BayFish – Tonle Sap: A Bayesian Model of The Fish Production In The Tonle Sap Great Lake, Cambodia

[1]Baran, E. and [2]T.A. Jantunen

[1]WorldFish Center, [2]Environmental Consultant, E-Mail: e.baran@cgiar.org

Keywords: *Fisheries, Natural Resources Management, Bayesian Belief Networks, Modelling*

EXTENDED ABSTRACT

Probabilistic Bayesian networks have been used as decision making tools mainly in medicine and economics, but recently they have also been utilized in natural resource management. In this study a Bayesian model is built as a decision-making tool for fisheries management in the Tonle Sap Lake in Cambodia. As the largest freshwater fishery in the world, the Tonle Sap has enormous importance to both Cambodia and the whole of Lower Mekong Basin. The Tonle Sap Lake is an extremely dynamic and unique system due to the annual reversal of its outlet, which inundates large areas of the floodplain. In this case Bayesian network was used as the modelling tool for fisheries production because it was readily available and easily adaptable for this system. No effort was taken to test other modelling options or compare this model to other models, which also could have been used for this study. As in many developing countries there is a serious lack of data and knowledge of the system. To compensate for this deficiency, extensive stakeholders consultation was undertaken. The model is designed as a tool for decision makers responsible for the management of the Tonle Sap Lake fisheries.

The Bayesian network models have interacting variables, which can be defined with quantitative data or qualitative stakeholders consultations. Some 38 stakeholders were consulted during the model-building phase from various international, national and local organizations. The consultations resulted in the progressive building and defining of a model of the Tonle Sap inland fisheries, whereas the integration of hydrological, water quality and environmental databases contributed to the reliability and accuracy of the model. Elicitation of probabilities for variables without data and setting of numeric thresholds for data supported variables was also done by stakeholders.

Examples of model structure construction and development as well as variable defining and parameterization are given mainly from hydrological and environmental aspects of the model.

The results from a comparison between the model's predicted fish stocks and the actual fish catch data from Dai fisheries at the Tonle Sap River between 1995 and 2003 show a very close correlation. A set of scenarios were developed based on published literature to test impacts of dam building, land use change and deterioration of water quality. High development in upstream countries as well as within the Tonle Sap Basin could reduce fish catches by 1/3. Obviously this would have serious deteriorating impact not only on biodiversity and environment, but also on livelihoods of people and sustainable development of Cambodia.

Even though it is impossible to incorporate dynamic spatial and temporal scales into the model it can be a very useful tool for predictions, impact analysis and understanding the natural system. The model can also be used to inform decision makers, civil servants and other stakeholders about linkages and trade-offs of the system. A drawback of the approach is the limited number of variables and parameters that can be included in the model, and the method does not allow feedback mechanisms. Another drawback is the model output data, which is only provided as indicative probabilities, not direct quantities or values. On the other hand, the model is much less data-hungry than traditional fish production models.

This paper presents the results and experiences from the model construction, development and results as well as reviews the advantages and drawbacks of the Bayesian modelling approach.

1. INTRODUCTION

As demand for freshwater steadily increases, decision makers at a national as well at basin level require information on the role of river flow in sustaining environmental benefits and tools to assess the necessary trade-offs between different water uses. River and floodplain fisheries are one of these benefits, and in the case of Cambodia are assets of remarkably high importance for the country.

Inland fisheries amounted to 360,000 tons in 2002 (Department of Fisheries), contributing up to 16% of the GDP. Roughly 60 % is derived from the Tonle Sap Lake. However, detailed scientific monitoring shows that this annual catch varies from year to year (Ngor Peng Bun 2000, Baran et al. 2001a, 2001b). Recent studies have also shown that the fish production in the Mekong Basin is dependant upon a number of hydrological, environmental and ecological factors (Baran et al. 2001a). A modelling approach is the only possible way to integrate all these parameters because the intricacy of factors is beyond the reach of individual experts and the number of interacting variables would require decades of data for a standard statistical approach (Baran and Baird 2003). For example, 60 annual cycles would be required to test all the interactions of four environmental variables on annual fish production with non parametric method, the least data-hungry, approach (Sokal et al., 1981).

A number of reviews of modelling approaches and tools for tropical floodplain rivers management have demonstrated an interest in modelling approaches based on Bayesian networks (Baran 2002; Arthington et al. 2004). As some other models they also allow the integration of quantitative as well of qualitative information (databases or expert knowledge), and they are intuitive, flexible and powerful. However, this paper only aims to demonstrate Bayesian modelling approach for fisheries, and does not compare other modelling options.

For the purposes of illustration, the model nodes are referred in the text in Boxes and parameters of these nodes in *"Italics"*. The final full model can be seen in Appendix A.

2. BAYESIAN NETWORKS

In the mid-90s Bayesian networks were developed as Decision Support Systems (DSS) for medical diagnostics and financial risk assessment. Since late '90s Bayesian networks have been applied more widely, particularly in natural resources management.

The model consists in defining the system studied as a network of variables, which are connected by links expressed in terms of probabilities. These variables can be either quantitative (e.g. Flooded vegetation) or qualitative (e.g. Subsistence fisher activity). Parent nodes (driving variables) connect to child nodes (driven variables).

One of the challenges when building a network is in defining enough, but not too many, variables. For each variable a small number of parameters are defined (e.g. Flood duration "*More than 11 weeks*" or "*Less than 6 weeks*"). If data is available, then the software automatically converts the quantified relationship between two variables into probabilities. If data is not available, then expert knowledge from stakeholders consultations can be used to elicit the known relationship between two variables in terms of probabilities. Once the network has been parameterized (an operation called "elicitation of probabilities"), the software can calculate the probability of an output or event based on the weight given to each variable and parameter of the system.

3. STAKEHOLDERS CONSULTATION

Although sometimes overlooked in models developed so far, a stakeholder consultation is useful in identifying the relevant variables of the system to be managed, to determine links and cause-effect relationships, and to elicit probabilities within the networks. This is especially important in developing countries where quantitative data is often scarce and scientific knowledge of the structure of the modelled natural system is poor.

In using Bayesian networks for environmental management, the consultation of experts and stakeholders is acknowledged as being of critical importance (Borsuk et al. 2001; Cain et al. 2003). Many studies have described stakeholders consultations, but on the practical side, the guidelines provided by Cain (2001), and Ravnborg and Westermann (2002) for stakeholders consultations are among the most detailed; and acknowledging the lack of concise and pragmatic recommendations, Baran and Jantunen (2004) have prepared guidelines regarding stakeholders consultation for Bayesian modelling in environmental management.

The BayFish - Tonle Sap model has been built completely following the recommendations of 38 stakeholders overall, met during four one-day workshops. The stakeholders were selected by the modellers with an intention of having a balanced representation between the groups. The composition was improved during the process. The meetings were attended by a majority of stakeholders involved in the fisheries sector, from national agencies to local organizations. Environmental and socio-economic disciplines were also represented, in particular hydrology, water quality and environmental valuation. Among other disciplines, managers and policy-makers were also present. The attendance of independent scientists and representatives from fisher organizations balanced the number of specialists from the governmental agencies, which is coherent with the target of the tool developed.

Several consultations were necessary so that the modellers could progressively convert the information provided by stakeholders into a computer model. This back-and-forth process also permitted the identification of missing variables, incoherencies and mistakes. The three main steps of the consultation consisted of; a) building the model framework, b) defining the model variables and c) parameterizing the variables. In addition to this consultative process, available databases on hydrology and land use were integrated into the model (Jantunen 2004) in order to support it with as much quantitative information as possible.

4. BUILDING THE MODEL FRAMEWORK AND DEFINING THE VARIABLES

The decision was made to concentrate on only a few sets of variables in order to present all the stages of the model development in this paper. Examples are provided from two of the main variable categories: hydrology and environment. A detailed description of the model can be found at Baran et al. (2005).

4.1. Model objective

The overall goal of the BayFish – Tonle Sap model network is to model the total production of the Tonle Sap Lake and its floodplain. Its food production results from two separate variables, or nodes, Tonle Sap agricultural production and Tonle Sap fish harvest (Figure 1). Fish harvest depends on hydrology, environment, fish migrations and fishing; whereas agricultural production depends on hydrology and other agronomic factors. Agricultural production was included in order to take into account the major aspects of food production in the floodplain. While detailed modelling of all agricultural variables was outside the scope of the present study, it may be included in the future.

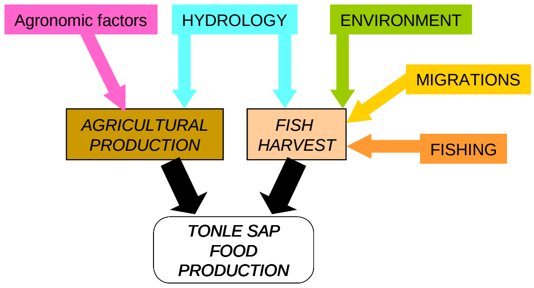


Figure 1. Main variables contributing to Tonle Sap food production

4.2. Hydrology

Flooding for fish is understood as the combination of three interacting hydrological variables: Date of floodplain flooding, Flood duration and Floodplain flood level. Of these variables Flood duration is affected by Date of floodplain flooding and Floodplain flood level. I.e. Earlier and higher floods cause flood duration probabilities to increase. Similarly, Floodplain flood level is affected by Date of floodplain flooding because earlier floods have a greater possibility of causing higher floods. The date of floodplain flooding is based on the moment when water level at the Lake reaches 4m and flood duration on the length of time from the flood beginning to the reversal of the flow in the Tonle Sap River. Flooding for fish is simply qualified as "*Good*" or "*Bad*", as this adequately describes the quality of a hydrological year from a fisheries perspective. Data is used with stakeholders consultations to define most of the other variables. E.g. Flood duration parameters are "*Less than 6 weeks*" (short flood), "*Around 8 weeks*" (6 to 11 weeks, normal flood) and "*More than 11 weeks*" (long flood).

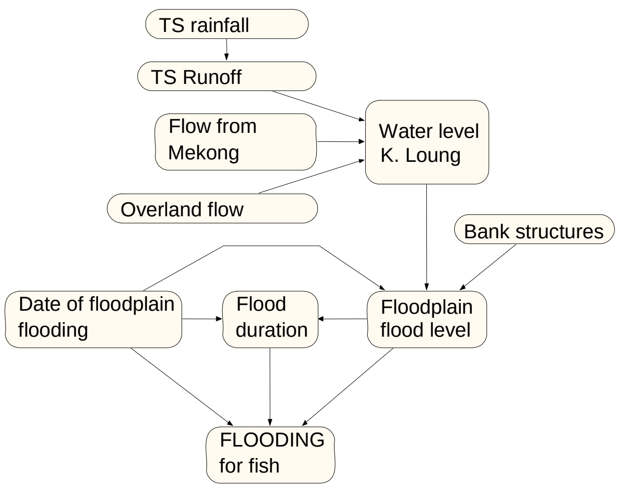


Figure 2. Variables contributing to Flooding for fish

Further up the hydrology network structure Floodplain flood level is affected by Bank structures and Water level at K. Loung. Water level is then determined by the combination of Flow from Mekong, Overland flow and Tonle Sap runoff from Tonle Sap basin tributaries. Tonle Sap Rainfall is the parent node for runoff. Most of these variables are defined in terms "*Above*" or "*Below*" their respective input dataset average. Defining more than two states would generate a non-manageable complexity in the probability table (e.g. Table 1), with incompatible combinations and unrealistic data requirements.

4.3. Environment

Habitat for Black fish and Habitat for White fish are understood as the quality of the environment used by fishes in these groups (Figure 3). Stakeholders and recent studies show that the parameters of critical importance to both major fish types are the Dissolved Oxygen (DO) levels in the floodplain (Oxygen for Black fish and Oxygen for White fish) and the nature of the vegetation in the floodplain (Flooded vegetation). Incidentally DO is the only indicator of proven importance to fish production as other chemical parameters could not be related to fish production. In general the Lake is well oxygenated due to wind and wave induced aeration, but parts of floodplain are largely anoxic due to the decay of vegetation (Sarkkula et al. 2003).

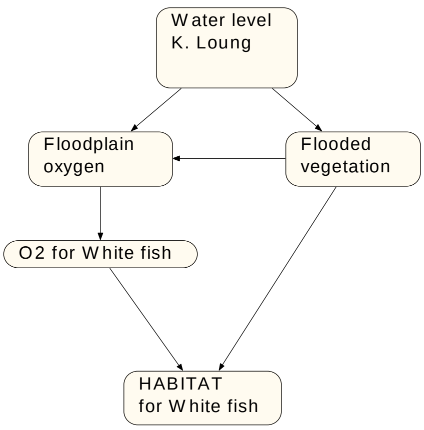


Figure 3. Variables contributing to Habitat for White (or Black) fish

Floodplain oxygen has been defined, after a review of literature and of FishBase 2004, as being "*Above 4 mg/l*" (value acceptable to almost all fish), "*Between 2 and 4 mg/l*" (values acceptable to Black fishe but too low for White fishe) and "*Below 2 mg/l*" (values too low for any fish). This rough classification is based in particular on a consultation of local aqua-culturists. Flooded vegetation is defined in terms of land use type "*Grass*", "*Shrub*" or "*Forest*", whereas the other variables are loosely defined as "*Good*" or "*Bad*" (Habitat) and "*Acceptable*" or "*Impossible*" (DO).

5. PARAMETERIZING THE VARIABLES

Parameterizing the variables consists, for driving variables, of attributing a probability to each state and, for driven variables, to each combination of states originating from driving variables. For all these parameters the user can refer to the model itself, in which all probability tables are open to viewing and modification. Thresholds for data series were set by stakeholders after initial data analysis. The software computes the probabilities from data after an input file is downloaded into the model.

5.1. Hydrology

For Water level at K. Loung the reference is the maximum annual water level at Kampong Loung and the parameterization (Table 1) is derived from MIKE11 model simulated output results for the time period of 1985 to 2003. The same dataset was used for Water level at K. Loung parent nodes.

Table 1. Water level at Kampong Loung probability table

Variable parameters			Water level at K. Loung		
Flow from Mekong	Overland Flow	TS runoff	Above 10m	Between 8 and 10m	Below 8m
Above 37000	Above 7600	Above 30000	42.85	42.85	14.3
Above 37000	Above 7600	Below 30000	40	40	20
Above 37000	Below 7600	Above 30000	25	50	25
Above 37000	Below 7600	Below 30000	20	60	20
Below 37000	Above 7600	Above 30000	40	40	20
Below 37000	Above 7600	Below 30000	15	55	30
Below 37000	Below 7600	Above 30000	16.6	66.8	16.6
Below 37000	Below 7600	Below 30000	12.5	37.5	50

5.2. Environment

Habitat for Black fish and Habitat for White fish were elicited by a fisheries expert. "*Impossible*" DO is always 100% detrimental for fish. Although

"*Forest*" is in the Mekong Basin considered as the best habitat for fish, the fish catch has not declined dramatically after considerable deforestation, and therefore "*Shrub*" is also regarded as a "*Good*" habitat (90%). "*Grass*" does not provide shelter and sustenance in the way that "*Shrub*" and "*Forest*" do, and therefore it is elicited as only 50% "*Good*".

6. RESULTS

6.1. Baseline scenario

The baseline scenario was compared with Dai fisheries annual fish catch data. This data is regarded as the best fish catch data available in Cambodia, and it reflects Tonle Sap Lake fish production quite well. The years 1995 to 2003 were used for the comparison, because pre-1995 data is unreliable. Flood properties for TS Runoff, Flow from Mekong, Overland Flow, Date of floodplain flooding and Flood duration for each year were used to obtain fish stock probabilities. The years 1996 to 1997 and 2000 to 2001 have exactly same flood properties, and so fish stock model results for these years are the same.

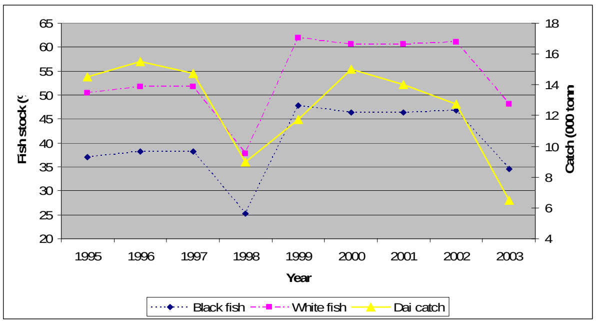


Figure 4. Comparison between Dai fisheries catch data and model baseline output data for Black and White fish stocks (Dai fish catch from Starr 2004)

When compared to published Dai fisheries catch data for the same period of time the shape of the curve fits with Black and White fish stocks remarkably well (Figure 4). The model results show that the best year for the Tonle Sap fish stock according to the model was 1999, closely followed by 2002, 2000 and 2001. This does not correspond with the Dai fisheries data, but it is likely that the extremely low fish catch in 1998 also had an effect on the year 1999, which is not taken into account in the model. The model not allowing feedback mechanisms is, however, a major drawback. The years 2000 and 2001 are somewhat lower in the model results because in the current model very high flood levels reduce DO levels in the floodplain due to large anoxic areas in the floodplain periphery. This causes "*Good*" habitat probabilities to decrease, thereby decreasing fish stocks as well.

6.2. Development scenarios

Largely based on WorldBank (2004) publication scenarios were developed for High and Low upstream and Tonle Sap Basin development. Basically all upstream development options, such as hydropower and irrigation schemes, would delay flood beginning ad reduce its height and duration. Within Tonle Sap Basin floodplain shrub and forest cover would decrease undermining habitat quality. Construction of bank structures would also delay flooding of the floodplains.

Similar analysis was performed as with Baseline scenario for years 1998 to 2000. Low development scenario resulted in 6.2% and 7.3% unit reduction in average Black and White fish stocks, whereas High development scenario resulted in 10.6% and 13.4% reduction respectively. Direct calculation of fish catch reduction in tons from the probabilities provided by the model is impossible. However, given that in the Baseline scenario (Figure 4) minor variations from year to year in fish stocks are matched with variations of several thousand tons in fish catch, it could be argued that over 10% unit reduction in fish stock probabilities could mean 5,000 tons reduction in fish catch. This would be approximately 1/3 of the average annual Dai fisheries fish catch.

7. CONCLUSIONS

The Bayesian approach is a vibrant modelling option for situations where the structure of the system is not well known or the available data nonexistent. Including years of expert knowledge, often untapped in scientific studies and model building, can substantially improve and develop a model. A Bayesian model can be used as a teaching and training tool for decision makers, civil servants and other stakeholders to improve their understanding of the linkages and trade-offs of a given system. However, the model output is in probabilities which can only be used indicatively for management decisions and scientific predictions. Other modelling approaches could also be used for the same purpose though.

BayFish – Tonle Sap model has proven in scenario analysis the accuracy obtainable with the combination of data integration and extensive stakeholders consultations into a Bayesian Belief Network. Even though the model is simplified it can be used as an efficient management and planning tool for the Tonle Sap fisheries and environment. The next step of the model development involves training of decision makers in the using the model and fine tuning the model according to their feedback.

8. REFERENCES

Arthington, A.H., E. Baran, C.A. Brown, P. Dugan, A.S. Halls, J.M. King, C.V. Minte-Vera, R.E. Tharme, and R.L. Welcomme (2004), Water requirements of floodplain rivers and fisheries: existing decision-support tools and pathways for development, *Comprehensive Assessment of Water Management in Agriculture*, 59 pp., In press.

Baran, E. (2002), Modelling the Mekong fisheries: what can be done? In: Sverdrup-Jensen, S. (ed.), *Fisheries in the Lower Mekong Basin: status and perspectives*, MRC Technical paper nº 6, Mekong River Commission, 103 pp., Phnom Penh.

Baran, E., and I.G. Baird (2003), Approaches and tools for sustainable management of fish resources in the Mekong River Basin, In: Cao M., K., Woods, H., Hu, and L., Li (Eds.), *Biodiversity management and sustainable development*, China Forestry Publishing House, 246 pp., Kunming.

Baran, E., and T. Jantunen (2004), Stakeholder consultation for Bayesian Decision Support Systems in environmental management. *ECOMOD 2004*, Regional Conference on Environmental and Ecological Modelling, Penang, Malaysia, September 15-16.

Baran, E., T. Jantunen, S. Hort, P. Chheng (2005), *Building "BayFish-Tonle Sap", a model of the Tonle Sap fish resource*, ADB/WorldFish Center project "Technical Assistance for capacity building of IFReDI"; WorldFish Center and Inland Fisheries Research and Development Institute, Department of Fisheries, 50 pp., Phnom Penh.

Baran, E., N. van Zalinge, Ngor Peng Bun (2001a), Floods, floodplains and fish production in the Mekong Basin: present and past trends, *Proceedings of the Second Asian Wetlands Symposium*, Penang, Malaysia, August 27-30.

Baran, E., N. van Zalinge, Ngor Peng Bun (2001b), *Analysis of the Cambodian Bagnet ("dai") fishery data,* ICLARM, Mekong River Commission Secretariat and Department of Fisheries, 62 pp., Phnom Penh.

Borsuk, M.E., R.T. Clemen, L.A. Maguire, and K.H. Reckhow (2001), Stakeholder values and scientific modeling in the Neuse River watershed, *Group Decision and Negotiation*, 10, 355-373.

Cain, J. (2001), Planning improvements in natural resources management, *Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond*, Centre for Ecology and Hydrology (CEH), 124 pp., Wallingford.

Cain, J., K. Jinapala, I. Makin, P. Somaratna, B. Ariyaratna, and L. Perera (2003), Participatory decision support for agricultural management: A case study from Sri Lanka, *Agricultural Systems*, 76, 457–482, Elsevier Science.

Jantunen, T. (2004), *Integration of hydrological, land use and water quality data to a model of the Tonle Sap fish resource*. ADB/WorldFish Center project "Technical Assistance for capacity building of IFReDI"; WorldFish Center and Inland Fisheries Research and Development Institute, Department of Fisheries, 56 pp., Phnom Penh.

Ngor Peng Bun (2000), Dai fishery in the Tonle Sap River of Phnom Penh and Kandal province, In: van Zalinge, N., T. Nao, and S. Lieng (eds.), Management aspects of Cambodia's Freshwater fisheries. *Presentations at annual meeting of the Department of Fisheries*, Ministry of Agriculture, Forestry and Fisheries, Phnom Penh, Cambodia, January 27-28.

Ravnborg, H., and O. Westermann (2002), Understanding interdependencies: stakeholder identification and negotiation for collective natural resource management, *Agricultural Systems*, 73, 41–56.

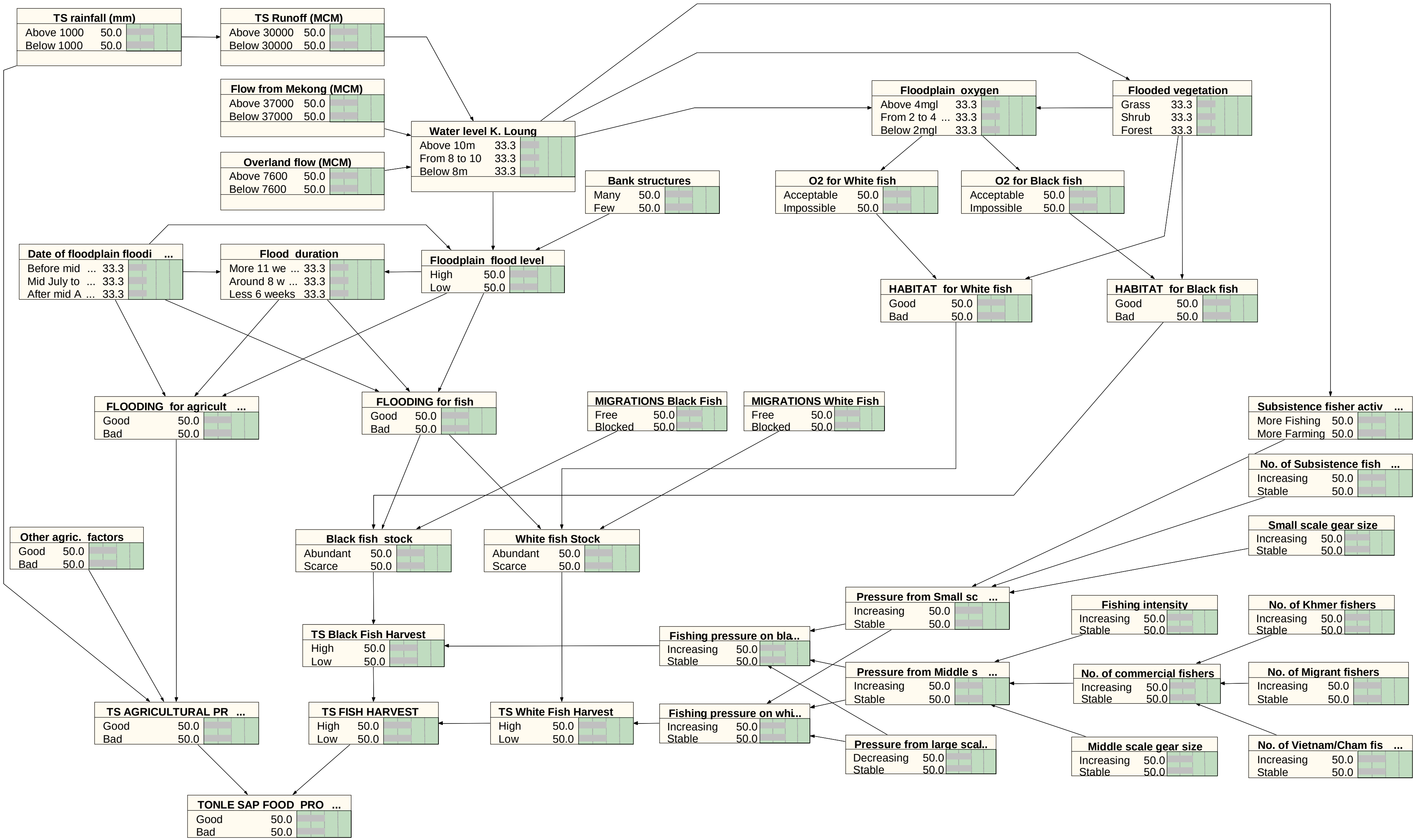
Sarkkula, J., Koponen J. et al. 2003. Modelling of the flow regime and water quality of the Tonle Sap. MRCS/WUP-FIN final project report, May 2003. Mekong River Commission and Finnish Environment Institute, 109 pp.

Sokal, R.R., and F.J. Rohlf (1981), The principles and practice of statistics in biological research, *Biometry*, WH Freeman and Company, 857 pp., San Francisco.

Starr, P. (2004), *Catch and culture*, Mekong Rriver Commission, 10(1).

WorldBank (2004), *Modelled Observations on Development Scenarios in the Lower Mekong Basin*, WorldBank, Vientiane.

9. APPENDIX A



TS rainfall (mm)
Above 1000 50.0
Below 1000 50.0
TS Runoff (MCM)
Above 30000 50.0
Below 30000 50.0
Flow from Mekong (MCM)
Above 37000 50.0
Below 37000 50.0
Overland flow (MCM)
Above 7600 50.0
Below 7600 50.0
Water level K. Loung
Above 10m 33.3
From 8 to 10 33.3
Below 8m 33.3
Bank structures
Many 50.0
Few 50.0
Floodplain oxygen
Above 4mgl 33.3
From 2 to 4 ... 33.3
Below 2mgl 33.3
Flooded vegetation
Grass 33.3
Shrub 33.3
Forest 33.3
O2 for White fish
Acceptable 50.0
Impossible 50.0
O2 for Black fish
Acceptable 50.0
Impossible 50.0
Date of floodplain floodi ...
Before mid ... 33.3
Mid July to ... 33.3
After mid A ... 33.3
Flood duration
More 11 we ... 33.3
Around 8 w ... 33.3
Less 6 weeks 33.3
Floodplain flood level
High 50.0
Low 50.0
HABITAT for White fish
Good 50.0
Bad 50.0
HABITAT for Black fish
Good 50.0
Bad 50.0
FLOODING for agricult ...
Good 50.0
Bad 50.0
FLOODING for fish
Good 50.0
Bad 50.0
MIGRATIONS Black Fish
Free 50.0
Blocked 50.0
MIGRATIONS White Fish
Free 50.0
Blocked 50.0
Subsistence fisher activ ...
More Fishing 50.0
More Farming 50.0
No. of Subsistence fish ...
Increasing 50.0
Stable 50.0
Small scale gear size
Increasing 50.0
Stable 50.0
Other agric. factors
Good 50.0
Bad 50.0
Black fish stock
Abundant 50.0
Scarce 50.0
White fish Stock
Abundant 50.0
Scarce 50.0
Pressure from Small sc ...
Increasing 50.0
Stable 50.0
Fishing intensity
Increasing 50.0
Stable 50.0
No. of Khmer fishers
Increasing 50.0
Stable 50.0
TS Black Fish Harvest
High 50.0
Low 50.0
Fishing pressure on bla..
Increasing 50.0
Stable 50.0
Pressure from Middle s ...
Increasing 50.0
Stable 50.0
No. of commercial fishers
Increasing 50.0
Stable 50.0
No. of Migrant fishers
Increasing 50.0
Stable 50.0
TS AGRICULTURAL PR ...
Good 50.0
Bad 50.0
TS FISH HARVEST
High 50.0
Low 50.0
TS White Fish Harvest
High 50.0
Low 50.0
Fishing pressure on whi..
Increasing 50.0
Stable 50.0
Pressure from large scal..
Decreasing 50.0
Stable 50.0
Middle scale gear size
Increasing 50.0
Stable 50.0
No. of Vietnam/Cham fis ...
Increasing 50.0
Stable 50.0
TONLE SAP FOOD PRO ...
Good 50.0
Bad 50.0