of the variables concerned, the specification of boundaries in the fuzzy domains of associated fuzzy subsets and the selection of the shape of the membership functions (MFs). For the concepts of the MFs, (Zadeh, 1973) and (Mathworks, 2005) are referred to. Here, an example is given to describe the steps to establish the MFs for one input variable (food crop price) and one output variable (the rate of change in forest area).

A major problem in establishing the possible numerical range of values for each of the input variables in the respective scenarios is that both the prices and the costs were subject to a high level of monetary inflation in the late 1990s. Consequently, these values are showing extreme fluctuations over time, which cannot simply be projected in the near future. For this reason we have presented these monetary values in terms of constant prices in 1993 (instead of current prices).

For the ranges of output variables, both statistical and spatial data obtained from survey and satellite images were used. For example, the yearly change in forested area would be negative (e.g. due to logging) or positive (e.g. due to reforestation). Data from the Division of Forestry and Land Conservation of Gowa district (2000) show an estimation of around 10 – 15% of the Jeneberang watershed area that was converted to other uses in the last 10 years. On the other hand, 2650 ha of forest was rehabilitated through replantation programmes. Taking 15% of the catchment area to represent the deforested area (5760 ha) during these 10 years, the net decrease in forest area is 3110 ha. From that, it is reasonable to have the maximum decrease of forest area for each year to be 400 ha/year. The maximum increase due to investment in reforestation, afforestation can be set at 400 ha/year, based on the same information.

In addition to the specification of the numerical ranges of variables, it is necessary to specify the boundaries of the associated fuzzy subsets. For example, from what value to what value the food crop prices can be considered to be “Low”, “Medium” or “High”. The boundaries of fuzzy subsets are allowed to have their intersection, i.e. one particular price can belong to both “Low” and “Medium” fuzzy subsets. These boundaries are often established subjectively from the experience of experts. This is the case adopted in this exercise. A less subjective example of specifying these boundaries, applying the statistical moving average technique, given the data available, was discussed by Draeseke and Giles (2002). Another requirement is the determination of the shapes of the MFs of the input and output variables. There are, in general, no rules for the selection of a shape of a membership function when little data and experts’ knowledge about a variable exist. Therefore, the symmetrically trapezoidal, triangular MFs (Aronica et al., 1998) and Gaussian MFs (De Kok, et al., 2001) are often chosen. In the present exercise the MFs of independent variables have Gaussian form, whereas trapezoidal functions are used for dependent variables. Four methods of building MFs using expert knowledge elicitation, if individual expert and groups of experts are present, are described in (Cornelissen et al., 2003).

4.5.2. Formulation of inference rules

A key step in the construction of the fuzzy system is the formulation of inference rules that reflect the mechanisms underlying the qualitative scenarios. For each scenario a set of all possible combinations of independent variables (or direct stimuli) has been defined, which may serve as a basis for assessing their impact on the five major land-use types. From these general sets a number of realistic combinations of independent variables, which are directly relevant for the dynamics in the respective land-use types are derived. The establishment of the direction and intensity of the impacts of these combinations on land-use through expert assessment is then conducted. For practical reasons, the full procedure for scenario A is presented:

Table 4.1: Set of possible combinations of independent variables for scenario A

Rule	Food crop Prices	Cash crop prices	Prod. cost	Public investment
1	M	L	L	L
2	M	L	L	M
3	M	L	L	H
4	M	L	M	L
5	M	L	M	M
6	M	L	M	H
7	M	L	H	L
8	M	L	H	M
9	M	L	H	H
10	M	M	L	L
11	M	M	L	M
12	M	M	L	H
13	M	M	M	L
14	M	M	M	M
15	M	M	M	H
16	M	M	H	L
17	M	M	H	M
18	M	M	H	H
19	M	H	L	L
20	M	H	L	M
21	M	H	L	H
22	M	H	M	L
23	M	H	M	M
24	M	H	M	H
25	M	H	H	L
26	M	H	H	M
27	M	H	H	H

In scenario A food prices are maintained at a stable Medium (M) level in order to guarantee a sufficient food supply at reasonable prices. This is achieved through import controls, input subsidies and marketing boards. Cash crop prices are fluctuating between Low (L) and Medium (M) levels, due to the suppressing impact of marketing imperfections on higher price levels. Production costs are gradually rising from L to M through the abolishment of subsidies for agricultural inputs. The labour costs are kept at a low level through the combined impact of a high rural labour surplus and a rigid control of trade union activities. Rural wages however, may increase near big cities through the impact on increasing rural-urban circulation opportunities. Public investments have been rising from L level to M level through special attention for rural public services, infrastructure and agricultural intensification programmes. But at the end of this period these investments may again decline to the L level, due to the rising importance of the urban-industrial sector. These parameters of the direct stimuli in scenario A are responsible for the fact that only rules 1, 2, 4, 5, 10, 11, 13 and 14 are relevant for this scenario. With this reduced set of rules we will finally assess their impact on the dynamics of the area expansion of the respective types of land use.

Table 4.2. Reduced set of inference rules in scenario A

Rule	Forest	Shrub and Grassland	Dry Upland Farming	Mixed Forest Gardening	Paddy fields
1	-	±	±	0	0
2	0	0	0	±	0
4	-	±	+	0	±
5	0	0	±	±	0
10	-	±	0	±	0
11	0	0	-	+	0
13	-	±	±	0	±
14	0	0	0	±	0

Notation: (+) = strong increase; (±) = weak increase; (0) = stagnant; (-) = weak decrease; (--) = strong decrease.

In a similar way we have established the relevant inference rules for the scenarios B and C, as well as their impacts on the areas of the respective land use types (Tables 4.3 and 4.4).

Table 4.3. Reduced set of inference rules in scenario B

Rule	Food crop prices	Cash crop prices	Production costs	Public Invest.	Forest	Shrub and Grassland	Dry Upland Farming	Mixed Forest Gardening	Paddy field
4	L	L	M	L	-	0	±	0	0
7	L	L	H	L	--	0	±	0	±
13	L	M	M	L	-	0	0	±	0
16	L	M	H	L	--	0	±	±	±
22	L	H	M	L	-	-	0	+	0
25	L	H	H	L	--	-	±	+	±

Table 4.4. Reduced set of inference rules in scenario C

Rule	Food crop prices	Cash crop prices	Production costs	Public invest.	Forest	Shrub and Grassland	Dry Upland Farming	Mixed Forest Gardening	Paddy field
14	M	M	M	M	0	0	0	±	0
15	M	M	M	H	±	-	-	+	0
17	M	M	H	M	±	0	0	0	±
18	M	M	H	H	+	-	-	±	0
23	M	H	M	M	±	-	-	+	0
24	M	H	M	H	+	--	--	+	0
26	M	H	H	M	+	-	-	±	±
27	M	H	H	H	+	--	-	+	0

4.5.3. Application of the inference rules

In the next two steps the calculations of the values for the output variable, which are concerned with fuzzy logic operation, are conducted with the Fuzzy Logic Toolbox embedded in MATLAP. The method adopted here is referred to as Mamdani inference (Mamdani and Assilian, 1975) and will be illustrated for an example. Considering inference rule 27 (Table 4.4):

IF (food crop price is medium) AND (cash crop price is high) AND (production cost is high) AND (public investment is high) THEN (the rate of change in forest area is strongly increased).

First the fuzzy value for the *rule antecedent*, which is the condition preceding the THEN statement, must be determined by calculating the corresponding membership function. The AND operation is implemented by taking the minimum value of the membership values for the four independent values:

$$\mu_{AND} = \min[\mu_1(x_1), \mu_2(x_2), \mu_3(x_3), \mu_4(x_4)] \tag{4.8}$$

where μ_{AND} is the membership value for the rule antecedent and $\mu_1(x_1)$ is the membership value for the food crop price corresponding to numerical value x_1.

4.5.4. Calculation of the output value

In the next step, the fuzzy value of the THEN part of the rule or the *rule consequent* must be determined. This is done by truncating the MF for the fuzzy output value (the rate of change in forest area) at the value $\mu_{AND.}$ The result is a new MF $\mu_{CONS}(y)$ for the rule consequent, where y is the value for the rate of change in forest area in the numerical domain. This procedure is repeated for each inference rule, after which the results are aggregated to a single MF by taking the maximum value of the membership values for the entire set of inference rules:

$$\mu_{OUT}(y) = \max[\mu^i_{CONS}(y)] \quad ; \; i = 1,\ldots,n \tag{4.9}$$

where n is the number of the inference rules (for example, n = 8 in Table 4.4).
The result is a single MF for the output variable which must now be translated from the fuzzy to the numerical domain (defuzzification) to allow for the comparison with the quantitative values produced by RaMCo. This defuzzification can take place in different ways. Here the output corresponding to the centroid of the output MF is used.

4.5.5. Testing the consistency of scenarios

To increase the credibility of the outputs produced by the experts' system, both actor-testing and quantification-testing were carried out. First, the land-use types, drivers, driving mechanism and inference rules of the land-use system were presented at a symposium in Makassar city.. This symposium was attended by local officials, stakeholders and scientists, ranging from forestry experts, agronomists, economists and sociologists to mathematicians, marine biologists and other natural scientists. These participants indulged in a lively debate on the merits and limitations of the respective scenarios and their assumptions, but in general recognised their local relevance and supported their main lines of reasoning. Second, the physical constraint of the total area of the basin is used to check the consistencies of inference rules and the numerical ranges of the outputs. The differences between the basin area and the total of computed land demands do not exceed ten percent of the basin area in any of the three scenarios. To use the quantitative changes in the micro-scale model (mentioned below), Eq.(4.4) in Section 4.3 is used to scale up and down so that the total computed land demands are always equal to the basin area.

4.6. Results

Quantitative changes in all land-use types projected by the experts' system need to be spatially allocated in order to compute the total soil loss and the sediment yield at the inlet of Bili-Bili dam. However, experts in socio-economic sciences have difficulty in speculating with regard to the locations where the changes should take place. This is due to the fact that the spatial distributions of land-use changes depend on biophysical aspects of the basin such as geomorphology and transportation networks. Fortunately, the Research Institute for Knowledge Systems (RIKS) has recently developed a generic tool – GEONAMICA (Engelen et al., 2004) which aims to represent spatially the quantitative changes of land-use systems in land-use maps. GEONAMICA, which adopts the constrained cellular automata approach, makes flexible use of the minimum available information such as: suitability maps, zoning maps, accessibility maps and cellular automata transition rules. The use of this tool allows the same spatial distribution mechanism as that adopted by RaMCo and the hypothesised system.

Maps of the land-cover changes produced by RaMCo and the experts' system under three scenarios were used to compute soil losses and sediment yields using the approach mentioned in Section 4.3. The final results of the dynamic development of sediment yields- the information needed by storage lake managers - are presented in figure 4.2.

It can be seen from Figure 4.2 that RaMCo can produce the trend lines of increasing sediment yields in scenarios A and B. However, for scenario C, RaMCo gives results which are contradictory to those produced by the experts' system, in terms of trend lines. It also means that RaMCo is incapable of differentiating between the consequences of scenario B and C, which, according to the experts's system, are opposite in direction.

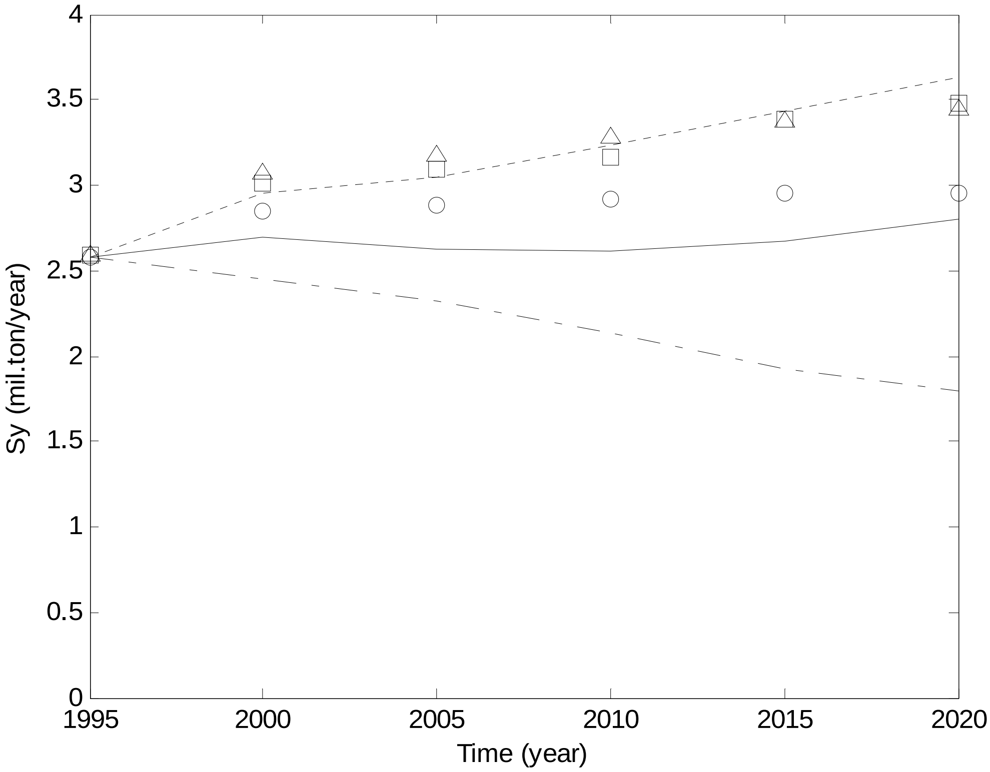

Figure 4.2. Comparisons between the sediment yields computed by RaMCo (with fixed α) under three scenarios A-guided market economy (o), B-maximum growth (□) and C-sustainable development (Δ) and the sediment yields computed by the hypothesised system for scenarios A (solid line), B (dotted line) and C (dash-dotted line).

To determine whether the aggregation level of land-use types adopted by the RaMCo model is the cause of the problem, a further analysis is conducted. The five active land-use types classified by the expert are quantitatively aggregated into three land-use types classified by RaMCo. In comparisons between the two predictions in scenario C, the land demand of land-use type "nature" is underpredicted while the land demand of "agriculture" is overpredicted by the RaMCo model. An attempt was made to reduce the growth coefficient (α) in Eq.(4.1), which was originally assumed to be constant. The reason to adjust it is that the growth coefficient originally reflected the stakeholders' "reaction" to the change in the net benefit obtained per unit area. This should take into account the control exerted by the government through environmental law, giving credits to farmers to convert from upland farming to mixed forest garden, and launching intensification of agriculture programs. It turned out that when α in scenario A is reduced slightly and α in scenario C is reduced strongly in comparison with the one in scenario B, the projections made by the RaMCo model are mostly the same as the projections made by the hypothesised system (after five land-use types are aggregated into three land-use types). The new land demands produced by the re-calibrated RaMCo are put to a micro-scale model and then to the USLE to compute the sediment yields. The new results are presented in Figure 4.3.

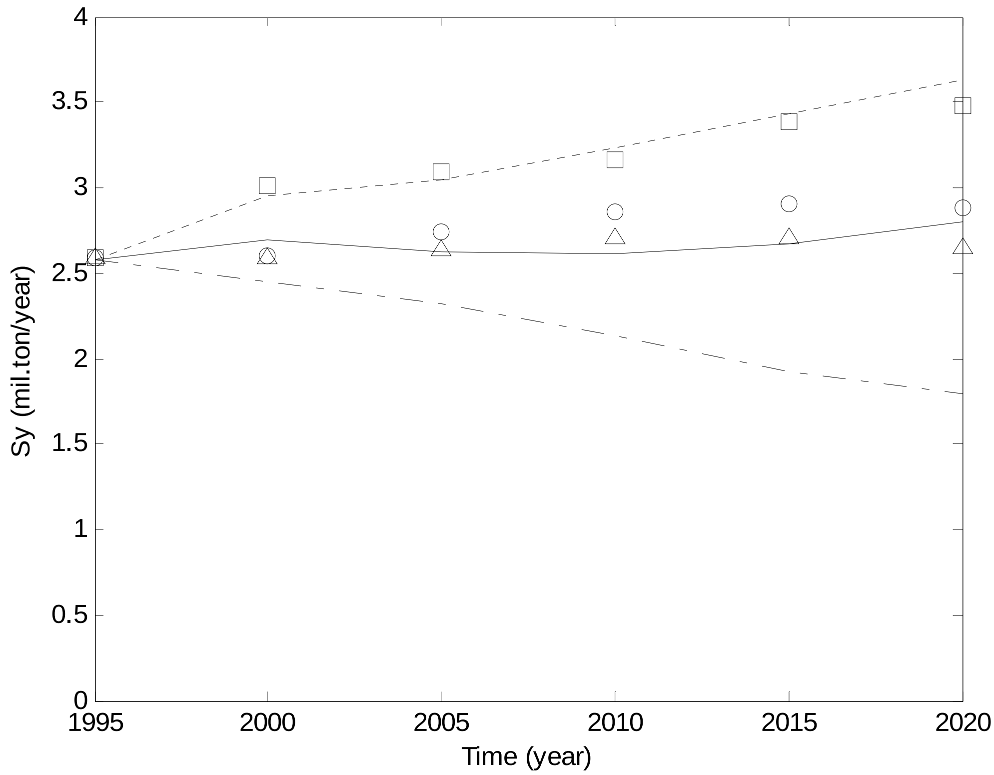

Figure 4.3. Comparisons between the sediment yields computed by RaMCo (with adjusted α) under three scenarios A-guided market economy (o), B-maximum growth (□) and C-sustainable development (Δ) and the sediment yields computed by the hypothesised system for scenarios A (solid line), B (dotted line) and C (dash-dotted line).

It can be seen from Fig. 4.3 that even with the same quantitative changes in the three land-use types, produced by the two systems, the RaMCo model is still unable to produce the trend line which was produced by the experts' system, in scenario C. It can be concluded that the too coarse aggregation level of land-use makes the model fall short of describing the consequences of the change in socio-economic factors and policy options produced by the expert on the sediment yields.

4.7. Discussion and conclusions

In the field of integrated systems modelling, researchers often encounter the dilemma that the model should not be so complex that it is unmanageable in terms of data collection, uncertainty propagation, yet not so simple that it cannot give useful information to the decision-makers.

In this paper, a novel approach to testing integrated systems models using qualitative scenarios has been presented. The approach is established to determine whether a model is ill or well designed, with regard to the purpose of an IWS model as a tool capable of reflecting the system experts' consensus about the dynamic behaviours of system output variables, under a set of possible socio-economic scenarios and policy options. The design of this approach was motivated by the three reasons that limit the relevance of the conventional approaches to the validation of ISM: the limited predictive ability of historical data to describe the future behaviour of interactive natural-human systems, the qualitative nature of the social sciences and the scarcity of field data for validation.

The application of the new approach to validation of the RaMCo model suggests several interesting points about the validity of the RaMCo model in particular and about IWS models in general. First, the RaMCo model is able to describe the dynamic developments of the three aggregated land-use types under three scenarios if the growth coefficient α in Eq.(4.1) is adjusted accordingly to each scenario. The argument for this adjustment is that it should be dependent on additional policy variables such as: environmental law, agricultural intensification programmes and cheap credits which have not been explicitly included in Eq.(4.1). Second, without a refinement of land-use types the RaMCo model fails to produce satisfactory consequences of policy options on sediment yields. The lesson learned is that a model can be valid for one purpose but invalid for another. Therefore, the validity of any IWS model should be assessed in accordance with a clearly specified management variable.

One important aspect related to the proposed approach is the consideration of the uncertainties related to the system output projections made by the fuzzy rule-based model (i.e. experts' system) and the RaMCo model (i.e. model system). The very sound question is 'Are the differences in the projections produced by each system model under the three scenarios overwhelmed by the uncertainty of internal factors (parameters) and external factors (stimuli)?' To answer this question, the propagation of the uncertainty of the system inputs (e.g. the costs and prices of rice culture) and of the system parameters (e.g. soil erodability factor and cover management factor) to the system output (e.g. sediment yield) can be carried out. This is a very useful test to determine how informative the policy recommendations are to the decision-makers, given the model structure and the numerical uncertainty of the model's inputs and parameters. The width of the uncertainty bounds around each future output projection of a system model reflects the ability of the modellers or analysts to predict or describe the future inputs and the ability to determine the numerical values of the model parameters. For the current approach, we consider and compare the systematic and casual changes in system output behaviour, which resulted from the changes in the future system inputs, produced by the two different structural systems (model system and experts' system). Under the same set of realistic (but uncertain) parameters and system inputs, both systems should produce similar trends (signs and magnitudes) of the system output under each scenario. The uncertainty propagation mentioned above, together with an evaluation of the accuracy of the model behaviour, given that empirical data are available, may be implemented after a certain level of confidence in the model structure is gained. These two topics are addressed in Chapters 3 and 5 of the thesis for other case examples. After all, our approach is inspired by the realisation that the systematic input-output changes is equally useful information to the decision-makers, besides the typical numerical uncertainty bounds around the prediction made by an integrated systems model.

The advantage of the proposed approach is that it opens a new direction for the validation of IWS models using qualitative hypotheses formulated by system experts on future trends for which conventional techniques fail. It makes the assumptions and reasoning processes that lead to expert's judgments more transparent to modellers. Thus, not only can the final quality of the model be assessed but also possible structural errors can be unravelled. It helps to reduce the possible bias of the experts through the process of documentation and communication between modellers, system experts and stakeholders.

However, some limitations of the new approach should also be mentioned. The first practical difficulty is that it is difficult to find system experts who are knowledgeable about both field and scientific research. In this application, to counter the expert's bias, a workshop with the participation of local scientist experts, resources managers and stakeholders, was held at the step of testing the consistency of the scenarios. In the situations where multiple system experts (scientist experts and resources managers) and stakeholders (other than resource managers) are available at earlier stages (e.g. structuring scenarios), several techniques could be implemented, which may facilitate the process of identifying key drivers and the driving mechanism (reflected by the inference rules) underlying the system studied. Elicitation techniques such as Analytical Hierarchy Process (Zio, E., 1996), Adaptive Conjoint Analysis (Van ders Fels-Klerx et al., 2000) and Simple or Weighted Average technique (Chapter 3) can be applied to elicit expert's and stakeholders' opinions on key system drivers. A data mining technique applied to establish the inference rules from a multiple expert's opinions was presented by Kawano et al. (2005). The second difficulty is that the estimations of the quantitative ranges of the inputs and outputs in the hypothesised system are difficult when data are lacking and surrounded with uncertainty.

From a philosophical perspective, the current approach acknowledges that the process of communicating, persuading and convincing groups of modellers, experts and end-users plays a vital role in the process of validating IWS models (Pahl-Wostl, 2002; Poch et al., 2004). The complexity of the environmental problems makes necessary the development and application of new tools capable of processing not only the numerical aspects, but also the experience of experts and wide public participation, which are all needed in the decision-making process. In parallel to this development, the use of the historical data and comparing them with model outputs (empirical test) is of vital importance. This comparison should be included when possible. However, new methods, focusing, for example, on the trend comparison might be promising and so need to be further developed for the validation of IWS models. Our new approach to testing IWS models may be useful in both situations where measured data are unavailable and where data are available for the empirical test.

Chapter 5

Validation of a fisheries model for coastal zone management in Spermonde Archipelago using observed data

Abstract: RaMCo (Rapid Assessment Model for Coastal Zone Management) consists of different process-based models and the linkages between them. The equilibrium Fox (1970) model is used in RaMCo to assess the fish stock and to predict catches from the anticipated effort development. In this chapter, validation of this model is presented with the availability of seven-year historical data on catch and effort. Residual analysis is used to test the pattern replication ability of the Fox model as well as the two alternative fisheries models (the equilibrium Schaefer model and the non-equilibrium Walters and Hilborn model). The Mitchell (1997) test is included to examine the predictive accuracy of these three models. The extreme condition test is also conducted to test biological plausibility. It can be concluded from the results that the validity of any biological model should not be judged solely on the basis of predictive accuracy but also on its pattern replication ability and biological plausibility. In this respect, the Fox model solved under equilibrium assumption gives more accurate results on the predicted values of catches, given the available measured efforts, comparing with its two alternatives. However, the two equilibrium production models fail to be useful tools as long-term predictors of future fisheries scenarios because they violate the biological stocks-and-flows principle.

5.1. Introduction

In RaMCo, the equilibrium Fox (1970) model (referred to as Fox model from now on) is used to assess the fish stock abundance from annual statistical data on effort and catch. The Fox model predicts the values of the Maximum Sustainable Yield (MSY) and the corresponding sustainable effort. With this information, decision-makers can determine at what exploitation rate the optimum yield is achieved while avoiding depletion of the fisheries stock (Mace, P.M., 2001). Besides, the anticipated fishing effort for the coming twenty-five years can be derived from predefined scenarios. The anticipated time series of the fishing effort is used as input for the Fox model to compute the catch. From that, the regional production of ocean fisheries, income and protein supply from ocean fisheries can be obtained.

In validating RaMCo, a framework for the validation of Integrated Systems Models has been established (Chapter 2). A new approach using qualitative scenarios was derived to validate the subsystem of land-use change and the resulting sediment yield (Chapter 4). The bio-economic fisheries model can be validated in a similar way, but here the fishing effort is substituted by land-use change and the resulting sediment yield is substituted by the catch or catch per unit effort (CPUE). In the example in Chapter 4, the expert found the results supplied by the USLE model satisfactory. Otherwise, it is still required to make sure that the soil loss model is correct, thus leaving room for the validation of the process-based model. For the Fox model, the nineteen-year statistical data (1977-1995) on effort and catch were used to calibrate the model. At that time, the district catch data in the study area were not available. The data of the regional catch were computed by adopting the values of CPUE(t) for the province. The availability of more recent seven-year statistical data (1996-2002) and new district data for the catch (1977-2002) enables us to recalibrate and empirically validate that model.

The purpose of this chapter is threefold. First, it searches for an appropriate method to validate a process-based model embedded in an Integrated Systems Model if measured data are available. Second, this method is applied to test the validity of the Fox model, considering it as the core component of a bio-economic fisheries model in RaMCo. Last, with the examinations of the two alternatives of the Fox model, recommendations for model improvement can be proposed.

This chapter is organised as follows. Section 5.2 is devoted to a description of the case study, which includes: fisheries in the study area, the bio-economic fisheries model in RaMCo and the processing of the statistical fishery data. Section 5.3 consists of a literature review of validation techniques which use observed data and the techniques adopted to assess the usefulness of the Fox model. Mathematical descriptions of the Fox model as well as the two alternative fisheries models are given at the end of this section. Section 5.4 presents the results obtained from the calibration and validation of the three models. The strong and weak aspects of the Fox model and the two alternatives, with respect to intended purposes, are discussed in Section 5.5. The recommendations for model improvement and for future research are also included in Section 5.5.

5.2. Case study

5.2.1. Fisheries in the Spermonde Archipelago, Southwest Sulawesi

The Spermonde Archipelago is an island group in the Makassar Strait west of Sulawesi. The islands lie to the west and northwest of Makassar (formally Ujung Pandang), the capital of the Indonesian province of Sulawesi Selatan (Renema and Troelstra, 2001). Administratively, the Archipelago is comprised of four districts: Takalar, Makassar, Maros and Pangkep. The coastal water covers about 400,000 ha with submerged coral reefs, coralline islands, sandy shallows and deeper waters up to a maximum depth of 60 m. The four districts mentioned had around 6500 fishing households that operated about 6700 boats (1995). The total effort in this shallow area was about 1.9 million trips/year, which resulted in 52,572 tons of fish and a total value of approximately $22.5 million. Small pelagic fish is the most important fish group in the study area, which contributed 59% to the total catch and counted for 68% of the total gear types. The prominent fishing boat is the non-powered canoe (75%), 20% were motorised boats with outboard engines, and the remaining 5% were motorised boats with inboard engines (Pet-Soede et al, 1999).

5.2.2. Fisheries modelling in RaMCo

Fishing effort *E* (trip/year), catch *C* (tonne/year), and the relative stock biomass CPUE form the key variables for the fisheries subsystem model (Fig. 5.1). Based on gear type and fish species, the model distinguishes four categories of fisheries: demersal (A), reef fish (B), small pelagics (C) and large pelagics (D). A portion of large pelagic fish is harvested outside while the other three are caught inside Spermonde.

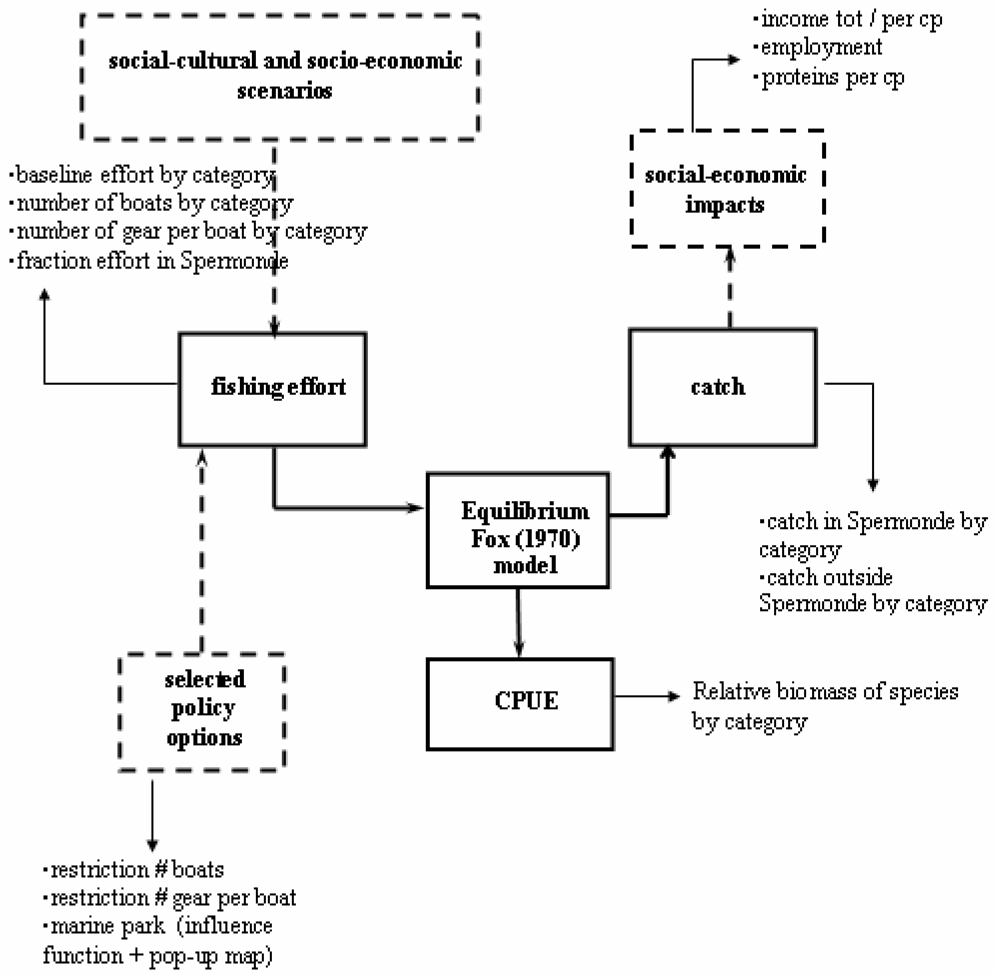

Fig. 5.1. Structure of the bio-economic fisheries model in RaMCo (after De Kok and Wind, 1999).

The model is driven by a user-defined scenario for the fishing effort *E*, which is specified by gear class, the demand based on consumption and export, and the size of the fishing community. The baseline scenario for the fishing effort corresponds to a qualitative description of the socio-cultural and socio-economic developments influencing the fishing effort (Fig. 5.1), which is selected by the user. The qualitative descriptions have been formulated by cultural and anthropological experts. Cognitive maps (Kosko, 1986) are provided in the model which describes the development of the fishing effort for the four fisheries categories. The integration of the qualitative scenarios with the quantitative simulation model using cognitive maps was described by

De Kok and Wind (1999). The user may intervene by three policy options: gear type restriction, gear number restriction, or the (permanent) installation of a marine park. A marine park (Man et al., 1995) is introduced by drawing cells on a pop-up map. The relative size of the marine park (compared to the total area of fishing grounds) determines its effectiveness, together with the location choice.

5.2.3. Data source and data processing

Indonesia's fisheries data have been collected since 1976 with a comprehensive Catch and Effort Data Recording System (CEDRS). Fisheries data for the province of South Sulawesi were obtained from 26 fisheries yearbooks (Anonymous, 1977-2002). Annual catch C (tonne /year) per fish species, effort (trip/year) per gear type and number of thirteen boat categories are available from 1977 to 2002 for each of the sixteen rural districts (Kabupaten) and two municipalities (Kotamadya) in the province.

Table 5.1. Categorisation of the four main fish categories by fish species and gear types (De Kok and Wind, 1999)

Fisheries Category	Fish species	Gear types
Demersal	Indian Halibut, flat fish, bombay duck, slip mouth, sea catfish, lizard fish, goat fish, giant seaperch, threadfin bream, big eye, yellow tail, croaker, shark, ray, black pomfret, silver pomfret, mullets, threadfin, hardtail cutlass fish, others	Other pole & line, guiding barrier, shrimp gillnet, beach seine, stownet, other, Danish seine, other liftnet, scoopnet, seaweed collector, trammelnet.
Reef fish	Sweetlip, red snapper, grouper, emperor	Other traps, shellfish collector, muro ami
Small pelagics	Scad, hardtail scad, flying fish, anchovies, rainbow sardine, fringescale sardinella, indian oil sardinella, wolf herring, toli scad, indian mackerel	Set bagan, payang, boat bagan, encircling gillnet, portable traps, purse seine
Large pelagics	Barracuda, trevally, jack, queenfish, rainbow runner, indo-pacific king mackerel, narrow barred king mackerel, tuna, skipjack tuna, eastern little tuna	Drift gillnet, set gillnet, trolline, set longline, other drift longline, skipjack pole & line, tune longline

As mentioned earlier in this section, RaMCo differentiates four main fish categories: demersal, reef fisheries, small pelagics and large pelagics. Therefore, the accumulation of statistical data on catch and effort for each main fish category in the study area is needed. The annual district fishing effort for each main category is accumulated from district effort data of 27 fishing gear types. For the annual district catch, catch data of 45 recognised fish species is accumulated. The selection of fish species and gear types, which belong to each of four main categories, is depicted in Table 5.1. These annual statistical data on catch and effort in the four districts mentioned are aggregated to obtain the annual data on catch and effort in the Spermonde Archipelago. Thirteen boat categories are accumulated into three main groups: nonpower, outboard engine and inboard engine. These three main categories of fishing boats are used to compute the correction factor for effort as described below.

Choosing an appropriate unit for fishing effort is of paramount importance in fish stock assessment which uses catch and effort statistical data. The use of fishing trips landed per gear type in Indonesian fishery statistics were criticized (Pet-Soede, 2000) because it does not take into account the effect of innovation of recent fishing techniques on the catch. Attempts to search for the right effort unit have recently been reported in the literature (Pascoe and Robinson, 1996; Moses et al., 2002; Van Oostenbrugge et al., 2002). In RaMCo, fleet motorisation is accounted for by using the following equation (private communication with Van Densen):

$$E(t) = E(t)_{uncorrected} \cdot \frac{N(t)_{nonpow} + 3N(t)_{outboard} + 5N(t)_{inboard}}{N(t)_{nonpow} + N(t)_{outboard} + N(t)_{inboard}} \qquad (5.1)$$

where $N(t)_{nonpow}$, $N(t)_{outboard}$ and $N(t)_{inboard}$ are numbers of nonpower units, outboard-engine units and inboard-engine units at year t, respectively; $E(t)_{uncorrected}$ is the number of fishing trips obtained from statistical data; $E(t)$ is the total number of trips at year t measured as the number of trips belonging to a “reference” fleet which has no engine.

5.3. Validation methodology

5.3.1. State of the art

For the last three decades, a controversial debate has been taking place on the validation and evaluation of physically-based process models (Konikow and Bredehoeft, 1992; Oreskes et al., 1994), bio-ecologically based process models (Rykiel, 1996), and integrated system dynamics models (Forrester and Senge, 1980; Parker et al. 2002). This lengthy debate has generated not only insight, but sometimes confusion for practitioners (e.g. semantic problems). More researchers, nowadays, are accustomed to the view that “all conceptual models are wrong” (Sterman, 2002; Refsgaard and Henriksen, 2004). It is widely accepted that any model is a tool designed for a specified purpose rather than a truth generator. Therefore, the validity of a model should be judged in the light of its usefulness (Nguyen et al., 2005). This does not mean that the requirement of a good model is less strict than it was before. Inventories of methodologies and techniques for model validation have been made (Shannon, 1981; Kleijnen, 1995). The following review focuses on the technical aspects of validation tests developed in the last decade.

Model validation tests can be divided into two groups (Nguyen and de Kok, 2003): data-driven tests (empirical tests) and expert-knowledge-based tests (rational tests). Empirical tests can be categorised, based on the three technical approaches: Goodness-Of-Fit (GOF) approach, hypothesis testing approach, and pattern testing (periodicity, trend) approach. The graphical approach to validation mentioned in Section 2.2 is not described here because it is very qualitative and subjective. The approach using confidence intervals is also excluded since it is similar to the hypothesis testing approach. In the first approach, a GOF statistic (based on some function of observed vs. predicted), such as root mean square and Nash-Sutchliffe coefficients (Nash and Sutcliffe, 1970), is used as the quality measure of a model. The advantage of this approach is that it gives a quantitative answer which can be used in comparing the alternative models. The disadvantage of using GOF statistics for model validation is the difficulty in finding an objective benchmark to distinguish between a useful and a useless model and the lack of diagnostic power. Besides, the GOF statistics has been argued to be inappropriate for the validation of ecological models (Loehle, 1997). Loehle demonstrates that a GOF statistic is not an adequate measure for an ecological model, due to the autocorrelation of the predicted variable over time, the sensitivity of state variables to the initial condition and the unequal importance of each point in predicted time series.

In the hypothesis testing approach, a popular technique was to use linear regression between predicted and observed values. Statistical tests such as the F-test and the t-test (Draper and Smith, 1981) were used to test the simultaneous null hypotheses that the regression line has an intercept of zero and a slope of one. However, the use of the F-test and regression analysis between observed and predicted for validation purpose is proven to be erroneous by Thornton (1996), Kleijnen et al. (1996) and Mitchell (1997). Kleijnen et al. (1996) propose an alternative hypothesis testing which tests the null hypothesis of zero intercept and zero slope of the regression line between the residual (observed minus predicted) and the sum (observed plus predicted). Mitchel (1997), in rejecting the use of regression for model validation, proposes the use of a graphical method and the residual (predicted minus observed) for the purpose of evaluating the model's adequacy. Criteria for model adequacy are defined as the width of the envelope of acceptable precision and the proportion of residual points lying within it. Both the width of the envelope of precision and the proportion of points should be chosen based on the purpose of the model and the precision of observations. The advantage of this method, in comparison to other hypothesis tests, is that it relaxes the assumption of independence of the predicted variable and the normality of the residuals. Similar approaches are set up (Scholten and Van de Tol, 1994; Ewen and Parkin, 1996; Scholten et al., 1998), adopting the concept of confidence bounds or confidence limits (both are analogous to the envelope of precision mentioned by Mitchell). In deprecating the use of GOF against time series data for the validation of dynamic ecological models, Loehle (1997) proposes a test statistic T, which has the same meaning as the proportion of the points lying within the confidence bounds mentioned previously. Similar to the approach of testing system dynamics models (Forrester and Senge, 1980), Loehle suggests using statistic T under biological and ecological realisms as well as the extreme condition test and pattern replication test. He also proposes the use of sensitivity and uncertainty analyses, and structural analysis (i.e. comparison of different models with different structures, algorithms, and aggregation schemes) for validation.

The third approach (pattern testing) tries to compare the trend line and the periodicity predicted by the model with the corresponding ones observed in reality. It is very useful for testing dynamic models, but received less attention in the literature. Some test statistics developed for this type of test are described by Barlas (1994). It is believed that the pattern replication ability of a model can be conveniently examined with residual analysis so that discussion on these statistics is not mentioned further here. Van Tongeren (1995), in discussing the strength and weakness of regression models versus simulation models for limnological modelling, demonstrates the importance of residual analysis for the evaluation of these two model types. He concludes that both model types are shown to fail for prediction purposes, mainly because the error variance in the data used for parameter estimation is large compared to the variance that can be explained. Kirchner et al. (1996) claim the poor diagnostic power of the conventional validation methods, which use predicted versus observed plots. Validation tests are often divorced from the conditions under which a model will be used, particularly when the model is designed to forecast beyond the range of historical experience. They mention that inspection of residuals is a standard step in evaluating both statistical and simulation models. Performance criteria and benchmarks are required to be explicitly stated and documented, and the validity of a model is most meaningfully judged by explicit comparison with the available alternatives. Fraedrich and Goldberg (2000) design a framework for the validation of predictive simulations, in which variance of residual (predicted minus observed), variance due to measurement accuracy of the output, variance of predicted values due to estimation of parameters and inputs, variance due to specification errors in the conceptual model, and the accuracy requirement of a model are compared in pair-wise fashion to assess simulation models. In short, all the authors mentioned agree on the important role of residual analysis for the validation of ecological models in particular and all models in general. It is vital because it has a diagnostic power; it is not based on any assumption about the normality and independence of the interested variable. It does not falsify the valid models just because of a phase shift between predicted and observed time series. This means that the periodicity and trend can also be examined within the framework of residual analysis.

5.3.2. The proposed method

In view of the technical problems related to model validation mentioned in Subsection 5.3.1, an appropriate approach is proposed to validate a biologically process-based model (the biological fisheries model in this case example). Here model validation is defined as the process of examining the ability of a model to fulfil its designed purposes. Since the validity of a model is always relative to different models and different purposes, a good validation test should not only indicate how bad or how good the model is, but also why it is bad or good and how the model can be improved. The proposed approach to validate the biological fisheries model includes the following four steps:

- A calibration is carried out to establish the parameter values for the model’s prediction.

- A visual inspection of residuals (observed minus computed) is conducted to examine the ability of the model to replicate long-term and short-term patterns of the real system.

- An examination of the model predictive accuracy with the inclusion of the confidence bounds around the residuals (i.e. Mitchell (1997) test) is carried out. As mentioned before, these conclusions are only provisionally valid in the light of currently available validation data.

- Judgement on the biological plausibility of the model is concluded, using the extreme condition test, in view of the biological, ecological and dynamic realisms.

To identify possible improvements for RaMCo, the equilibrium Schaefer model and the non-equilibrium Walters and Hilborn model (Hilborn and Walters, 1992) were also examined. These two models were chosen since they belong to the same group of fisheries surplus production models, which require data only on catch and effort.

The application of these four steps to the validation of the Fox model and of its two alternatives is described in Section 5.4. The following are summarised descriptions of the three models to be tested.

5.3.3. Fishery production models

Fox (1970) model

The Fox (1970) model is based on a continuous time-differential equation:

$$\frac{dB(t)}{dt} = rB(t)\ln\left(\frac{k}{B(t)}\right) - C(t) \tag{5.2}$$

where *B(t)* and *C(t)* are the biomass and catch rate at time *t*, respectively; *r* is an intrinsic rate of population growth; *k* is a parameter corresponding to the unfished equilibrium stock size. The catch rate *C(t)* is assumed to be proportional to biomass and effort, following the equation:

$$C(t) = qE(t)B(t) \tag{5.3}$$

In equation 5.3, *q* is a parameter to describe the effectiveness of each unit of effort and *E(t)* is fishing effort at time *t*. Under the condition of equilibrium biomass (rate of change of the biomass equals zero), combining Eqs (5.2) and (5.3), gives:

$$\ln(U_e(t)) = a + bE(t) \tag{5.4}$$

Or equivalently:

$$C_e(t) = E(t)\exp[a + bE(t)] \tag{5.5}$$

where, $U_e(t) = C_e(t)/E(t)$, is the catch per unit effort (CPUE) at time *t* under equilibrium biomass assumption; the two coefficients $a = \ln(kq)$ and $b = -q/r$.

Taking the first derivative of Eq (5.5) results in the effort corresponding to Maximum Sustainable Yield (E_{MSY}): E_{MSY} = -1/b and MSY = -[exp(a-1)]/b. The *a* and *b* coefficients can be calculated by regressing Eq. (5.4) with the annual statistical data on catch and effort, assuming that the catch rate at equilibrium $C_e(t)$ equals the observed catch rate *C(t)*.

Schaefer (1954) model

The Schaefer (1954) model has the following form:

$$\frac{dB(t)}{dt} = rB(t)\left(1 - \frac{B(t)}{k}\right) - C(t) \tag{5.6}$$

Under the same assumption as made for solving Eq. (5.2), one obtains:

$$\frac{C_e(t)}{E(t)} = U_e(t) = a + bE(t) \tag{5.7}$$

where $a = qk$ and $b = - q^2k/g$. Taking the derivative of Eq. 5.7, one can get the $E_{MSY} = -a/2b$ and $MSY = - a^2/4b$. The *r, k, B(t), $C_e(t)$, $U_e(t)$* have the identical meanings as those explained previously in the Fox model.

Walters and Hilborn (1976) model

The difference equation model of Walters and Hilborn (Hilborn and Walters, 1992) is the integral form of the Schaefer (1954) equation over the time step of one year:

$$B(t) - B(t-1) = rB(t-1)\left(1 - \frac{B(t-1)}{k}\right) - C(t-1) \tag{5.8}$$

Substituting Eq. (5.3) into Eq. (5.8) gives:

$$\frac{U(t) - U(t-1)}{U(t-1)} = r - \frac{r}{qk}U(t-1) - qE(t-1) \tag{5.9}$$

Or equivalently:
$$\frac{U(t) - U(t-1)}{U(t-1)} = a + bU(t-1) + cE(t-1) \tag{5.10}$$

By regressing Eq. (5.10), the regression coefficients *a*, *b*, and *c* can be estimated. The MSY and the corresponding effort can be obtained as: $E_{MSY} = -a/2c$ and $MSY = a^2/4bc$.

5.4. Results

5.4.1. Calibration

Table 5.2 and Fig. 5.2 present the results of regressions (calibrations) of the three mentioned models using the observed catch and effort data from 1977 to 1995. It is noted that the R^2 and the confidence limit estimates of the regression coefficients in the Fox model result from the natural logarithm transformation of the dependent variable (CPUE). Thus, the goodness of fit should be checked by examining the residuals rather than looking at the values of R^2 themselves. Furthermore, good fitting does not necessarily imply good forecasting (Stergiou and Christou, 1996). The predictive ability of the three models will be examined by inspecting the residuals between the predictions and the validation data (1996-2002). Fig. 5.2.b shows that the equilibrium Schaefer model has a better fit to calibrated data, comparing with the Walters and Hilborn (W&H) model, but turns out to be a very poor predictor of small pelagics.

One purpose of the Fox model is to predict two important variables of interest for the fishery managers (MSY and E_{MSY}). The confidence limits of the regression coefficients can give an indication of the magnitude of the confidence limits for these two management variables. Because of the transformation problem mentioned before, the confidence limits of the regression coefficients in the Fox model can be inferred from its two alternatives if they have a similar goodness of fit. Table 5.2 shows that confidence limits are approximately equal to the estimates of the corresponding parameters in most cases in both the Schaefer and W&H models. It can be expected that the uncertainty bounds around the predicted regression coefficient values and around the predicted values of MSY and E_{MSY} also are in the same order of magnitude. This makes the predicted values of these two variables less informative to decision-makers. Plots of catches against efforts using observed time series (not shown here) reveal that the available data do not allow any positive judgments on the accuracy of the predicted values of these two management variables. These plots falsify the two values predicted by the Fox model for demersal fish, the two values predicted by all three models for large pelagics and these two predicted by the Schaefer and W&H models for reef fish.

Table 5.2. Results of regression and the predicted values of fishery management variables

Model	Demersal fish	Reef fish	Small pelagic fish	Large pelagic fish
Fox (1970) (equilibrium)				
a	2.476 (0.4639)	6.523 (0.6316)	4.856 (0.180)	2.270 (0.326)
b	-0.000413 (0.000378)	-0.050751(0.014257)	-0.001868(0.000419)	-0.000767(0.000289)
R^2 (%)	23.8	76.8	83.9	64.8
E_{MSY} (10^3 trip)	2421.3	19.7	535.3	1303.8
MSY (tonne/yr)	10594	4934	25308	4643
Schaefer (1954) (equilibrium)				
a	10.940 (3.259)	1045.76 (639.11)	105.329 (13.284)	8.571 (1.550)
b	-0.00301(0.00265)	-15.336 (14.426)	-0.1083 (0.0309)	-0.00387 (0.00137)
R^2	25.2	22.8	76.3	67.5
E_{MSY} (10^3 trip)	1817.3	34.1	486.3	1107.4
MSY (tonne/yr)	9940	17828	25610	4746
Walters & Hilborn (1976)				
a	2.187 (0.990)	1.966 (2.12)	1.827 (1.696)	1.281 (1.129)
b	-0.182 (0.078)	-0.000734 (0.001301)	-0.02168 (0.01564)	-0.15749 (0.12425)
c	-0.000699 (0.000469)	-0.016709 (0.041769)	-0.001237(0.001938)	-0.000547(0.000585)
R^2 (%)	60.8	24.1	48.0	31.3
E_{MSY} (10^3 trip)	1564.4	58.8	738.5	1170.9
MSY (tonne/yr)	9399	78788	31116	4762

Notation: a, b, c are regression coefficients; the numbers in brackets are the 95% confident limits for the estimates of the regression coefficients next to them; R^2 is the percentage of variance explained by regression; MSY and E_{MSY} are the maximum sustainable yield and the corresponding effort.

5.4.2. The pattern test

The capability of the three models to predict catches from anticipated effort are examined by an inspection of residuals and the Mitchel (1997) test. For the purpose of methodological demonstration, only two fish categories are examined. The small pelagics and demersal are selected because the former is the most economically important in the study area and the latter is claimed by fishermen to be decreasing in fish stock size (Pet-Soede, 2000). The large pelagic fish category is excluded because it is only in the immature stage of exploitation. The obtained data do not allow an understanding of the stock development (Hilborn and Walters, 1992). Reef fish is also excluded because of the poor descriptive characteristic of the data (Pet-Soede et al., 1999).

Following Draper and Smith (1981), the time sequence plots of the residuals (observed minus predicted) of CPUE for the two fisheries categories are presented in Fig. 5.2. A distinction between the results, obtained from calibration data (1977-1995) and from validation data (1996-2002), is indicated by using different symbols in each figure. For time sequence plot inspection, it is important to differentiate between the long-term and short-term time trends. When validation data cover a limited time period, calibration data can be included to examine the long-term time trend. With the Fox model, the horizontal bands of residuals obtained (considering the whole time series from 1977-2002) for both fisheries categories indicate an adequacy of the long-term pattern replication ability of the model. However, the model somewhat under-predicts small pelagics since the residuals (stars) are all positive. An example of violating the horizontal band requirement is demonstrated by the equilibrium Schaefer model for small pelagics (Fig. 5.2.b). The residuals obtained during validation (stars) deviate from the horizontal band, indicating that the model is unable to replicate the long-term time trend of the observed data. It may be caused by the lack of a linear term in time in the model or by the effect of an independent variable subject to a time effect. This can be made clear in the next examination of the residuals (Fig. 5.3). Considering the short-term time trend (only data during the validation period is considered), the equilibrium Fox model unsatisfactorily replicates for demersal and successfully replicates for small pelagics. For the non-equilibrium W&H model, similar to the conclusions made for the Fox model, it is able to reserve the long-term time trends for both demersal and small pelagics. The comparison of predictive accuracies between the non-equilibrium W&H and the equilibrium Fox model will be mentioned later. Pattern test statistics or hypothesis tests, such as an F-test to test the null-hypothesis of the zero slope of the regression line between residuals and the time during the whole period of 1977-2002 (does not require a zero intercept for pattern test), can be used at this stage. However, they are not included since a detailed examination of the corresponding residuals plot is usually far more informative (Draper and Smith, 1981), and the plots will almost certainly reveal any violations of assumptions serious enough to require corrective action.

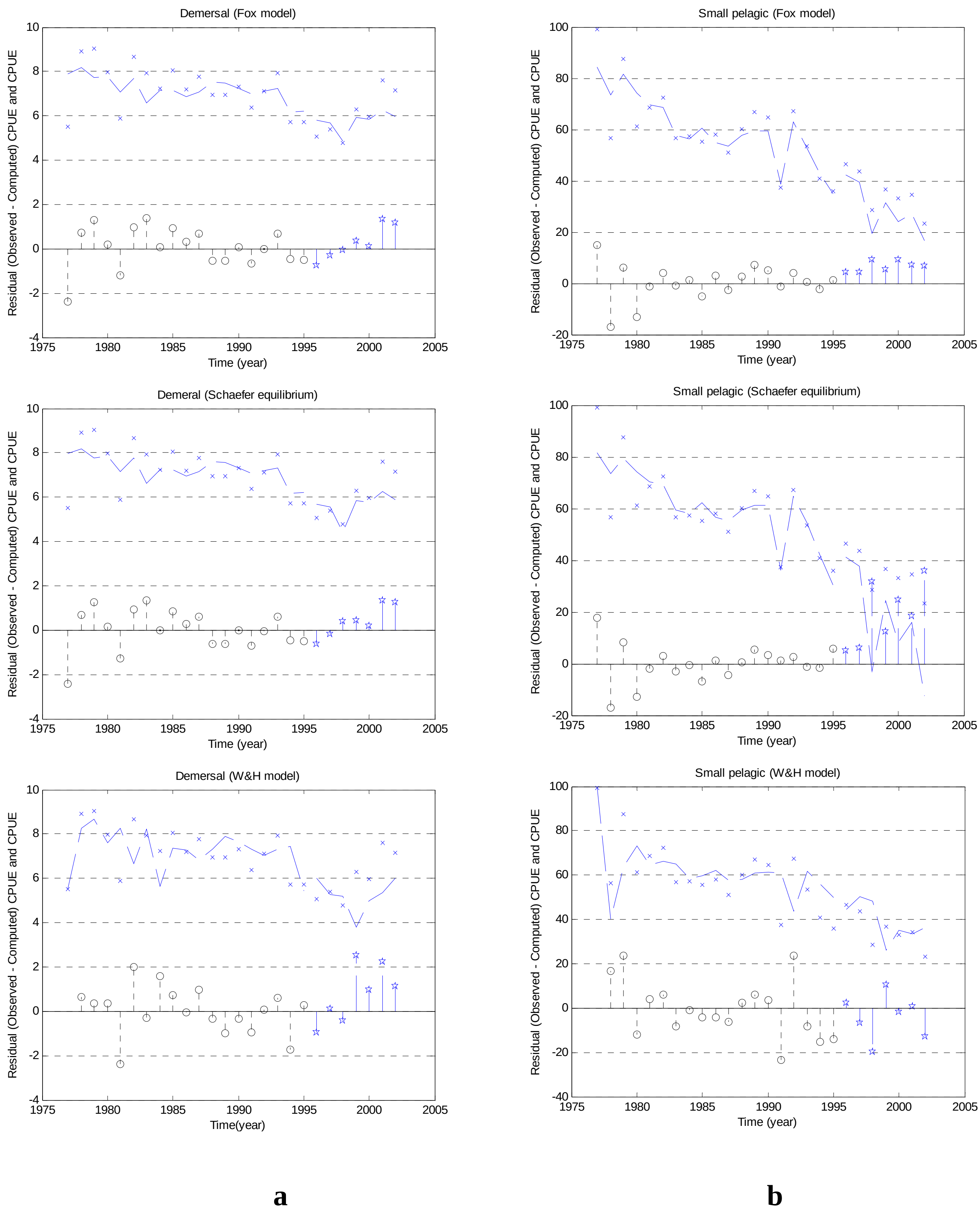

Fig. 5.2. Time sequence plot of the CPUE residuals against time, t. (x): observed data; (dash line): computed; (o): residuals during calibration (1977-1995); (☆): residuals during validation (1996-2002).

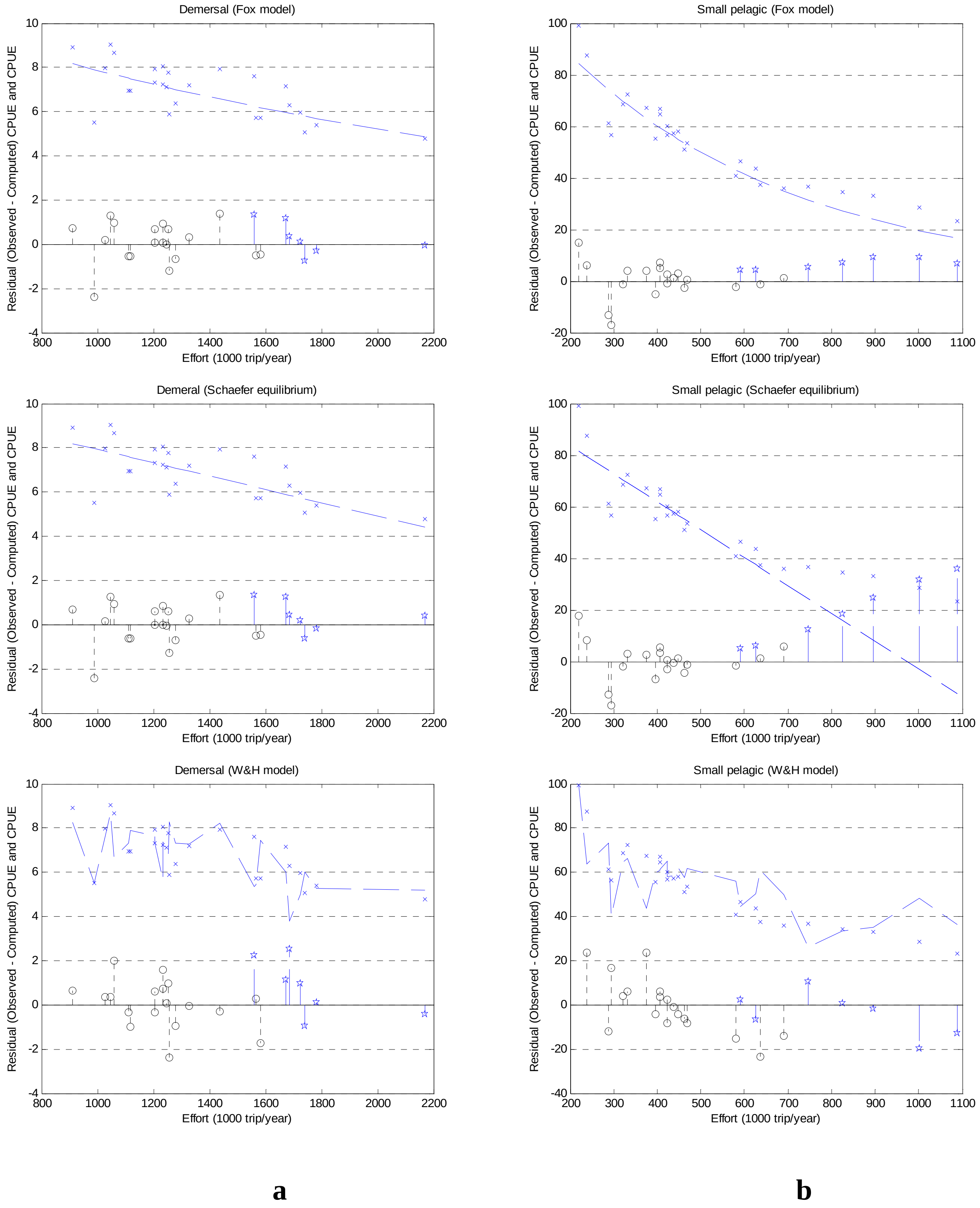

Fig. 5.3. Plot of the *CPUE (t)* residuals against independent variable, *E(t)*. (x): observed data; (dash line): computed; (o): residuals during calibration (1977-1995; (☆): residuals during validation (1996-2002).

Fig. 5.3 depicts the plot of residuals of dependent variable (CPUE) against the independent variable (E). A similar approach as described above can be followed to examine the quality of the three models. The results show that the equilibrium Fox model and the non-equilibrium Schaefer model mimic satisfactorily the pattern of the observed data. The added value to the first analysis is that the second one reveals that the equilibrium Schaefer model fails because it does not take into account the nonlinear relationship between CPUE and E for small pelagics. It explains the failure of the equilibrium Schaefer model to mimic the observed pattern, which was found by the first analysis, because of this reason. Thus, the improvement can be made by taking into account this nonlinear relationship rather than including a linear term in time.

5.4.3. The accuracy test

So far, the pattern tests via visual inspection of residuals have been discussed. The model used for predictive purpose also needed to be judged in terms of predictive accuracy. For this purpose, the concept of confidence bounds (Mitchell, 1997) is followed. However, the difficulty in finding an accuracy criterion was encountered since no observation error has been reported. For the sake of demonstration, we hypothetically assume that confidence limits of ± 2 units of CPUE for demersal and ± 20 units of CPUE for small pelagics were chosen to represent the accuracy requirement of the managers. These values were chosen to be 20 % of the maximum values of the historical data. For both fisheries categories, only the Fox model satisfies this condition. This conclusion is based on the availability of seven-year validation time series.

5.4.4. The extreme condition test

As mentioned previously, the Fox model is embedded in RaMCo to predict catches for given efforts in a 25-years time frame provided by different socio-economic scenarios. Thus, it is supposed to answer the what-if questions. For this purpose, the extreme condition test is very useful. The extreme condition test was originally proposed by Forrester and Senge (1980) to test system dynamics models, and later on was employed by Loehle (1997) to test ecological models. The purpose of this test is to determine whether a model behaves plausibly under an extreme condition.

The equilibrium Schaefer model has not been subject to this test because of its predictive inaccuracy and poor pattern replication. An extreme condition test is carried out to compare the Fox and the W&H models. For small pelagics, the Fox model predicts a value of 535,300 (trips/year) and the W&H model predicts a value of 738,500 (trips/year) as E_{MSY}. The two models are used to project CPUE from the year 2002 to the year 2020 with an extreme constant rate of exploitation corresponding to an effort of 1,477,000 (trips/year). This number corresponds to twice the value of E_{MSY} predicted by the non-equilibrium W&H model. As can be seen from Fig. 5.4, the W&H model predicts a sharp decline in CPUE for a certain period and a smaller decrease in the followed period. It can be explained by the combination of the stocks-and-flows principle in system dynamics (Sterman, 2002) and the biological logistic function of fish biomass (Eq. 5.6). For the Fox model, it keeps the constant value of CPUE as time propagates. This is an implausible behaviour which has been caused by the equilibrium biomass assumption. Thus, the Fox model is not useful with respect to its purpose as a predictor of the long-term behaviour of the system and to answer the what-if questions.

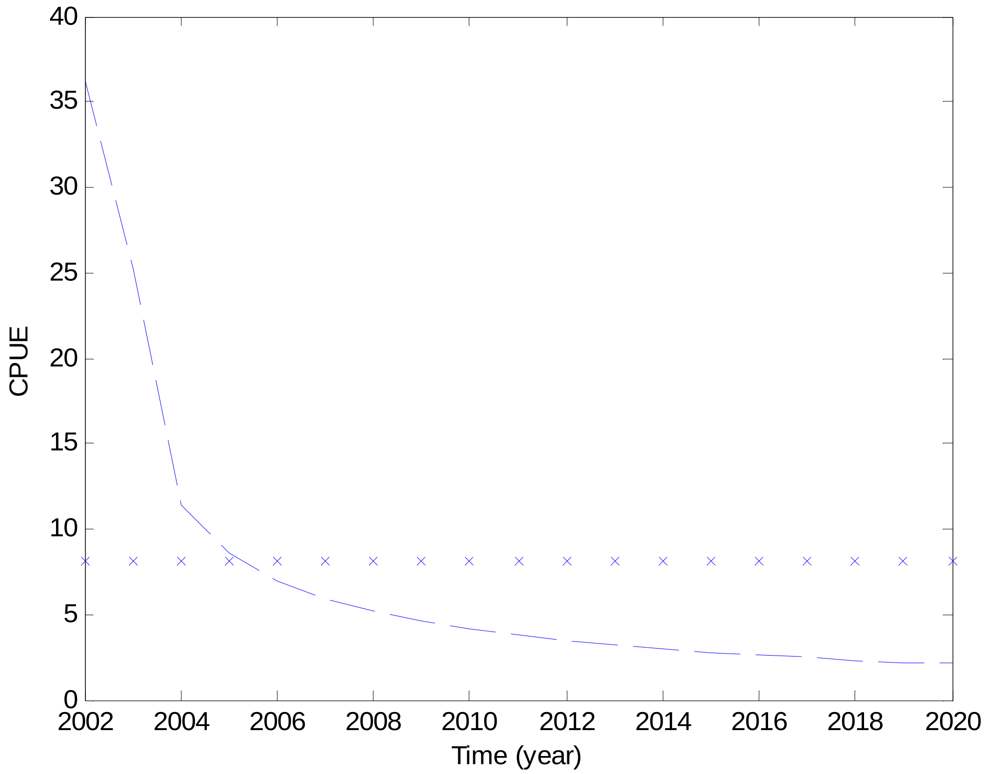

Fig. 5.4. Results from an extreme condition test for small pelagics.
(x): predicted by the Fox model; (dash line): predicted by the W&H model

5.5. Discussion and conclusions

In this chapter we present an appropriate approach to the validation of a biologically process-based (the fisheries) model embedded in a larger bio-economic fisheries model. First, the calibration is conducted to establish the parameter values of the model. Second, the residual analysis is used to test the pattern replication ability of a model. Third, the Mitchell (1997) test is adopted to test the accuracy adequacy. Last, the extreme condition test is used to test the model plausibility. The visual analysis of residuals is demonstrated to be essential for mode validation using observed data because of its diagnostic power. The results of the current exercise show that the empirical validity of a predictive model is always provisional given the available observed data. It is especially true when the model is used outside the range of calibration data. The overall model's validity should be judged with respect to its intended use. Therefore, the predictive accuracy test, the pattern replication test and the extreme condition test should altogether be taken into consideration. Besides, the information from the past (observed data for the calibration) may be taken into account when a model is validated. Although the proposed approach is designed to test a biologically process-based model, it is applicable to test a physically based process model, like the USLE used in RaMCO (Nguyen et al., 2005) as well.

The results obtained by applying this approach to the Fox model, and the two alternative models show that: i) the Fox model fails to give useful information on the management variables such as MSY and E_{MSY}, because of the wide uncertainty bounds surrounding the predicted values; ii) the same conclusion is applied to each of the two alternative models; iii) for short-term prediction, the Fox model seems to be outperformed over the two alternatives in terms of predictive accuracy; iv) given the model purpose, the quality of the data and the models available, the non-equilibrium W&H model is proposed to be used to alert the decision makers about the long-term depressing of fish stock if overexploitation occurs.

For future research, it is recommended spending more time and effort to examine the non-equilibrium Fox model and non-equilibrium Schaefer model with observation error/time-series fitting method for parameter estimation. A detailed study of the effect of technological innovation to standardise the effort should be conducted for the study area. Finally, the poor contrast of the present data hinders the effectiveness of any model, it is recommended to update the results as soon as new data are obtained.

Chapter 6

Discussions, Conclusions and Recommendations

6.1. Introduction

Systems approach and integrated approach to the planning and management of natural resources and the environment are currently perceived as promising approaches to achieve the sustainable development of a region, of a country, and of our common world. Consequently, an increasing number of Integrated Systems Models (ISMs) have been developed, which have different focal purposes. However, the scarcity of data for model development and validation, human behaviour involvements and other factors create a number of critical questions such as: to what extent can such models contribute to our knowledge and ability to manage our environment? Do they have an added value in comparison with conventional process models? Centred in these questions are the two underlying questions: How can the validity of an ISM be defined? How to determine and establish this validity? This chapter focuses on the discussion and conclusion on the most important findings toward addressing the last two questions.

The remainder of this chapter is organised as follows. Section 6.2 is dedicated to a discussion about the innovative aspects of the research as well as the generality of the validation methodology. Also included in this section is a discussion about the limitations of the validation methodology proposed. Section 6.3 gives conclusive answers to all the major research questions formulated in Chapter 1. The last section, section 6.4, gives recommendations for future research on validation of Integrated Systems Models and the proper use of these models for practical applications.

6.2. Discussions

6.2.1. Innovative aspects

- A systematic methodology to validate Integrated Systems Models has been established (Chapter 2), which has the following new aspects:

- An appropriate testing procedure has been established to test the completeness and soundness of the model structure and its elements, based on sensitivity and uncertainty analyses (Chapter 3).

- A new approach (Chapter 4) has been derived to test integrated models using qualitative scenarios, dealing with the problem of uncertain future conditions and the qualitative nature of social science.

- An appropriate testing procedure has been formulated to test the validity of the quantitative model behaviour when measured data are partially available (Chapter 5). This consists of the three tests: pattern test using visual inspection of residuals (Draper and Smith, 1981), accuracy test using Mitchell (1997), plausibility test using extreme condition test (Forrester and Senge, 1980). They are considered to be necessary and appropriate to validate ISMs when data are available.

6.2.2. Generic applicability of the methodology

The limited time does not allow for the application of this validation methodology to another case study. However, the proposed methodology for validation is supposed to be widely applicable to other integrated system models which have the same characteristics as RaMCo. These characteristics include: the model complexness, the inclusion of social science, the lack and large uncertainty of field data for model development and validation, and the high level of aggregation. The proposed methodology is not only applicable to the validation of ISMs but it can also contribute important aspects to the *quality insurance guidelines* for integrated modelling practices. The two approaches proposed in Chapter 3 and Chapter 4, which allow the participation of experts and stakeholders in the model building process, can be used to select and refine the conceptual model as well as to calibrate the model code (i.e. the numerical computer program) in a participatory manner. The quantitative testing procedure, which is described in Chapter 5, can be applied to validate other types of models such as ecological models, hydrological models and hydrodynamic models. Nevertheless, this thesis is not expected to give a full guideline with a complete set of tests. We acknowledge that it does not matter how many tests a model has passed, the very next test and/or next data can falsify the validity of a certain model. The successful tests proposed in this thesis should be considered as the minimum or necessary (instead of sufficient) conditions prior to any application of Integrated System Models.

6.2.3. Limitations

Although several innovative aspects have been incorporated into a methodology established for the validation of Integrated System Models the limitations of the methodology and its application to the case study should be mentioned.

- In Chapter 3, a simple average method was adopted to combine the opinions of experts and stakeholders. Although the use of multidisciplinary experts and stakeholders can reduce the disciplinary bias (i.e. only the issue related to his/her disciplinary domain is important), the political bias, institutional bias and inconsistency of experts were not dealt with. Suggestions were made to adopt the Adaptive Conjoint Analysis (Van der Fels-Klerx et al, 2000) and the Analytical Hierarchy Process technique (Zio, 1996). However, these techniques have not been applied yet due to some practical reasons, such as time constraint and the willingness of the participants.

- In Chapter 4, even though we have tested the consistency of the scenarios developed by the expert by using physical constraint and the workshop with local experts, the testing procedure might be more rigorous if there had been more experts, decision-makers and stakeholders involved at the earlier stages (i.e. specifying the driving factors, creating the qualitative scenarios and formulating inference rules).

- In Chapter 5, the accuracy testing of the fisheries model adopted the value of 20% of maximum observed data values to establish the confidence bounds. This value (20%) was based on neither a measurement error analysis nor an analysis of the model purpose, but a hypothetical example of the authors. Therefore, a recommendation for the values of these confidence bounds has not been made in the thesis.

6.3. Conclusions

6.3.1. Concept definition

This section is about the answers to the first main research question:

Question 1: *How can the validity and validation of an ISM be defined?*

Since a model is an abstract simplification of a real system, which is designed for some specified purposes, the validity of any model should be judged with respect to these purposes. Therefore, we start with an inventory of the common purposes of ISMs. In light of the validation problems (Chapter 1) and purposes of an ISM, the definitions of the validity, validation and validity criteria of this model are presented.

Question 1.1: *What are the purposes of Integrated System Models?*

The literature and our own experiences (Chapter 2) indicate the following main functions of an ISM: 1) Database and library function; 2) Educational function; 3) Research prioritising function; 4) Scenario building and discovering our ignorance function; 5) Communication and discussion function; 6) Decision support function (answering the what-if questions).

Validation of an ISM is always important, but is essential with respect to the last four purposes.

Questions 1.2: *What are the appropriate definitions of the validity, validation and validation criteria of an Integrated System Model with respect to these purposes?*

In view of the above-mentioned purposes of ISMs, the validity of an integrated system model should comprise four aspects (Chapter 2): the *soundness and completeness of the model structure, the plausibility and correctness of the model behaviour.* The *soundness* is understood to be based on valid reasoning and free from logical flaws. Its *completeness* means that the models should include all elements relevant to defined problems and their causal relationships which concern the stakeholders. *Plausibility* of the model behaviour means that the behaviour should not contradict to general scientific laws, well-founded scientific knowledge and practical knowledge concerning the system studied. The *correctness* is understood as the extent to which the modelled behaviour and the measured behaviour are in agreement. This correctness should be within an allowable permit, which again depends on the purpose of the ISM and requirement of the end-users. These four aspects and the allowable permit lead to the following definitions of the validity, validation and validity criteria of ISMs:

'*The validity of an Integrated Systems Model is the soundness and completeness of the model structure together with the plausibility and correctness of the model behaviour.'*

We argued in the previous chapters that absolute validity of an ISM cannot be obtained and validation is a process rather than a final product. The ultimate purpose of the model validation is to establish the confidence of the end-users in the model's usefulness. This leads to our definition of the validation of an ISM:

'*The validation of an Integrated Systems Model is the process of establishing the soundness and completeness of the model structure together with the plausibility and correctness of the model behaviour.'*

A performance criterion defines what aspect (e.g. correctness) of the model we want to examine and what references (e.g. field data or expert experience) are used for this examination. *A validity criterion* is a benchmark needed to determine whether a model is good enough for its designed purposes. This criterion can be either qualitative or quantitative. For instance, Mitchell (mentioned in Chapter 5) proposed a quantitative validity criterion for a predictive model as "ninety-five per cent of the total residual points should lie within the acceptable bound". A qualitative criterion, for example, is 'the modelled behaviour should correspond to the stock-and-flow principle" as mentioned in Chapter 5.

6.3.2. Methodology

This section concerns the answers to the second main research question:

Question 2: *How can the validity of an ISM be established?*

To answer this question, a general framework for the validation of ISMs and its realisation in the detailed steps are described first. The ongoing conclusions are about the methods established to address the three methodological sub-questions: How to establish the validity of the model *system elements and its structure?* How to establish the validity of the *system future behaviour qualitatively*? And how to establish the validity of the *system behaviour quantitatively,* when quantitative data are available only to a limited extent?

In establishing a conceptual framework for the validation of integrated system models, three types of systems are distinguished: the *real system, the model system and the hypothesised system*. The *real system* includes existing components, interactions, causal linkages between those components and the resulting behaviour of the system in reality. The *model system* is the system built by the modellers to simulate the real system, which can help managers in decision-making processes. The *hypothesised system* is the counterpart of the real system, which is constructed for the purpose of model validation. The hypothesised system is created from the readily available data, available knowledge of scientific experts and/or the experience of stakeholders with respect to the real system, which are obtained through a process of learning, observation and reasoning. With the above classification, we can carry out two categories of validation tests: empirical and rational with and without real field data, respectively.

The validation procedure has been organized systematically in 16 steps, ordered in 4 phases in Chapter 2 (Fig. 2.3, Chapter 2). Phase 1 is aimed at specifying relevant inputs, parameters, sub-models, and clusters to the management objective variables (MOVs) under concern. This is conducted by using screening sensitivity analysis. Phase 2 is related to collecting and evaluating validation data from the field, expert knowledge, the experience of stakeholders and literature. A special feature of it is involved with the formulation of a hypothesis with which validation can be carried out in the absence of field data. The use of fuzzy set theory helps to quantify the future behaviour of this hypothesised system under different scenarios. Phase 3 is the testing phase, where the concepts of performance criteria and validity criteria are important. A set of appropriate validation tests for Integrated Systems Model validation is selected, derived and carried out. The procedure ends with phase 4: assessing and reporting. It can be concluded that this systematic procedure helps to reduce the workload and overcome the problem of complexity of Integrated Systems Models.

Question 2.1: *How can the validity of the elements and the structure of an ISM be established?*

To establish the validity of the *model elements and structure*, it is necessary to find an appropriate approach to solve several common problems such as: the lack of field data for model development and validation, the uncertainty involved with these data and the differences between the perceptions of resource managers, stakeholders and of the modellers about problems of concern. The ultimate purpose of this validation is to obtain a model with a complete set of relevant elements (key issues and their causal components that concern stakeholders) and a sound structure. The model should serve as a good tool for discussion between scientific experts and stakeholders. For example, the exclusion of the poison fishing types, discussed in Chapter 3, limits the ability of RaMCo as a discussion tool for describing the future state of the live coral reef with and without solutions.

Approach: to extract the knowledge and experience of *local experts and stakeholders* in term of the key elements and underlying causal relationships to a set of issues; to use sensitivity analysis for determining the model key elements (parameters, inputs, measures) and key assumptions underlying their relationships; to compare the two; to use practical and scientific knowledge gained from the literature to establish the soundness and the completeness of the model components, assumptions and the correctness of the values of parameters and inputs. We conclude that one main purpose of ISM validation is to show transparently both the strengths and weaknesses of a model to the model developers and its users. Furthermore, validation of integrated assessment models is a process, not a final product of integrated assessment. One important component of validation is an adaptive feedback between stakeholders and researchers.

Question 2.2: *How can the validity of the future behaviour descried by an ISM be established?*

To establish the validity of the model qualitative *future behaviour*, it is necessary to find an appropriate approach to solve several key problems, such as: human behaviour (social science) is complex and qualitative in nature; the system under consideration is open, characterised by the uncertain future exogenous variables (e.g. effect of an advance in fishing techniques or the changing subsidy policies for rice culture). This limits the predictive value of the historical data (of internal and external parameters and input variables) for describing the future state of the system. This means also that a good agreement between simulated behaviour and historical data does not guarantee a good agreement between the simulated behaviour and future data. The purpose of this validation approach is to obtain a model which serves as a good tool for facilitating discussion between experts and experts about the system future behaviour. The model should be sound and complete enough to be able to reflect different views on the problems and different solutions to solve them. This is demonstrated by the example discussed in Chapter 4. Due to the too coarse aggregation level of land-use types (i.e. lack of elements), RaMCo fails to describe the consequences of possible future changes in socio-economic factors and policy options on the sediment yields at the inlet of a storage lake.

Approach: to extract *expert knowledge* in the form of qualitative scenarios for building a hypothesised system and qualitative responses of the output variables in the form of inference rules; to use fuzzy set theory to project the future behaviour of the hypothesised system made by the expert; to compare this behaviour with the behaviour produced by the model system in terms of trend lines (i.e. system behaviour modes and patterns). It can be concluded that the process of communicating, persuading, and convincing groups of modellers, experts, and end-users plays a vital role in the process of validating ISMs. One of the possible ways of facilitating this process is to use the historical data and compare them with model outputs (empirical testing). This comparison should be included when possible. However, new methods, focusing for example, on the trend comparison, might be promising and so need to be developed further for validation of ISMs. Our new approach to testing ISM may be useful in situations both where measured data are unavailable and where data are available for the empirical testing.

Question 2.3*: How can the validity of the model behaviour be established if the observed data for validation are available only to a limited extent?*

In establishing the validity of the quantitative model system behaviour when observed data are available, an appropriate approach has been found to address two key problems: the uncertainty involved with calibration and validation data and the philosophical issue related to the criteria for the validity of ISMs. The ultimate purpose of this validation is to obtain a model which can provide a correct trend and a reasonable magnitude of change of the key management variables if some policies or measure are applied. This means that the model should be good enough to provide plausible and accurate behaviour to satisfy these requirements. An example given in Chapter 5 demonstrates that the good fit between observed data and predicted data does not guarantee the plausibility of a model. When the model lacks plausibility, it fails to be a useful tool for policy formulation.

Approach: We proposed using three tests in order to establish the conditional validity of ISMs when data are partially available. The *pattern* test, which uses the inspection of residuals between observed and predicted values. The *accuracy* test, which adopts the concept of confidence bounds based on the analysis of measurement error and the requirements of the decision-makers. The *behaviour plausibility* test is based on the extreme condition test. We conclude that the empirical validity of an ISM is always provisional to the available observed data. The model is safest to be used inside the range of calibrated data. The results of the three above tests should be considered simultaneously in order to conclude about the validity of an ISM.

6.4. Recommendations

6.4.1. Other directions for the validation of integrated system models

Besides the methods set out in this thesis, the following approaches to the validation of ISMs are worth being applied and investigated further:

Quantitative tests of behaviour reproduction: in this thesis, we use a qualitative method (i.e. visual inspection of residual) to measure the pattern replication capability of the model. Quantitative tests of the pattern reproduction, such as the test of the zero slope of regression line of residual (Chapter 5) and the test of the non-difference between the trend coefficient estimators of the measured and computed time data series (Barlas, 1989) could be applied.

Evidence theory: integrated systems validation is hindered by the lack of field data and knowledge about the structure and parameters of the real system. It is beneficial to have a methodological framework that can combine evidence coming from different sources (data, expert knowledge, literature) to support our decision on choosing between the alternative models, structures and values of the parameters. Evidence Theory (Shafer, 1976) can give us such a combining tool. A good explanation with an example of this theory is given by Guan and Bell (1997). Two applications of evidence theory for decision-making can be found in (Caselton and Luo, 1992) and (Beynon et al., 2000).

Tests for the spatial distribution of land cover change: in validating the land-use and land cover change model we see the need to follow the two-stage approach (Veldkamp and Lambin, 2001). This approach separately validates quantitative land use (quantity) first and spatial distribution (location) later. With regard to the validation of spatial distribution, the two techniques, which use longitudinal transecting and Kappa fuzzy statistics, are appropriate. The former method has been applied to validate RaMCo within this validation project (Wismadi, 2003). The latter technique, which is derived by Hagen (2003), has not been used to test RaMCo because of the limited time budget.

Experimental tests of sub-models: in contrast to testing an ISM as a whole model against the observed data and expert knowledge, we can design real experiments in the field to test the parameters and hypotheses of the sub-models, since conducting experiments to test the whole system is too expensive or impossible. An example of conducting an experiment to validate the hydrological and sediment modules in RamCo is given by Huizer and Nieuwenhuis (2003). This approach can be further carried out to validate, for example, the motorisation factor of fishing fleet mentioned in Chapter 5.

6.4.2. Proper use of Integrated Systems Models

As we argued before that Integrated Systems Models always have an added value to other conventional process models, the usefulness of them is confirmed. Nevertheless, the proper use of an ISM is recommended. The functions 3 to 5 (i.e. research prioritising, scenario building and communication functions) can be obtained after a thorough validation process by conducting the tests described in Chapters 3, 4 and 5. For the last function (decision support function), even with these tests, it is safe to use the model to answer qualitatively what-if questions under the contexts similar to the future contexts under which the model was validated (Chapter 4). The quantitative interpretation of the model results can be safely used under the range of conditions under which the model was calibrated and validated against the field data (Chapter 5). As Asian people have a proverb which can be literally translated as ‘a good sword needs a good knight’, well-trained people should be a prerequisite for model usage.

References

Ackoff, R.L., 1981. The art and science of mess management. TIMS Interfaces 2 (1), 20-26.

Alcamo, J., 2001. Scenarios as tools for international environmental assessments. Environmental issue report, No. 24. Experts' corner report. Prospects and Scenarios No. 5. European Environment Agency. Copenhagen, Denmark.

Anonymous, 1977-2002. Laporan Statistik Perikanan Sulawesi Seletan. Dinas Perikanan Propinsi (Statistical fisheries year-books of South Sulawesi).

Aronica, G., Hankin, B., and Beven, K., 1998. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. Advances in Water Resources 22 (4), 349-365.

Ayyub, B.M., 2001. A Practical Guide on Conducting Expert-Opinion Elicitation of Probabilities and Consequences for Corp Facilities. IWR Report 01-R-1. Prepared for U.S. Army Corps of Engineering, Institute for Water Resources, Alexandria, VA 22315-3868.

Balci, O., 1995. Principles and techniques of simulation validation, verification, and testing. Proceedings of the 1995 Winter Simulation Conference. Alexopoulos, C., Kang, K., Lilegdon, W.R., and Goldsman, D. (Ed.). 147-154.

Barlas, Y., 1989. Multiple tests for Validation of System Dynamics Type of Simulation Models. European Journal of Operational Research 42 (1), 59-87.

Barlas, Y., 1994. Model Validation in System Dynamics. Proceedings of the 1994 International System Dynamics Conference. Methodological Issues. Vol., pp. 1-10, Stirling, Scotland.

Barlas, Y. and Kanar, K., 1999. A Dynamic Pattern-oriented Test for Model Validation. Proceedings of 4th Systems Science European Congress, Valencia, Spain, Sept. 1999, pp. 269-286.

Beck, B., 2002. Model evaluation and performance. Encyclopedia of the Environmetrics 3. El-Shaarawi, A.H. and Piegorsch, W.W. (Ed.). 1275-1279.

Beck, M.B. and Chen, J., 2000. Assuring the Quality of Models Designed for Predictive Tasks. In: A. Saltelli, K. Chan, & E.M. Scott, (Eds.), Sensitivity Analysis. Wiley, Chichester, pp. 401-420.

Bennett, J., Lawrence, P., Johnstone, R. and Shaw, R., 2004. Adaptive management and its role in managing Great Barrier Reef water quality. Marine Pollution Bulletin. (Article in Press)

Beynon, M., Curry, B. and Morgan, P., 2000. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling. Omega 28, 37-50.

Campolongo, F. and Saltelli, A., 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. Reliability Eng. & System Safety 57, 49-69.

Caselton, W.F. and Luo, W., 1992. Decision Making with Imprecise Probability: Dempster-Shafer Theory and Application. Water Resources Research 28 (12), 3071-3083.

Checkland, P., 1981. Systems thinking, systems practice. Wiley, Chichester, UK.

Clark, M.J., 2002. Dealing with uncertainty: adaptive approaches to sustainable river management. Aquatic Conservation: Marine and Freshwater Ecosystems 12, 347-363.

Cocks, A.T., Rodgers, I.R., Skeffington R.A. and Webb, A.H., 1998. The limitations of integrated assessment modelling in developing air pollution control policies. Environmental Pollution 102, 635-639.

Comenges, J-M. Z. and Campolongo, F., 2000. An Application of Sensitivity Analysis to Fish Population Dynamics. In: A. Saltelli, K. Chan, & E.M. Scott, (eds.), Sensitivity Analysis. Wiley, Chichester, pp. 367-383.

Cooke, R.M., 1991. Experts In Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press, New York.

Cornelissen, A.M.G., Berg, J.V.D, Koops, W.J., and Kaymak, U. 2003. Elicitation of expert knowledge for fuzzy evaluation of agricultural production systems. Agriculture, Ecosystems and Environment 95 (1), 1-18.

CTI., 1994. The detailed design of environmental improvement works and raw water transmission main in the Bili-Bili multipurpose dam project (phase 2). Design Report. Part 1. Report prepared for the Government of the Republic of Indonesia. Ministry of Public works.

De Kok, J. L. and Wind, H. G., 2002. Rapid assessment of water systems based on internal consistency. Journal of Water Resources Planning and Management, 128(4), 240-247.

De Kok, J.L. and Wind, H.G., 1999. Methodology for sustainable coastal-zone management in the tropics: The integrative research. Final draft report. Dept. of Civil Engineering Technology & Management, Twente University, P.O. Box 217, 7500 AE Enschede, The Netherlands.

De Kok, J.L. and Wind, H.G., 2002. Rapid assessment of water systems based on internal consistency. Journal of Water Resources Planning and Management 128 (4), 240-247.

De Kok, J.L., Engelen, G., White, R., and Wind, H.G., 2001. Modeling land-use change in a decision-support system for coastal zone management. Environmental Modeling and Assessment 6, 123-132.

De Kok, J.L., Titus, M.J. and Wind, H.G., 2000. Application of fuzzy sets and cognitive maps to incorporate social science scenarios in integrated assessment models: A case study of urbanization in Ujung Pandang, Indonesia. Integrated Assessment 1, 177-188.

Dowlatabadi, H., 1995. Integrated assessment models of climate change: An incomplete overview. Energy Policy 23 (4/5), 289-296.

Draeseke, R. and Giles, D.E.A., 2002. A fuzzy logic approach to modeling the New Zealand underground economy. Mathematics and Computers in Simulation 59, 115-123.

Draper, N. R. and Smith, H., 1981 Applied Regression Analysis, 2nd ed. Wiley, New York.

Engelen, G., White, R., and Uljee, I., 2004. The MOLAND for Urban and Regional Growth. Draft Final Report of contract nr. 21512-2003-12 F1SP ISP nl. Submitted to European Commission Joint Research Center, Institute for Environment and Sustainability, Ispra, Italy. Research Institute for Knowledge System bv. P.O. Box 463, 6200 AL Maastricht, The Netherlands *(www.riks.nl)*.

Ewen, J. and Parkin, G., 1996. Validation of catchment models for predicting land-use and climate change impacts. 1. Method. Journal of Hydrology 175, 583-594.

Finlay, P.N. and Wilson, J.M., 1997. Validity of decision support systems: towards a validation methodology. System Research and Behavioral Science 14 (3), 169-182.

Flavelle, P., 1992. A quantitative measure of model validation and its potential use for regulatory purposes. Advances in Water Resources 15, 5-13.

Forrester, J.W. and Senge, P.M., 1980. Tests for building confidence in system dynamics models. TIMS studies in Management Sciences, 14, 209-228. A.A. Legasto, J.R, J.W. Forrester, J.M. Lyneis eds. North Holland.

Forrester, J.W., 1961. Industrial Dynamics. The MIT Press. Cambridge. Massachusetts.

Forrester, J.W., 1969. Urban Dynamics. The MIT Press. Cambridge. Massachusetts.

Forrester, J.W., 1971. World Dynamics. The MIT Press. Cambridge. Massachusetts.

Fox, H.E., Pet, J.S., Dahuri, R. and Caldwell, R.L., 2003. Recovery in rubble fields: long-term impacts of blast fishing. Marine Pollution Bulletin 46, 1024-1031.

Fox, W.W., 1970. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99, 80-88.

Fraedrich, D. and Goldberg, A., 2000. A methodological framework for the validation of predictive simulations. European Journal of Operational Research 124, 55-62.

Fung-Smith, S.J. and Briggs, M.R.P., 1996. Water quality and nutrient discharge of intensive marine shrimp ponds in Thailand and their relationships to pond productivity. Institute of Aquaculture, University of Stirling, Stirling, FK9 4 LA, Scotland, UK. (submitted to Aquaculture)

Gigch, J.P.V., 1974. Applied General Systems Theory. Harper & Row Publishers. NY.

Gonet, M. and Roubelat, F., 1996. Creating the Future: The Use and Misuse of Scenarios. Long Rang Planning 29 (2), 164-171.

Guan, J.W. and Bell, D.A., 1997. Approximate Reasoning and Evidence Theory. Information Science 96, 207-235.

Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. Int. J. Geographical Information Science 17 (3), 235-249.

Helmer, O., 1970. Long-range forecasting: Role and methods. Paper presented at the conference on Forecasting the Future, Harrogate, England.

Hilborn, R. and Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman & Hall, New York, 570 pp.

Hoekstra, A.Y., 1998. Perspectives on water: A model-based exploration of the future. International Books, Utrecht, the Netherlands.

Huizer, K. and Nieuwenhuis, M.J.C., 2003. Verification of the hydrological and erosion section of the RaMCo model Suggested improvements based on a characterisation of the subcatchments: Batulapisi and Bontotence (Sulawesi, Indonesia). Master Thesis. Utrecht University, Utrecht, The Netherlands.

Hulme, M. and Raper, C. B. S., 1995. An integrated framework to address climate change (ESCAPE) and further developments of the global and regional climate modules (MAGICC). *Energy Policy* 23 (4/5), 347-355.

Iman, R.L. and Helton, J.C., 1988. An investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. Risk Analysis 8 (1), 71-90.

Jakeman, A.J. and Letcher, R.A., 2003. Integrated assessment and modelling: features, principles and examples for catchment management. Environmental modelling & Software 18, 491-501.

Jansen, M. and De Vries, B., 1999. Global modeling: managing uncertainty, complexity and incomplete information. In: van Dijkum, C., de Tombe, D. and E. van Kuijck, D. (Eds.), Validation of Simulation Models. Amsterdam: SISWO, pp. 45-69.

Janssen, M. and de Vries., B., 1998. The battle of perspectives: a multi-agent model with adaptive responses to climate change. Ecological Economics 26, 43-65.

Janssen, W. and Goldsworthy, P., 1996. Multidisciplinary Research for Natural Resource Management : Conceptual and Practical Implications. Agricultural Systems 51, 259-279.

JICA, 1994. Master Plan and Feasibility Study on Wastewater and Solid Waste Management for The City of Ujung Pandang in The Republic of Indonesia. Progress Report No. 1. Pacific Consultants International, Tokyo and Yachiyo Engineering Co., LTD., Tokyo.

Kawano, S., Huynh, V.N., Ryoke, M. and Nakamori, Y., 2005. A context-dependent knowledge model for evaluation of regional environment. Environmental Modelling & Software 20, 343-352.

Kirchner, J.W., Hooper, R.P., Kendall, C. et al., 1996. Testing and validating environmental model. The Science of the Total Environment 183, 33-47.

Kleijnen, J. P. C., 1995. Verification and validation of simulation models. European J. Operational Research, 82, 145-162.

Kleijnen, J. P. C., Bettonvil, B. and Groenendaal, W., 1996. Validation of trace-driven simulation models: regression analysis revisited. Proceedings of the 1996 Winter Simulation Conference. Charnes, J.M., Norrice, D.J., Brunner, D.T. and Swain, J.J. (Eds.), 532-359.

Kleindorfer, J.B., O'Neill, L., and Ganeshan, R., 1998. Validation in Simulation: Various Positions in the Philosophy of Science. Management Science 44 (8), 1087-1099.

Konikow, L.P. and Bredehoeft, J.D., 1992. Growth-water models cannot be validated. Advance in Water Resources 15, 75-83.

Kosko, B., 1986. Fuzzy Cognitive Maps. International Journal of Man-Machine Studies 24, 65-75.

Kramer, N.J.T.A. and De Smit, J., 1991. Systeemdenken, Stenfert-Kroese, Leiden, The Nederlands (in Dutch).

Kuhn, T.S., 1970. The Structure of Scientific Revolutions. 2nd ed., University of Chicago Press, Chicago, IL.

Lockeretz, W., 1991. Multidisciplinary research and sustainable agriculture. Biol. Agric.

Hortic. 8, 101-122.

Loehle, C., 1997. A hypothesis testing framework for evaluating ecosystem model performance. Ecological modeling 97, 153-165.

Mace, P.M., 2001. A new role for MSY in single-species and ecosystem approaches to fisheries stock assessment and management. Fish and Fisheries 2, 2-32.

Mamdani, E.H. and Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Study 7, 1-13.

Man, A., Law, R. and Polunin, N.V.C., 1995. Role of marine reserves in recruitment to reef fisheries: a metapopulation model. Biological Conservation 71, 197-204).

Mathworks, 2005. Fuzzy Logic Toolbox User's Guide (For use with MATLAB®) (*http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf*).

Meadows, Donella H., Dennis L. Meadows, J. Randers, and W. W. Behrens, 1972. The Limits to Growth. Universe Books, New York.

Meijerink, S.V., 1995. Cooperation in River Basins; The Scheldt Case. Phys. Chem. 20 (3-4), 215-220.

Mintzberg, H., Raisinghani, D. and Theoret, A., 1976. The structure of 'unstructured' decision processes. Amd. Sci. Q., 21 (June), 246-275.

Miser, H.J. and Quade, E.S., 1985. Handbook of Systems Analysis: Overview of Uses, Procedures, Applications, and Practice. John Wiley and Sons, Chichester.

Miser, H.J. and Quade, E.S., 1988. Handbook of Systems Analysis: Craft Issues and procedural choices. John Wiley and Sons, Chichester.

Mitchell, P. L., 1997. Misuse of regression for empirical validation of models. Agricultural Systems 54, (3), 313-326.

Morgan, M. G. and Henrion, M., 1990. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press.

Morris, D. M., 1991. Factorial sampling plans for preliminary computational experiment. Technometrics, 33(2), 161-174.

Moses, B.S., Udoidiong, O.M. and Okon, O., 2002. A statistical survey of the artisanal fisheries of south-easten Nigeria and the influence of hydroclimatic factors on catch and resource productivity. Fisheries Research 57, 267-278.

Nakamura, T., 2003. Ecosystem-based River Basin Management: its approach and policy-level-application. Hydrological processes 17, 2711-2725.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual model. Part I - A discussion of principles. Journal of Hydrology 10, 282-290.

Nguyen, T.G. and De Kok, J.L., 2003. Application of sensitivity and uncertainty analysis for validation of an Integrated Systems Model for coastal zone management. Proceedings of the International congress on modelling and simulation*,* MODSIM 2003 (ed. Post, D.A.), Modelling & Simulation Society of Australia & New Zealand Inc., Townsville, Australia, 542-547.

Nguyen, T.G., De Kok, J.L. and Titus, M.J., 2005. A new approach to testing and integrated water systems model using qualitative scenarios. Submitted to Environmental Modelling and Software (in revision).

Oreskes, N., 1998. "Evaluation (not validation) of quantitative models, *Environmental Health Perspectives*, 106, supp. 6, 1453-1460.

Oreskes, N., Shrader-Frechette, K. and Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263, 641-646.

Otte, Y., 1997. Impact of shrimp culture development on the water quality at the Southwest coast of Sulawesi. Reports Environmental Studies no. 148. Department of Environmental Studies, Faculty of Sciences, University of Nijmegen, The Netherlands.

Pahl-Wostl, C., 2002. Towards sustainability in the water sector – the importance of human actors and processes of social learning. Aquatic Sciences 64, 394-411

Parker, P., Letcher, R., Jakeman, A.J., et al., 2002. Progress in integrated assessment and modelling. Environmental modelling & Software 17, 209-217.

Pascoe, S. and Robinson, C., 1996. Measuring changes in technical efficiency over time using catch and stock information. Fisheries Research 28, 305-319.

Pet-Soede, C., 2000. Options for co-management of an Indonesian coastal fishery. Ph.D. Thesis. Wageningen University, Wageningen, The Netherlands.

Pet-Soede, C., Cesar, H.S.J and Pet, J.S., 1999. An economic analysis of blast fishing on Indonesian coral reefs. Environmental Conservation 26 (2), 83-93.

Pet-Soede, C., Machiels, M.A.M., Stam, M.A. and Densen, W.L.T., 1999. Trends in an Indonesian coastal fishery based on catch and effort statistics and implications for the perception of the state of the stocks by fisheries officials. Fisheries Research 42 ,41-56.

Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. and Cortés, U., 2004. Designing and building real environmental decision support systems. Environmental Modelling & Software 19, 857-873.

Popper, K.R., 1959. The logic of scientific discovery. Hutching and Son Company. London, UK.

Randers, J., 1980. Elements of the System Dynamics Method. The MIT Press, Cambridge, Massachusetts.

Reckhow, K. H., Clements, J. T. and Dodds, R. C., 1990. Statistical evaluation of mechanistic water-quality model. *J. Env. Eng.,* 116, 250-268.

Refsgaard, J.C and Henriksen, H.J., 2004. Modelling guidelines-terminology and guiding principles. Advance in Water Resources 27, 71-82.

Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholten, H. and Kassahun, A., 2005. Quality assurance in model based water management – review of existing practice and outline of new approaches. Environmental Modelling & Software 20, 1201-1215.

Renema, W. and Troelstra, S.R., 2001. Larger foraminifera distribution on a mesotrophic carbonate shelf in SW Sulawesi (Indonesia). Palaeogeography, Palaeoclimatology, Palaeoecology 175, 125-146.

Rosnay, J.D., 1979. The Macroscope A new world scientific system. Translated from the French by Robert Edwards. Harper & Row, Publishers New York, Hagerstown, San Francisco, London. *(http://pespmc1.vub.ac.be/macroscope/default.html)*

Rotmans, J. and de Vries, B., 1997. Perspectives on Global Change: The TARGETS Approach. Cambridge University Press, Cambridge, UK.

Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90, 229-244.

Saila, S.B., Kocic, V.Lj. and McManus, J.W., 1993. Modelling the effects of destructive fishing practices on tropical coral reefs. Marine Ecology Progress Series 94, 51-60.

Saltelli, A. and Scott, M. 1997. Guest Editorial: The role of sensitivity analysis in the corroboration of models and its link to model structural and parametric uncertainty. Reliability Engineering and System Safety 57, 1-4.

Sargent, R. G., 1984. A Tutorial on Verification and Validation of Simulation Models. *Proceedings of the 1984 Winter Simulation Conference*, 115-121.

Sargent, R. G., 1991. Simulation model verification and validation. *Proceedings of the 1991 Winter Simulation Conference*, 37-47.

Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries Bull. I ATTC / Bol. CIAT 1, 27-56.

Scholten H., Van der Tol, M.W.M. and Smaal, A.C., 1998. Models or measurements? Quantitative validation of an ecophysiological model of mussel growth and reproduction. Paper presented at the ICES Annual Science Conference, Cascais, Portugal.

Scholten, H. and Cate, A.J.U., 1999. Quality assessment of the simulation modeling process. Computers and Electronics in Agricultures 22, 199-208.

Scholten, H. and Van der Tol, M.W.M., 1994. Towards a metrics for simulation model validation. In: Predictability and nonlinear modeling in natural sciences and economics. Grassman, J. and Van Straten, G.(Eds.) Kluwer Academic Publishers, Dordrecht, The Netherlands, 398-410.

Schreider, S.Y. and Mostovaia, A.D., 2001. Model sustainability in DSS design and scenario formulation What are the right scenarios?. Environment International 27, 97-102.

Schwab, P., Cerutti, F. and von Reibnitz, C., 2003. Foresight-using scenarios to shape the future of agriculture research. Foresight 5 (1), 55-61.

Scrase, J.I. and Sheate, W.R., 2002. Integration and Integrated Approach to Assessment: What do they mean for the environment?. Journal of Environmental Policy & Planning 4, 275-294.

SCS., 1971. Sediment sources, yields, and delivery ratios. In National Engineering Handbook, Section 3: Sedimentation. U.S. Department of Agriculture, Soil Conservation Service.

Shafer, G., 1976. A mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey.

Shannon, C. E., 1948. A Mathematical Theory of Communication. The Bell System Technical J. 27, 379-423 and 623-656, July and Oct. 1948.

Shannon, E. R., 1975. *System Simulation: the art and science.* Prentice-Hall, Englewood Cliffs, New Jersey.

Shannon, E.R., 1981. Tests for verification and validation of computer simulation models. Proceedings of the 1998 Winter Simulation Conference, 573-577.

Stergiou, K.I. and Christou, E.D., 1996. Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods. Fisheries Research 25, 105-138.

Sterman, J.D., 2002. All models are wrong: reflections on becoming a system scientist. System Dynamics Review 18, 501-531.

Suriamihardja, D.A. et al., 2001. Study of Integrated Management on Jeneberang Watershed, Phase 2. Final Report made by the cooperation program between Canadian International Development Agency and Center for Environmental Study, Hasanuddin University, Makassar, Indonesia.

Tarantola, S., Jesinghaus, J. and Poulamaa, M., 2000. Global Sensitivity Analysis: A Quality Assurance Tool in Environmental Policy Modelling. In: A. Saltelli, K. Chan, & E.M. Scott, (eds.), Sensitivity Analysis. Wiley, Chichester, pp. 385-397.

Thornton, P.K., 1996. A note on Regression Real-world Data on Model Output. Agricultural systems (50), 411-414.

Turner, R. K., 2000. Integrated natural and socio-economic science in coastal management. J. Marine Systems 25, 447-460.

Uljee I., Engelen, G. and White, R., 1996. RAMCO Demo guide. Modelling and Simulation Research Group, Research Institute for Knowledge Systems b.v. P.O. Box 463, 6200 AL Maastricht, The Netherlands.

Van Asselt, M.B.A. and Rotmans, J., 2002. Uncertainty in Integrated Assessment Modelling: From Positivism to Pluralism. Climate Chang 54, 75-105.

Van Asselt, M.B.A., 2000. Perspectives on Uncertainty and Risk: The PRIMA Approach to Decision Support. Kluwer Academic Publishers. Boston/Dordrecht/London.

Van Ast, J.A., 1999. Trends Towards Interactive Water Management; Developments in International River Basin Management. Phys. Chem. Earth (B) 24 (6), 597-602.

Van der Fels-Klerx, H.J., Horst, H.S. and Dijkhuizen, A.A., 2000. Risk factors for bovine respiratory disease in dairy youngstock in The Netherlands: the perception of experts. Livestock Production Science 66, 35-46.

Van der Heijden, K., 1996. Scenarios: The Art of Strategic Conversation. John Wiley & Sons. Chichester.

Van Gigch, J.P., 1974. Applied General Systems Theory. Harper & Row Publishers. N.Y.

Van Oostenbrugge, J.A.E, Poos, J.J., Van Densen., W.L.T. and Machiels, M.A.M., 2002. In search of a better unit effort in the coastal liftnet fishery with lights for small pelagics in Indonesia. Fisheries Research 59, 43-56.

Van Tongeren, O.F.R., 1995. Data analysis or simulation model: a critical evaluation of some methods. Ecological modeling 78, 51-60.

Veldkamp, A. and Lambin, E.F., 2001. Editorial: Predicting land-use change. Agriculture, Ecosystems and Environment 85, 1-6.

Von Bertalanffy, L. (1968). *General systems theory: Foundations, development, applications*. New York: Braziller.

Von Reibnitz, U., 1988. Scenario techniques. McGraw-Hill, Hamburg.

Walker, W.E., Harremoes, P., Rotmans, J. et al., 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment 4 (1), 5-17.

WCC, 1993. World Coast Conference 1993, Preparing to meet the coastal challenges of the 21st century, Conference Report, Noordwijk, The Netherlands, 1-5 November 1993. Ministry of Transport, Public Works and Water Management. National Institute for Coastal and Marine Management, Coastal Zone Management Centre (RIKZ).

WCED,1987. Our Common Future. The report of the World commission on environment and Development, Oxford University Press, Oxford, UK.

White, R. and Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environment and Planning B: Planning and Design. 4, 235-246.

Wiener, N. 1948. Cybernetics or Control and Communication in the Animal and the Machine, Hermann et Cie, Paris, The MIT Press, Cambridge (Mass.), Wiley and Sons, New York.

Wischmeier, W.H. and Smith, D.D., 1965. Predicting rainfall-erosion losses from cropland east of the Rocky Mountains: Guide for Selection of Practices for Soil and Water Conservation. Agriculture Handbook No. 282. Washington. U.S. Printing Office.

Wischmeier, W.H. and Smith, D.D., 1978. Predicting rainfall-erosion losses. Agriculture Handbook No. 537. U.S. Department of Agriculture. Washington D.C.

Wismadi, A., 2003. Validation of Urban Growth Modelling by GIS and Earth Observation Data Case Study: RaMCo Model for Makassar, Indonesia. Master Thesis. International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands.

Zadeh L.A., 1973. Outline of a New Approach to the Analysis of Complex Systems and Decision Process. IEEE Trans. Systems, Man, Cybernetics, SMC-3, 28-34.

Zio, E. and Apostolakis, G.E., 1997. Accounting for Expert-To-Expert Variability: A Potential Source of Bias in Performance Assessments of High-Level Radioactive Waste Repositories. Ann. Nucl. Energy 24 (10), 751-762.

Zio, E., 1996. On the use of the analytical hierarchy process in the aggregation of expert judgments. Reliability Engineering and System Safety 53, 127-138.

Symbols

A	computed soil loss per unit area
$A(t)$	area of mixed agriculture or rice culture at the year t
a, b, c	regression coefficient of fisheries models
$A_{available}$	total available land in the study area
$A_i(t)$	allocated land demand for land-use type i
$A_n(t)$	area of 'nature' land-use type at the year t
Ave.	average of the ranking of a factor/process given by all experts/stakeholders in combining expert's opinion
$\overline{A_i(t)}$	computed land demand for land-use type i
$B(t)$	fish biomass at time t
C	cover management factor in USLE equation
$c(t)$	system control variable at time t; production cost per unit area of mixed agriculture or rice
$C(t)$	catch rate at time t
$C_e(t)$	catch rate at equilibrium biomass
$d_i(X)$	elementary effect attributable to the input x_i in the Morris design
E	gross Erosion in sediment equation
$E(t)$	total number of trips at year t measured as the number of trips belonging to a "reference" fleet which has no engine; also the fishing effort at time t for fisheries model computations
$E(t)_{uncorrected}$	number of fishing trips obtained from statistical data
E_{MSY}	fishing effort corresponding to Maximum Sustainable Yield
$f(V)$	a dimensionless function whose value ranging form 0 to 1, reflecting the irrigation of the storage lake to compute rice yield
F_i	frequency distribution of elementary effects of factor x_i
k	number of factors considered in the Morris analysis; a parameter corresponding to the unfished equilibrium stock size in fisheries

K	soil erodibility factor
KF_i	fields of expertise, expert *i* has knowledge about
L	slope length factor
LI_i	level of interest of expert *i* on a certain set of questions
LS	the production of slope length factor and slope steepness factor
m	power factor in USLE equation
MSY	maximum sustainable yield
n	number of the inference rules
$N(t)_{inboard}$	number of inboard-engine fishing trips at year t
$N(t)_{nonpow}$	number of nonpower fishing trips at year t
$N(t)_{outboard}$	numbers outboard-engine fishing trips at year t
p	number of levels chosen for each factor in the Morris design
P	support practice factor
$p(t)$	price per tonne of mixed agriculture or rice
PT_i	professional title of expert *i* on a certain set of questions
q	parameter to describe the effectiveness of each unit of effort
r	number of elementary effects to construct a frequency distribution F_i; also the intrinsic rate of fish population growth
R	rainfall factor
R^2	percentage of variance explained by regression
S	slope steepness factor
s	slope percentage in USLE
$s(t)$	system state variable at time t
SDR	sediment delivery ratio
SK_i	source of knowledge of expert *i* on a certain set of questions
Std.	standard deviation of the ranking of a factor given by all experts/stakeholders for combining expert’s opinion
S_y	sediment yield
t	time
$U(t)$, $CPUE(t)$	catch per unit effort (CPUE) at time t
$U_e(t)$	catch per unit effort (CPUE) at time t under equilibrium assumption
w_i	weight assigned to the expert *i*
X	a vector containing k inputs or factors ($x_1,\ldots x_i,\ldots x_k$)

x	Aggregated ranking of a factor given by all experts/stakeholders in weight average method for combining expert's opinion
$x(t)$	system input variable at time t
x_1	numerical value of food crop price in the fuzzy set
x_2	numerical value of cash crop price in the fuzzy set
x_3	numerical value of production cost in the fuzzy set
x_4	numerical value of public investment in the fuzzy set
x_i	input or factor i in Morris design; ranking of a factor/process with regard to its relatively influential importance, given by the expert *i*
x_i^1	minimum values of the uncertainty range of factor x_i
x_i^p	maximum values of the uncertainty range of factor x_i
y	values of a fuzzy output in the numerical domain
$y(t)$	system output variable at time t; yield of mixed agriculture which can accommodate technological changes in RaMCo
$y(X)$	output as a function of the input vector X in Morris analysis
YE_i	years of experience of expert *i* on a certain set of questions
y_{ir}	maximum yields of rice culture with irrigation
y_{nir}	maximum yields of rice culture without irrigation
y_{rice}	yield of rice culture
$Z(t)$	sum of suitability for mixed agriculture or rice culture over all cells occupied by mixed agriculture or rice culture at the year t
$Z_n(t)$	sum of geographical suitability for 'nature' land-use type over all cells occupied by mixed agriculture or rice culture at the year t
$Z_{n,tot}$	extended sum of geographical suitability for 'nature' land-use type over all cells on the map
Z_{tot}	extended sum of geographical suitability for mixed agriculture or rice culture over all cells on the map
α	spatial growth coefficient of rice culture, mixed agriculture or nature
Δ	predetermined increment of factors in the Morris design
$\Delta A(t)$	rate of change of the area of mixed agriculture (rather than rice) and rice culture in RaMCo at year t
$\Delta A_n(t)$	rate of change of the area of 'nature' land-use type in RaMCo
λ	field slope length in USLE

$\mu_1(x_1)$	membership value for the food crop price corresponding to numerical value x_1
$\mu_2(x_2)$	membership value for the cash crop price corresponding to numerical value x_2
$\mu_3(x_3)$	membership value for the production cost corresponding to numerical value x_3
$\mu_4(x_4)$	membership value for the public investment corresponding to numerical value x_4
μ_{AND}	membership value for the rule antecedent
$\mu_{CONS}(y)$	membership function of the fuzzy rule consequent
$\mu^i_{CONS}(y)$	membership function of the fuzzy rule consequent i
μ_I	mean of the frequency distribution F_i in Morris analysis
$\mu_{OUT}(y)$	aggregated membership function of n inference rules
$\eta(t)$	spatial fraction of rice fields which can be irrigated
σ_i	the standard deviation of the frequency distribution F_i
$\delta_n(t)$	area of reforestation at the year t

Acronyms and abbreviations

AMSL	Above Mean Sea Level
Ave.	Average
BOD	Biological Oxygen Demand
CCA	Constrained Cellular Automata
CEDRS	Catch and Effort Data Recording System
CPUE	Catch per Unit Effort
GOF	Goodness-Of-Fit
H	High
IA	Integrated Assessment
IAM	Integrated Assessment Modelling
ISM(s)	Integrated Systems Model(s)
ISW	Integrated Water System
ITC	International Institute for Geo-information Science and Earth observation
KF	Knowledgeable Fields
L	Low
LI	Level of Interest
M	Medium
MF(s)	Membership Functions
MOV(s)	Management Objective Variables
MSY	Maximum Sustainable Yield
RaMCo	Rapid Assessment Model for Coastal zone Management
RIKS	Research Institute for Knowledge System
SA	Sensitivity Analysis
SD	System Dynamics
SDR	Sediment Delivery Ratio

SE(s)	Scientific experts
SK	Source of Knowledge
Std.	Standard deviation
SUA	Sensitivity and Uncertainty Analyses
UK	United Kingdom
UNHAS	Hasanuddin University
US	United State
USLE	Universal Soil Loss Equation
W&H Model	Walters and Hilborn Model
YE	Years of Experience

Summary

The history of validation, verification or evaluation of scientific models is probably as long as that of science itself. Nevertheless, the controversial debate pertaining to the terminology and methodology used to determine the truthfulness, usefulness, trustworthiness and validity of scientific models has not ended yet. The shift of model purpose from describing nature to supporting the human process of regulating and controlling nature is an obvious necessity, because making predictions for open systems is arguably not possible. Describing nature and providing a fit to measurements is necessary but not sufficient. New conditions and future situations, which the model is intended to describe, should be taken into account in order to draw conclusions about the validity of the model.

Integrated Systems Models (ISMs) have been developed over decades to support the planning and management of natural resources and the environment. The development of these models is based on the concepts of systems approach and integrated approach. However, the lack of a generally accepted definition of model validity and model validation, the inherent complexity of ISMs, the poor predictive value of historical data related to the natural-human system, the scarcity of field data and the high level of aggregation of ISMs make the validation of ISMs an extremely difficult task (Chapter 1). These problems raise a number of important questions, such as: to what extent can such models contribute to our knowledge and ability to manage the environment? Do they have added value in comparison with conventional process models? Centred in these questions are the two questions: how can the validity of an ISM be defined? How can this validity be determined? This thesis is aimed at answering these two questions.

The Rapid Assessment Model for Coastal-zone Management (RaMCo), which was developed by a Dutch-Indonesian multidisciplinary team, serves as a case study to achieve the objective of the thesis. The theoretical justification for this choice is that RaMCo contains the typical characteristics of an Integrated Systems Model. First, RaMCo has the ability to take into account the interactions of socio-economic developments, biophysical conditions and policy options. Second, the model includes linkages between many processes pertaining to different scientific fields, such as marine pollution, land-use change, urbanisation, catchment hydrology, coastal hydrodynamics, fisheries, and regional economic development. A practical justification for choosing this model is that validation did not take place during the project. In addition, the availability of new measured data (from 1996 until now) allows for the application of the quantitative techniques which are suitable for the validation of ISMs.

In Chapter 1, the principles and concepts of the systems approach, system dynamics modelling and integrated approach are explained in the context of integrated water

management. These concepts, together with the review of the purposes of ISMs (Chapter 2), form the background that leads to a definition of validity and validation of ISMs. The fundamental characteristic of the integrated systems modelling approach, which differs from traditional modelling, is the focus on model structure, the interaction between system elements and the behaviour (patterns) of the system. The function of an ISM can range from data base and library - for which validation is less important - to systems analysis and decision support, for which validation prior to any practical application is of vital importance.

In the light of the concepts of system modelling and the purposes of ISMs, the validity of ISMs is proposed to pertain to four aspects: the *soundness and completeness of the model structure, the plausibility and the correctness of model behaviour.* S*oundness* of the model's structure is understood to mean that the model's structure should be based on valid reasoning and be free from logical flaws. C*ompleteness* of the structure means that the model includes all elements relevant to the defined problems and their underlying causes which concern the decision-makers and stakeholders. *Plausibility* of model behaviour means that the model behaviour should not contradict general scientific laws and established knowledge. Behaviour *correctness* is understood as the extent to which computed behaviour and measured behaviour are in agreement. Therefore, *the validity of an Integrated Systems Model is* defined as *the soundness and completeness of the model structure together with the plausibility and correctness of the model behaviour.* As a consequence, the validation of ISMs is defined as the process of determining the model validity as defined above.

In view of the definition of model validity and the problems related to the validation of ISMs just mentioned, a conceptual framework and a detailed procedure for the validation of ISMs have been established. These reflect the philosophical position taken in this thesis, which lies somewhere between objectivism (in the sense that there is an ultimate truth) and relativism (one model is as good as any other) and beyond rationalism and empiricism. In the conceptual framework for ISM validation (Fig. 2.2, Chapter 2), three types of systems are distinguished: the *real system, the model system* and *the hypothesised system*. The *real system* includes the components, their interactions, the causal linkages between those components and the resulting behaviour of the system in reality. The *model system* is the system built by the modellers to simulate the real system, which may be used to support the decision-making process. The *hypothesised system* is the counterpart of the real system, which is constructed for the purpose of model validation. It is created from readily available data, available knowledge of the system experts (scientific researchers and decision-makers) and the experience of stakeholders. With this classification, we can carry out two categories of validation tests: empirical and rational. These tests are selected and designed to answer three research questions: how can the validity of the ISM *elements and structure* be determined? How can the validity of ISM *future behaviour* be determined qualitatively? How can the validity of ISM *quantitative behaviour* be determined if measured data are available to a limited extent?

A realisation of the conceptual framework in the form of a general validation procedure is organised in sixteen steps, ordered in four phases (Fig. 2.3, Chapter 2). Phase 1 is aimed at specifying the inputs, parameters, sub-models and clusters of processes that are relevant to the Management Objective Variables (MOVs) of concern. This is done by using screening sensitivity analysis. Phase 2 is related to collecting and evaluating validation data from the field, expert knowledge, experience of stakeholders and literature. A special aspect is the formulation of system experts' hypotheses with which validation can be carried out in the absence of field data. Phase 3 is the testing phase, where the concepts of performance criteria and validity criteria are important. A set of appropriate tests for ISM validation is selected, developed and applied. The procedure ends with Phase 4, assessing and reporting. This general procedure helps to reduce the workload and overcome the problem of the complexity of ISMs.

To determine the validity of the *model elements and structure*, it is necessary to find an appropriate approach to solve several typical problems, such as the lack of field data for model calibration and validation, the uncertainty of these data and the differences in perception between resource managers, stakeholders and modellers about the problems of concern. The ultimate purpose of this validation is to obtain an ISM with a complete set of relevant elements (key issues and causally linked components) that are important to stakeholders and decision-makers, and a sound structure. A validation procedure, which has the above-mentioned characteristics, is described in Chapter 3. The approach is based on the Morris sensitivity analysis, a simple expert elicitation technique, and Monte Carlo analysis to facilitate three validation tests, namely Parameter-Verification, Behaviour-Anomaly and Policy Sensitivity. Two management variables, the living coral reef area and the total pollution load into the coastal waters, expressed in the Biological Oxygen Demand (BOD), are selected as case examples. The application of this validation procedure shows that omitting poison fishing limits the ability of RaMCo as a discussion tool for describing the future state of the living coral.

To determine qualitatively the validity with respect to the system's *future behaviour*, it is necessary to find an appropriate approach to solve a number of particular problems. The system under consideration is open, which is characterised by the uncertain future exogenous variables, for example the effect of an advance in fishing techniques or the changing subsidy policy for rice culture. Human behaviour is complex and therefore the social sciences are largely qualitative by nature. This limits the predictive value of historical data for describing the future state of the system. Agreement between simulated behaviour and historical data does not guarantee agreement between the simulated behaviour and future data. The purpose of this validation approach is to obtain a model which serves well as a tool for facilitating discussions between system experts about the future behaviour of the system. The model should be sound and complete enough to reflect the system experts' consensus about the behaviours of the system, under a chosen set of possible socio-economic and policy scenarios. Chapter 4 describes such an approach. Within this approach, expert knowledge is elicited in the form of qualitative scenarios. These qualitative scenarios are translated into quantitative projections using fuzzy set theory, which is very suitable to deal with the ambiguity and imprecision related to humanistic systems. Trend line comparison between the behaviour projections made by the model and projections based on expert knowledge can reveal structural faults of the model. This is demonstrated by the example discussed in Chapter 4. Due to the too coarse aggregation level of the land-use model (reflected by a lack of erosion-sensitive land-use types), RaMCo fails to describe the consequences of possible future changes in socio-economic factors and policy options on the sediment yield to a storage lake.

In order to determine the validity of the *model behaviour* quantitatively, i.e. if quantitative observations are available, an appropriate approach has been formulated to address two problems: the uncertainty of the data for the model calibration and validation, and the problem of defining quantitative criteria for measuring the validity of an ISM. The ultimate purpose of this kind of validation is to obtain a model which can provide a correct trend and a reasonable magnitude of change in the key management variables under a selected combination of measures and scenarios. This means that the ISM should be good enough to provide plausible and accurate behaviour to satisfy these requirements. Chapter 5 is devoted to the development of a procedure for this purpose, which is tested for the fisheries model of RaMCo. This model is chosen as a case example because the empirical data needed for quantitative validation are easier to obtain for a small-scale model. Residual analysis is proposed to examine the pattern replication ability of the model. The Mitchell (1997) test is used to test the predictive accuracy, and the extreme behaviour test is adopted to test the plausibility of the model behaviour. The example given in Chapter 5 demonstrates that a good fit between observed data and predicted data does not guarantee the plausibility of a model. When the model lacks plausibility, it fails to be a useful tool for policy formulation.

Summarising, this thesis presents a methodology to validate Integrated Systems Models, with three innovative aspects. An appropriate procedure has been established to test the completeness and the soundness of the model structure and elements based on sensitivity and uncertainty analyses. A new approach has been developed to test integrated models using qualitative scenarios, dealing with the problem of uncertain future conditions and the qualitative nature of social sciences and human behaviour. Finally, a procedure has been formulated to test the validity of the quantitative model behaviour when measured data are only available to a limited extent.

Although the proposed methodology has been applied to validate RaMCo, it is expected to be applicable to other ISMs which have the same characteristics as RaMCo. These characteristics include: model complexity, the inclusion of social science, the lack of and large uncertainty of field data for model calibration and validation and the high level of aggregation. The proposed methodology is not only applicable to the validation of ISMs but it can also contribute to the *quality assurance guidelines* for integrated modelling. The two approaches proposed in Chapter 3 and Chapter 4, which allow for the participation of system experts and stakeholders during the model design, can be used to select and refine the conceptual model and to calibrate a site-specific model. The quantitative testing procedure, which is described in Chapter 5, can be applied to validate process models, such as ecological models, hydrological models and hydrodynamic models.

Nevertheless, this thesis is not expected to provide a full guideline for ISM validation with a complete set of tests. Taking the philosophical standpoint, it does not matter how many tests a model has passed, the very next test and/or next data may falsify a model. The tests proposed in this thesis can be considered as the minimum necessary prior to any practical application of an ISM.

Samenvatting

Waarschijnlijk zijn de validatie, verificatie, en evaluatie van wetenschappelijke modellen net zo oud als de wetenschap zelf. Desalniettemin is er nog geen einde gekomen aan de controverse m.b.t. de terminologie en methodologie voor het bepalen van het waarheidsgehalte, de bruikbaarheid, betrouwbaarheid en validiteit van wetenschappelijke modellen. De accentverschuiving van het doel van modellen van beschrijving van de natuur naar de ondersteuning van de regulering en beheersing daarvan is duidelijk noodzakelijk, omdat het aantoonbaar onmogelijk is om voorspellingen te doen over open systemen. Het beschrijven van de natuur en bereiken van overeenstemming met metingen is noodzakelijk, maar niet voldoende. Nieuwe omstandigheden en toekomstige veranderingen, welke het model dient te beschrijven, zouden in acht genomen moeten worden om tot een oordeel te komen over de validiteit van een model.

Reeds gedurende een aantal decennia zijn er Integrale Systeem Modellen (ISMs) ontwikkeld om de beleidsvorming en het beheer van natuurlijke hulpbronnen en het milieu te ondersteunen. De ontwikkeling van deze modellen is gebaseerd op concepten ontleend aan de systeembenadering en integrale benadering. Desondanks leiden het gebrek aan een algemeen geaccepteerde definitie van model validiteit en model validatie, de complexiteit die inherent is aan ISMs, de slechte voorspellingswaarde van historische gegevens m.b.t. het natuur-mens systeem, de schaarsheid van veldgegevens en het hoge aggregatieniveau van ISMs ertoe dat de validatie van ISM's een zeer moeilijke taak is (Hoofdstuk 1). Deze problemen roepen een aantal belangrijke vragen op, zoals: "In welke mate kunnen dergelijke modellen een bijdrage leveren aan onze kennis en ons vermogen om de natuur te beheren?", "Hebben deze modellen een toegevoegde waarde in vergelijking met conventionele procesmodellen?". Besloten in deze vragen zijn de twee onderliggende vragen: "Hoe kan de validiteit van een ISM worden gedefinieerd?", "Hoe kan deze validiteit worden bepaald?". Het doel van dit proefschrift is een antwoord te vinden op deze twee vragen.

Het Rapid Assessment Model for Coastal-Zone Management (RaMCo) is ontwikkeld door een nederlands-indonesisch, multidisciplinair team, en dient als gevalsstudie voor dit proefschrift. De theoretische rechtvaardiging voor deze keuze is dat RaMCo de typische kenmerken van een Integraal Systeem Model heeft. In de eerste plaats biedt RaMCo de mogelijkheid om de interacties tussen sociaal-economische ontwikkelingen, biofysische omstandigheden, en beleidsopties in beschouwing te nemen. In de tweede plaatszijn in het model dwarsverbanden tussen vele processen, zoals de vervuiling van zeewater, verandering van landgebruik, verstedelijking, de hydrologie van stroomgebieden, de kusthydrodynamica, de visserij, en regionale economische ontwikkeling, opgenomen. Een practische rechtvaardiging voor de keuze voor dit model isdat gedurende het project geen validatie plaatsvond. Daarnaast biedt de beschikbaarheid van nieuwe meetgegevens (vanaf 1996 tot nu toe) de mogelijkheid om kwantitatieve technieken, die geschikt zijn voor de validatie van ISMs, toe te passen.

In Hoofdstuk 1 worden de principes en concepten van de systeembenadering, systeemdynamisch modelleren en de integrale aanpak uiteengezet tegen de achtergrond van integraal waterbeheer. Samen met een overzicht van de functies van ISMs (Hoofdstuk 2) vormen deze concepten de achtergrond, van waaruit een definitie van validiteit en validatie van ISMs is ontwikkeld. Het fundamentele kenmerk van de benadering van integrale systeemmodellering, die verschilt van de traditionele wijze van modelleren, is dat de nadruk ligt op modelstructuur, de wisselwerking tussen systeemelementen en het gedrag(spatroon) van het systeem. De functie van een ISM kan variëren van gegevensopslag en bibliotheek, waarvoor validatie minder belangrijk is, tot een systeemanalyse en beslissingsondersteuning, waarvoor validatie voorafgaand aan enige practische toepassing van het model van wezenlijk belang is.

In het licht van de concepten uit de systeemmodellering en de doelen van ISMs wordt voorgesteld de validiteit van ISMs te koppelen aan vier aspecten: de *juistheid en compleetheid van de modelstructuur, de aannemelijkheid en correctheid van het modelgedrag.* Onder de *juistheid* van een model wordt verstaan dat de modelstructuur gebaseerd is op geldige redeneringen en gevrijwaard is van logische tekortkomingen. *Compleetheid* van de structuur betekent dat het model alle elementen omvat, welke relevant zijn voor de gedefinieerde problemen en onderliggende oorzaken, die voor besluitvormers en belanghebbenden van belang zijn.. *Aannemelijkheid* van modelgedrag betekent dat het modelgedrag niet in strijd mag zijn met de regels der wetenschap en gevestigde kennis. Onder de *correctheid* van het modelgedrag wordt de mate waarin het berekende en het waargenomen gedrag overeenstemmen verstaan. Daarom wordt *de validiteit van een Integraal Systeem Model* gedefinieerd als de *juistheid en compleetheid van de modelstructuur, samen met de aannemelijkheid en correctheid van het modelgedrag.* Daaruitvolgend wordt de validatie van ISMs gedefinieerd als het proces dat leidt tot een bepaling van de modelvaliditeit.

Met het oog op de definitie van modelvaliditeit en de zojuist genoemde problemen met betrekking tot de validatie van ISMs, zijn een conceptueel raamwerk en een gedetailleerde procedure voor de validatie van ISMs opgezet. Deze weerspiegelen het filosofische uitgangspunt van dit proefschrift, dat zich ergens tussen het objectivisme (in de zin dat er een ultieme waarheid is) en het relativisme (het ene model is net zo goed als enig ander model), en voorbij rationalisme en empirisme, bevindt. Binnen het conceptuele raamwerk (Fig 2.2, Hoofdstuk 2), worden drie soorten systemen onderscheiden: het *werkelijke systeem, het modelsysteem,* en het *gehypothetiseerde systeem*. Het *werkelijke systeem* omvat de bestaande componenten met hun interactiesen causale verbanden tussen deze componenten, en het daaruit volgende gedrag van het systeem, zoals dat in werkelijkheid bestaat. Het *model systeem* is het systeem, dat door de modelontwikkelaars geconstrueerd is om het werkelijke systeem te simuleren, en kan worden ingezet om de besluitvorming te ondersteunen. Het *gehypothetiseerde systeem* is de tegenhanger van het werkelijke systeem, en is geconstrueerd voor modelvalidatie. Het is gebaseerd op eenvoudig beschikbare gegevens, de beschikbare kennis van systeemdeskundigen (wetenschappers en besluitnemers) en de ervaring van belanghebbenden. Met deze indeling kunnen we twee categorieën van validatietoetsen, empirische en rationele, uitvoeren. Deze testen zijn gekozen en ontworpen om drie onderzoeksvragen te beantwoorden: "Hoe kan de validiteit van *de elementen en structuur* van een ISM worden bepaald?", "Hoe kan kwalitatief de validiteit van het *toekomstgedrag* als beschreven door een ISM worden bepaald?", en "Hoe kan de validiteit van het *kwantitatieve modelgedrag* worden bepaald, indien meetgegevens in beperkte mate beschikbaar zijn?".

Een realisatie van het conceptueel raamwerk binnen een algemene validatie procedure is gevormd rond zestien stappen, die in vier fasen zijn geordend (Fig 2.3, Hoofstuk 2). De eerste fase is erop gericht de ingangsvariabelen, parameters, submodellen, en clusters van processen te specificeren, die relevant zijn voor de betreffende Management Doel Variabelen (MDVs). Dit gebeurt door een gevoeligheidsanalyse op basis van screening. De tweede fase houdt verband met het verzamelen en waarderen van validatiegegevens op basis van veldbezoek, expertkennis, de ervaring van belanghebbenden, en literatuur. Een bijzonder aspect is het formuleren van systeemhypotheses door experts, waarmee validatie zonder veldgegevens mogelijk wordt. De derde fase betreft het toetsen, waarbij de concepten van doelmatigheidscriteria en validiteitscriteria belangrijk zijn. Voor de validatie van Integrale Systeem Modellen zijn een aantal geschikte toetsen uitgekozen, ontwikkeld, en toegepast. De procedure eindigt met de vierde fase, die bestaat uit het waarderen en rapportage. Deze algemene procedure draagt bij aan de verlichting van de werklast en aan de aanpak van het probleem van de complexiteit van ISMs.

Teneinde de validiteit van *de elementen en structuur van modellen* te bepalen, is het noodzakelijk een geschikte benadering te vinden, waarmee een aantal kenmerkende problemen, zoals het gebrek aan veldgegevens voor de calibratie en validatie van modellen, de onzekerheid van deze gegevens, en het verschil in perceptie tussen de beheerders van hulpbronnen, belanghebbenden en de modelontwikkelaars met betrekking tot de problemen, die van belang zijn. Het uiteindelijke doel van deze validatie is te komen tot een ISM met een complete verzameling relevante elementen (hoofdkwesties en de oorzakelijk gekoppelde componenten), die voor belanghebbenden en besluitvormers een rol spelen, en een juiste structuur. In Hoofdstuk 3 wordt een validatie procedure met bovengenoemde eigenschappen beschreven. De benadering is gebaseerd op de gevoeligheidsanalyse volgens Morris, een eenvoudige techniek om informatie aan experts te onttrekken, en Monte Carlo analyse, waarmee drie validatietoetsen, namelijk Parameterverificatie, Gedragsanomalie, en Beleidsgevoeligheid, kunnen worden uitgevoerd. Twee beheersvariabelen, het oppervlak levend koraalrif, en de totale uitstoot van afvalstoffen in de kustwateren, uitgedrukt in de Biologische Zuurstofbehoefte (BZB), zijn als voorbeeld uitgekozen. De toepassing van deze validatie procedure toont aan dat het weglaten van gifvisserij het vermogen van RaMCo om als discussieinstrument voor de beschrijving van de toekomstige toestand van het levend koraal te dienen beperkt.

Teneinde kwalitatief de validiteit van een model met betrekking tot het *toekomstgedrag* te bepalen, is het noodzakelijk een geschikte benadering te vinden, waarmee een aantal bijzondere problemen kunnen worden opgelost. Het beschouwde systeem is open, en wordt gekenmerkt door onzekere, toekomstige, exogene variabelen, bijvoorbeeld het effect van een vooruitgang in visvangsttechniek of een veranderd subsidiebeleid t.a.v. de rijstcultuur.Het menselijk gedrag is complex en daarmee is de beschrijving daarvan door de sociale wetenschappen grotendeels kwalitatief van aard. Dit beperkt de voorspellende waarde van historische gegevens voor het beschrijven van de toekomstige systeemtoestand. Overeenstemming tussen het gesimuleerde gedrag en historische waarnemingen waarborgt niet een overeenstemming tussen het gesimuleerde gedrag en toekomstige waarnemingen. Het doel van deze validatiemethode is een model te verkrijgen, dat goed als instrument kan dienen ter vereenvoudiging van discussies tussen systeemexperts met betrekking tot het toekomstige gedrag van het systeem. Het model dient voldoende correct en compleet te zijn om de overeenstemming weer te geven, die onder deskundigen bestaat over de gedragspatronen van het systeem onder een gekozen verzameling mogelijke sociaal-economische en politieke scenarios. Hoofdstuk 4 beschrijft een dergelijke aanpak. Binnen deze aanpakwordt kennis in de vorm van kwalitatieve scenarios aan experts onttrokken. Deze kwalitatieve scenarios worden vertaald in kwantitatieve projecties door middel van de theorie van vage verzamelingen. Deze is zeer geschikt om de dubbelzinnigheid en onnauwkeurigheid, die samenhangen met menselijke systemen, aan te pakken. Structurele fouten van het model kunnen worden blootgelegd door de trendcurves tengevolge van het geprojecteerde modelgedrag te vergelijken met de projecties op basis van expertmeningen. Het voorbeeld dat in Hoofdstuk 4 wordt beschreven vormt een voorbeeld hiervan. Tengevolge van het te hoge aggregatieniveau van het landgebruiksmodel (weerspiegeld door het ontbreken van erosiegevoelige landgebruiksklassen) is RaMCo niet in staat om de gevolgen van toekomstige veranderingen in sociaal-economische factoren en beleidskeuzes op de sedimentvracht naar een stuwmeer te beschrijven.

Teneinde de validiteit van het *modelgedrag* kwantitatief te beschrijven, d.w.z. indien kwantitatieve waarnemingen beschikbaar zijn, is een geschikte methode geformuleerd, waarmee twee problemen kunnen worden aangepakt: de onzekerheid die samenhangt met de calibratie en validatie van modellen en het probleem dat samenhangt met de definitie van kwantitatieve criteria om de validiteit van ISMs te meten. Het uiteindelijke doel van dit type validatie is een model te verkrijgen, dat een correcte trend en een redelijke orde van grootte in de verandering van belangrijke beheersvariabelen kan geven, indien een bepaalde combinatie van maatregelen en scenarios wordt toegepast. Dit betekent dat het ISM zo goed dient te zijn dat aannemelijk en nauwkeurig gedrag wordt vertoond, zodat aan deze eisen voldaan kan worden. Hoofdstuk 5 is gewijd aan de ontwikkeling van een procedure hiervoor, die getoetst wordt methet visserijmodel van RamCo. Dit model is gekozen als voorbeeld omdat het voor kleinschalige modellen eenvoudiger is om de empirische gegevens, die nodig zijn voor kwantitatieve validatie, te vezamelen. Binnen deze methode wordt residuele analyse voorgesteld om het vermogen van het model om gedragspatronen te reproduceren, onderzocht. De Mitchell (1997) toets is ingezet om de voorspellende nauwkeurigheid te toetsen, en de extreme gedrags toets is gebruikt om de aannemelijkheid van het modelgedrag te toetsen. Het voorbeeld, dat in dit hoofdstuk wordt beschreven, toont aan dat een goede overeenstemming tussen waargenomen en voorspelde gegevens de aanemelijkheid van een model niet garandeert. Indien het model tekort schiet in aannemelijkheid is het niet bruikbaar als instrument voor beleidsvoorbereiding.

Samengevat wordt in dit proefschrift een methodologie voor de validatie van Integrale Systeem Modellen gepresenteerd, met drie vernieuwende aspecten. Op basis van gevoeligheids- onzekerheidsanalyses is een geschikte procedure voor het toetsen van de compleetheid en juistheid van de modelstructuur en elementen opgezet. Daarnaast is een nieuwe benadering ontwikkeld om integrale modellen met kwalitatieve scenarios te toetsen, waarmee het probleem van onzekere toekomstige omstandigheden en de kwalitatieve aard van de sociale wetenschappen en het menselijk gedrag kan worden aangepakt. Tenslotte is een procedure geformuleerd om de validiteit van het kwantitatieve modelgedrag te toetsen indien meetgegevens beperkt beschikbaar zijn.

Hoewel de voorgestelde methodologie is toegepast om RaMCo te valideren is de verwachting dat deze inzetbaar zal zijn voor andere ISMs, die dezelfde kenmerken hebben als RaMCo. Deze kenmerken zijn: model complexiteit, de rol van de sociale wetenschappen, het gebrek aan en de grote onzekerheid omtrent de veldgegevens voor de modelcalibratie en modelvalidatie, en het hoge aggregatieniveau. De voorgestelde methodologie is niet alleen toepasbaar voor de validatie van ISMs, maar kan ook een belangrijke bijdrage leveren aan *richtlijnen ter waarborging van de kwaliteit* van het integraal modelleren. De twee benaderingen die in Hoofdstuk 3 en Hoofdstuk 4 worden voorgesteld om systeemexperts en belanghebbenden bij de modelontwikkeling te betrekken, kan worden ingezet om het conceptuele model te kiezen en te verfijnen, alsmedeeen locatieafhankelijk model te calibreren. De kwantitatieve procedure voor het toetsen, die in Hoofdstuk 5 wordt beschreven, kan worden toegepast om andere soorten modellen, zoals ecologische, hydrologische, en hydrodynamische modellen, te valideren.

De verwachting is echter niet dat dit proefschrift een volledige richtsnoer voor de validatie van ISMs biedt met een complete verzameling toetsen. Filosofisch beschouwd maakt het geen verschil hoeveel toetsen een model doorstaat, de eerstvolgende toets en/of gegevens kunnen een model falsifiëren. De toetsen, die in dit proefschrift worden voorgesteld, zouden moeten worden beschouwd als het minimum dat noodzakelijk is, voorafgaand aan enige practische toepassing van een ISM.

About the author

Nguyen Tien Giang was born in Hanoi, Vietnam. He received his Engineering Degree, with distinction, majoring in Hydrology and Environment from Hanoi Water Resources University in 1997. His bachelor thesis was entitled ‘Water balance in the Upper Serepok Basin for socio-economic development up to the year 2010’. From 1997 to 1998 he worked as a lecturer at the Faculty of Hydro-meteorology and Oceanography of the University of Science, Hanoi National University.

He enrolled at the Asian Institute of Technology (AIT), in Bangkok, Thailand to follow a master programme, and majored in Water Resource Development (1998-2000). The fund for this two-year study period was granted by the Danish International Development Agency (DANIDA). The topic of his master thesis was ‘Sediment transport balance and bank erosion in the Son Tay curved bend, Red River, Vietnam’. He was awarded the M.E. degree in April, 2000, with an excellent grade for his thesis.

After completing the master programme, he received a grant from the Japan International Cooperation Agency (JICA) and worked at AIT for one year (2000-2001) as a research assistant in the department of Water Engineering and Management at the School of Civil Engineering. His research was involved with the development of a Two-dimensional Riverbed Evolution Model constructed in general non-orthogonal curvilinear coordinate system.

In August, 2001 he joined the group of Water Engineering and Management as a PhD student, at Twente University, Enschede, The Netherlands. The PhD programme has been funded by both the Netherlands Foundation for the Advancement of Tropical Research (WOTRO) and the University of Twente. Four years of working mainly with integrated systems models, uncertainty and sensitivity analyses, fuzzy logics, expert elicitation and regression analysis have resulted in this thesis.

After the completion of this PhD research, he will return to Vietnam and work at the Faculty of Hydro-meteorology and Oceanography as a lecturer.