STOCKHOLM ENVIRONMENT INSTITUTE

Stockholm Environment Institute, Working Paper – 2009

Assessing Water-related Poverty Using the Sustainable Livelihoods Framework

Eric Kemp-Benedict, Sukaina Bharwani, Elnora de la Rosa, Chayanis Krittasudthacheewa, Nilufar Matin

Assessing Water-related Poverty Using the Sustainable Livelihoods Framework

Eric Kemp-Benedict,* Sukaina Bharwani, Elnora de la Rosa,
Chayanis Krittasudthacheewa, Nilufar Matin

* Coordinating author; email: erickb@sei-us.org

Stockholm Environment Institute
Kräftriket 2B
SE 106 91 Stockholm
Sweden

Tel: +46 8 674 7070
Fax: +46 8 674 7020
Web: www.sei-international.org

Head of Communications: Robert Watt
Publications Manager: Erik Willis
Proofreading: Mark Hoffman
Layout: Tyler Kemp-Benedict

Cover Photo: © Eric Kemp-Benedict

SEI STOCKHOLM ENVIRONMENT INSTITUTE

CONTENTS

ABSTRACT

Local circumstances and adaptations affect water-related interventions and livelihood outcomes. This local variation is crucial for developing resilient livelihood strategies, but also creates significant analytical challenges to assessing the likely impacts of water-related interventions. This report presents an approach using a probabilistic, 'fuzzy' model of the links between water and livelihoods that takes these fundamental uncertainties into account. The model is grounded in the Sustainable Livelihoods framework, and is implemented as a Bayesian network. The approach is applied to data from a previous study in Northeast Thailand, and the research was supplemented by field visits and key informant interviews at farms, communities, and universities in Northeast Thailand. This report presents a conceptual framework for analysing water-related interventions on poverty, an elicitation approach, and an example application. Also, it presents three innovations that resulted from the project: a novel way to represent institutions within the Sustainable Livelihood framework, a Bayesian approach to representing indicators as indirect evidence of a quantity of interest, and an elicitation technique for the conditional probability tables within a Bayesian network model.

ACKNOWLEDGEMENTS

The research described in this working paper started within the Mekong Basin Focal Project (BFP) of the CGIAR Challenge Program on Water and Food. The authors wish to express their sincerest thanks to Simon Cook, who headed the BFP, and Mac Kirby, who led the Mekong BFP. Matt Chadwick, of SEI, was a champion of Bayesian methods within the BFP, and without his inspiration and enthusiasm this project would never have been undertaken. The authors also wish to express their gratitude to the professors, students, farmers, and others who provided insights for this work. Thanks to Dr. Vichien Kerdsuk, Dilok Sarawdi, and Patcharin Ruchuwararak of Khon Kaen University for taking the time to comment on the methods and model in this paper, and for taking us to an integrated farm in Pra Bhudhabath Village, whose owners generously showed us their farm and training center. Thanks also to Assistant Professor Yanyong Inmuong and Dr. Ajara Boobpapan of Mahasarakham University and their students for bringing together representatives from the Department of Water Resources and WWF to comment on this work. Dr. Ajara and the students kindly took us to two integrated farms, one in Don Daeng villlage and one in Don Man village, where our hosts patiently answered our many questions. Finally, thanks to Associate Professors Watcharapong Wattanakul, Kanjana Payooha, and Tuantong Jutagate of Ubon Ratchathani University for bringing together several of their colleagues and graduate students to critique the methods in this report. We were then shown Pak Mun Dam, and conversed with a fisherman who worked downstream of the dam and told us what impacts the dam had had on his livelihood.

1 INTRODUCTION

Communities and households are complex, and they can respond in surprising ways to interventions (Easterly, 2006; Scott, 1998). The challenge is to assess the likelihood of success of interventions using techniques that take into account the particulars of the community in which the intervention will be introduced but that do not create prohibitive or intrusive demands for data (Bharwani, 2006; Newton *et al*., 2006). In this report we describe such a technique for assessing the potential impact of interventions on water-related poverty.

The main assumption underlying this approach is that indicators of water-related poverty—numbers on or statements about the target community—are indirect evidence of an underlying reality that can be usefully, even if only approximately, described by a model. The model structure, as well as the links between the indicators and the model, are specific to the community and can be discovered through elicitation and by fitting to observed data. It is further argued that the model should be grounded on the concept of sustainable livelihoods, and this report makes use of the Sustainable Livelihoods framework of the UK Department for International Development (DFID, 1999).

The ideas presented in this report were implemented using data collected for the Challenge Program on Water and Food (CPWF) Mekong Basin Focal Project for a collection of villages in Si Sa Ket province in Northeast Thailand (SEI, 2008). To supplement the report, key informant interviews and field visits were carried out at farms, communities, and universities in the provinces of Khon Kaen, Mahasarakham, and Ubon Ratchathani in Northeast Thailand. Key informant interviews with farmers in Khon Kaen and Mahasarakham will be used throughout the report for illustration. The locations of the field visits are shown in Figure 1, with the location of Ubonrattana Dam and the centre of Khon Kaen indicated.

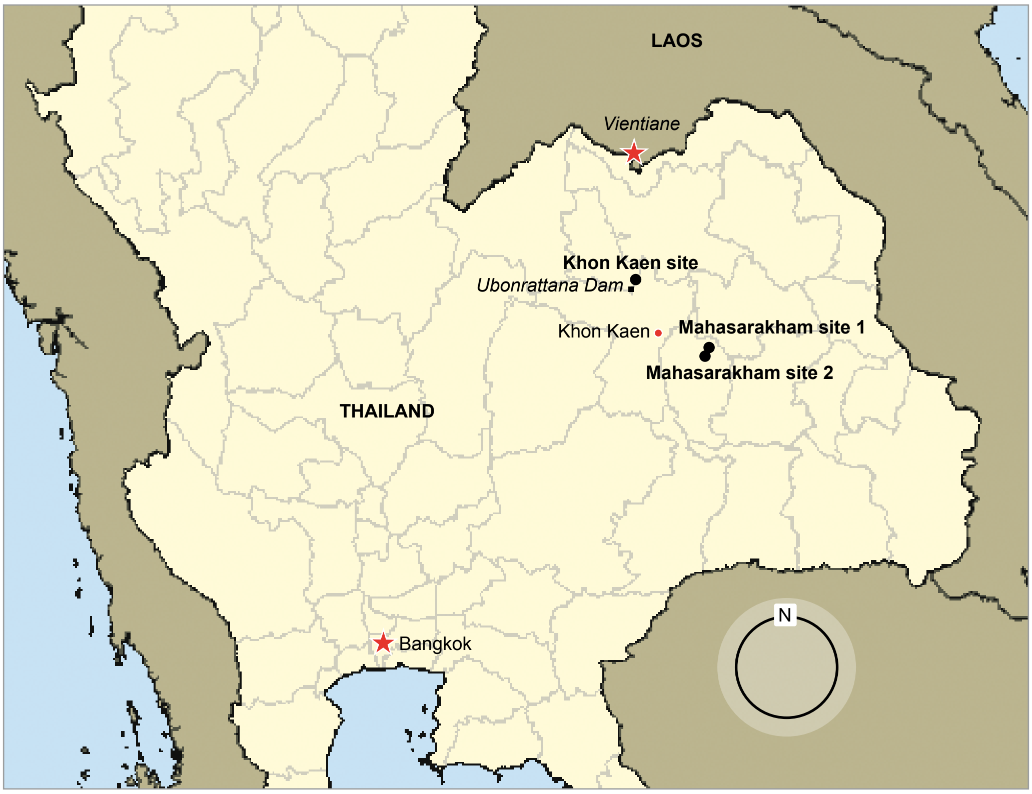

Figure 1: Location of Khon Kaen and Mahasarakham field visits

Water-related poverty

Water-related poverty emerges when water resources constrain or impact upon people's livelihood options and assets. This may occur because of flooding or drought, limited access, expense, water quality, or water-borne disease (Black & Hall, 2004) and depends upon both the characteristics of the resource and the capacity of the community making use of it (Sullivan & Meigh, 2003). Because this report has its roots in the CGIAR Challenge Program on Water and Food, we focus on the interactions between water resources, agriculture, and livelihoods that can contribute to water-related poverty (Cook & Gichuki, 2006).

Water and livelihoods

Water can be seen as flowing through three interlinked systems: a hydrologic system, a food production system, and a livelihood system (Cook & Gichuki, 2006). Interventions for poverty reduction can be targeted at any of these systems, and might include the provision of water resources, protection of environmental flows, protection from health hazards, and, especially for agricultural water use, increases in water productivity.

Water productivity expresses the socioeconomic and environmental benefits derived from the use of water (Molden *et al*., 2007). A system that can deliver more benefits with a given amount of water than another system has a higher water productivity, and broadly speaking there are three ways to accomplish this: by increasing socioeconomic or environmental services, by reducing agricultural water depletion, and by decreasing negative impacts on other systems. In this report the focus is on crop production, and the water productivity concept we employ is that of crop-water productivity. The basic links between water and crop-supported livelihoods are shown in Figure 2. As shown in the figure, the water available for agricultural production is determined partly by the natural water availability (the hydrological system) and partly by the available water infrastructure (part of the food production system). The available water is then used to produce crops, in an amount determined by the water productivity (the rest of the food production system). The produce is then used to support livelihood goals resulting in livelihood outcomes.

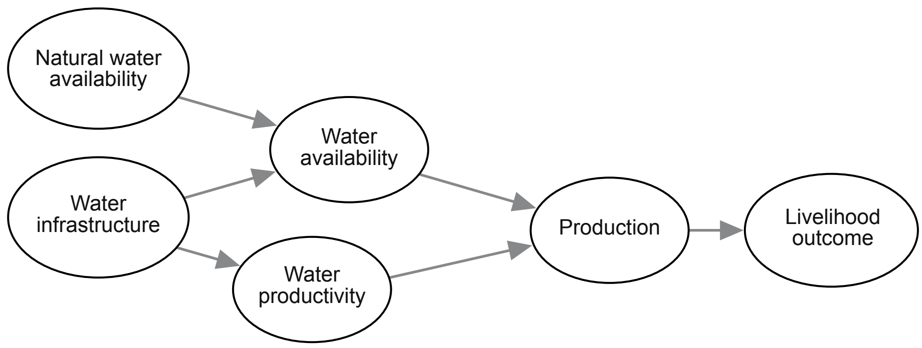

Figure 2: Basic links between water and livelihoods

The crucial final arrow in Figure 2 that links production and outcome is the livelihood system in Cook & Gichuki's formulation. Representation of this system and the livelihood aspects of the food production system are the focus of most of the rest of this report.

2 CONCEPTUAL FRAMEWORK

The approach described in this report is based on the Sustainable Livelihoods (SL) framework of the UK Department for International Development (DFID, 1999). The SL framework seeks to place people at the centre of development efforts, and views people as deploying assets to reach their goals within a context of vulnerability. At the present time the SL framework is widely used and accepted. Therefore, while we will describe its features, we will not present detailed arguments for using it in contrast to other approaches. For those wishing to understand the context in which the SL framework arose, Bebbington (1999) provides a sharp critique of then-current approaches, and argues persuasively for an asset-based livelihoods approach to developing effective rural poverty-reduction interventions, while Nicol (2000) argues that the sustainable livelihoods framework can help improve outcomes for water projects.

Sustainable Livelihoods framework

The Sustainable Livelihoods framework is an asset-based (Carter & Barrett, 2006) poverty and vulnerability analytical framework. It is described in DFID's SL Guidance Sheets as follows (DFID, 1999):

> In its simplest form, the framework views people as operating in a context of vulnerability. Within this context, they have access to certain assets or poverty-reducing factors. These gain their meaning and value through the prevailing social, institutional, and organisational environment. This environment also influences the livelihood strategies—ways of combining and using assets—that are open to people in pursuit of beneficial livelihood outcomes that meet their own livelihood objectives.

The components and relationships in the SL framework are shown in Figure 3. Five categories of assets, or 'capitals', are identified (DFID, 1999; Scoones, 1998). Social capital is the set of networks and relationships that support coordinated strategies for achieving livelihood goals. Human capital is individual skills and knowledge, as well as health and physical ability, that can be mobilized in livelihood strategies. Physical capital is the infrastructure, equipment, and other long-lived physical goods that people, households, and communities can bring into use. Financial capital is the pool of economic assets, including savings, cash or other liquid assets, and credit. Finally, natural capital is the natural resource stocks and services that can be used to support livelihood outcomes, including soil, water, genetic variability, and pollution sinks.

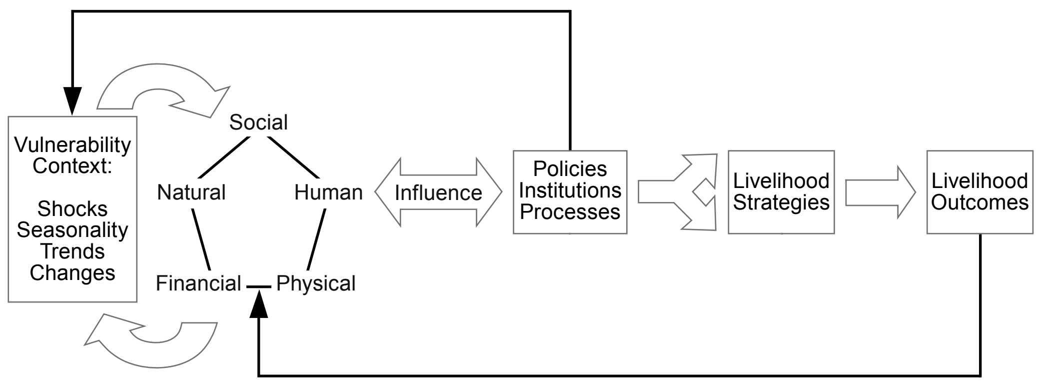

Figure 3: The DFID Sustainable Livelihoods framework

When responding to shocks, as well as during life transformations, households deploy their assets in different combinations to try to meet livelihood goals (Bharwani *et al*., 2008; Moench, 2005). Within the project summarized in this report, the interview with a farmer in Mahasarakham (Mahasarakham Site 1 in Figure 1) provides an illuminating example of a changing mix of assets. The farmer, referred to in this report as MS1, had been a mechanic, and wished to start his own company. He had some land—natural capital—and mortgaged it to obtain financial capital. He then bought tools necessary for his new business, converting some of his financial capital into physical capital. His timing was unfortunate, coinciding with an economic downturn, and his business did not do well. The bank was unsympathetic and planned to take his land. However, he realized that he could get a better price for his land than the bank had initially loaned to him. So, he managed to repay the loan to the bank and sell most of the land, while retaining a small plot of land for himself. Family members—representing his social capital—then lent him some further land. At this point he had very little to work with except for the loaned land, but he described how he and his wife had sat down and counted their assets, and pointed out that while they had little else, they still had their hands. That is, they had human capital. They used that remaining asset to begin working their farm. It is now a successful integrated farm, and serves as a teaching centre for other farmers wishing to learn integrated farming techniques. Due to MS1's work and that of his family, both the physical and natural capital of the farm is improved since he began to work it, and he has built up financial assets. He supports, and is supported by, an extended social network.

The links between the SL asset categories and the basic framework in Figure 2 are shown in Figure 4. The node 'natural assets' is meant to encompass all those natural assets other than natural water availability, which is included explicitly, while 'physical assets' encompasses all physical assets other than water infrastructure. Some aspects of the diagram merit discussion. The node 'production' has no links to the capitals. In this framework it is taken to be a direct consequence of water availability and water productivity, which are influenced by the capitals. Water availability is affected by natural water availability (natural capital) and water infrastructure (physical capital), as well as social capital. It is also influenced by institutions, as discussed below. Water productivity is affected by water infrastructure and other physical capital, as well as financial, natural, and human capital. The capitals then mediate between production and livelihood outcomes. The extent to which production is converted to livelihood outcomes depends in part on the assets available to households and the strategies they employ.

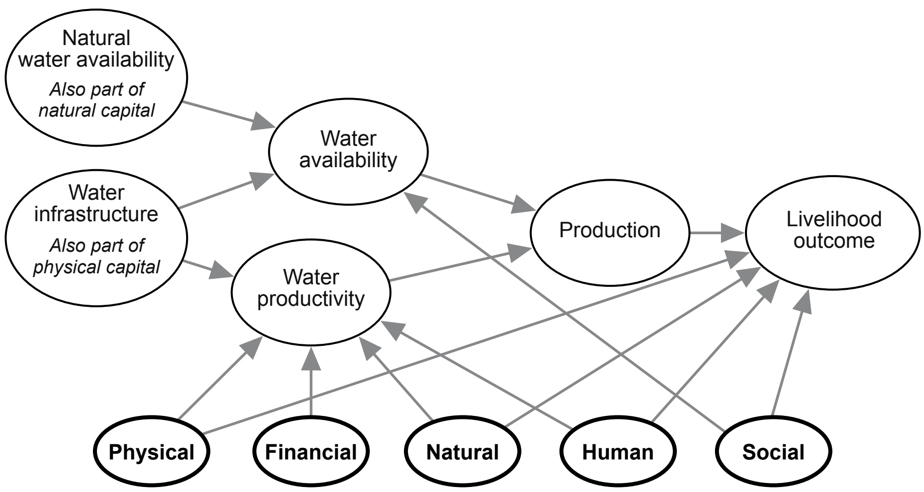

Figure 4: The SL livelihood assets added to the basic framework

Institutions

The network shown in Figure 4 does not reflect all of the processes and dynamics in the SL framework (shown in Figure 3). Instead, it focuses on the connection between livelihood assets, water, and livelihood outcomes. Both vulnerability and livelihood strategies are represented to some extent in Figure 4, through natural water availability—which may feature floods, droughts, or both—and the links between water availability, agricultural production, and outcomes. However, the institutional context is absent, and must be included (Carloni & Crowley, 2005; Saravanan, 2008). The capability approach to poverty provides a more useful frame for thinking about institutions than does an asset-based approach, and we turn to it for inspiration. Amartya Sen defines capabilities as, "the substantive freedoms [an individual] enjoys to lead the kind of life he or she has reason to value" (Sen, 1999). From a capability perspective, income and assets are instrumental, but not intrinsic, to the definition of poverty, and factors other than assets can contribute to, or subtract from, an individual's capabilities. Of particular importance are the formal and informal institutions within which people operate.

Institutions are also important in the DFID SL framework, in which people—with their livelihood assets and strategies—are viewed as embedded within a network of institutions (Scoones, 1998). Institutions affect the transformation of incomes and resources into capabilities and opportunities. This is shown in Figure 5, in which institutions influence the transition between water infrastructure and water availability, between assets and water productivity, between assets and access to markets, and between production and livelihood outcomes.

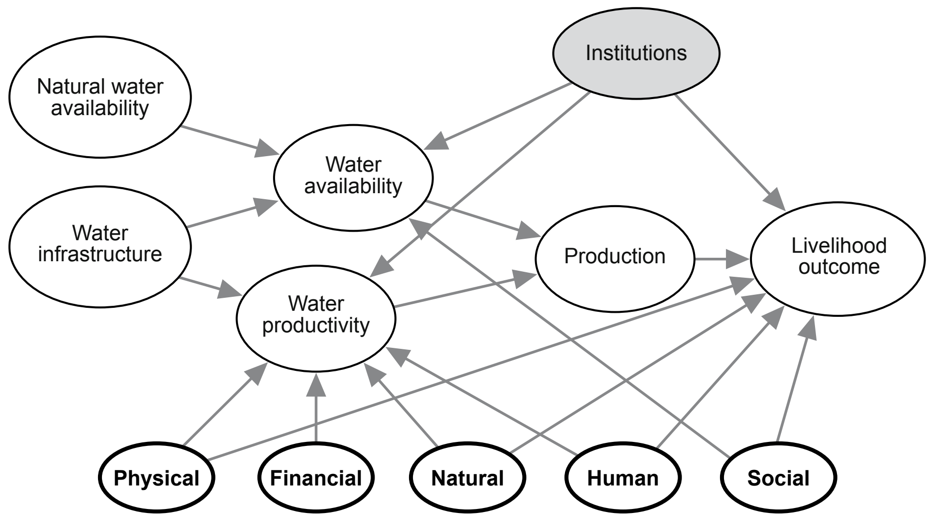

Figure 5: Adding institutions to the basic framework

Institutions are difficult to represent in any kind of analytical framework. From Scoones (1998):

> Institutions may be both formal and informal, often fluid and ambiguous, and usually subject to multiple interpretations by different actors. Power relations are embedded within institutional forms, making contestation over institutional practices, rules and norms always important. Institutions are also dynamic, continually being shaped and reshaped over time.

The fact that institutions can both mediate and consolidate power relations makes it difficult to elucidate and represent how they function. In order to represent institutions within our conceptual framework we turn to a theory of the sociologist Charles Tilly, which suggests that while institutions may change their form, the results they produce are comparatively stable and observable. Specifically, he argues (Tilly, 1998),

> Paired and unequal categories, consisting of asymmetrical relations across a socially recognized (and usually incomplete) dividing line between interpersonal networks, recur in a wide variety of situations, with the usual effect being the unequal exclusion of each network from resources controlled by the other… For these reasons, inequalities by race, gender, ethnicity, class, age, citizenship, educational level, and other apparently contradictory principles of differentiation form through similar social processes and are, to an important degree, organizationally interchangeable.

Tilly's basic argument is that institutions play a social role in distributing scarce resources—including access and prestige—and this distribution is made simpler by allocating resources on the basis of categorical pairs that are recognized in society at large (Tilly, 1998; Tomaskovic-Devey, Avent-Holt, Zimmer, & Harding, 2009). As a result, goods get distributed according to external categories, so that inequality and the categories themselves become mutually reinforcing, thus helping to explain the durability of categorical inequality.

The utility of this theory for our conceptual framework is that it suggests that for understanding poverty and livelihoods the details of institutions are less important than the effect they have on the conversion of incomes and resources into livelihood outcomes. To the extent that categorical inequality is ingrained in institutions, the effect is to align outcomes along the boundaries of socially-recognized categorical pairings. This leads to our final basic conceptual framework, shown in Figure 6, in which the 'institutions' node of Figure 5 has been replaced by the node 'Gender, etc.', which includes any relevant categorical pairs.

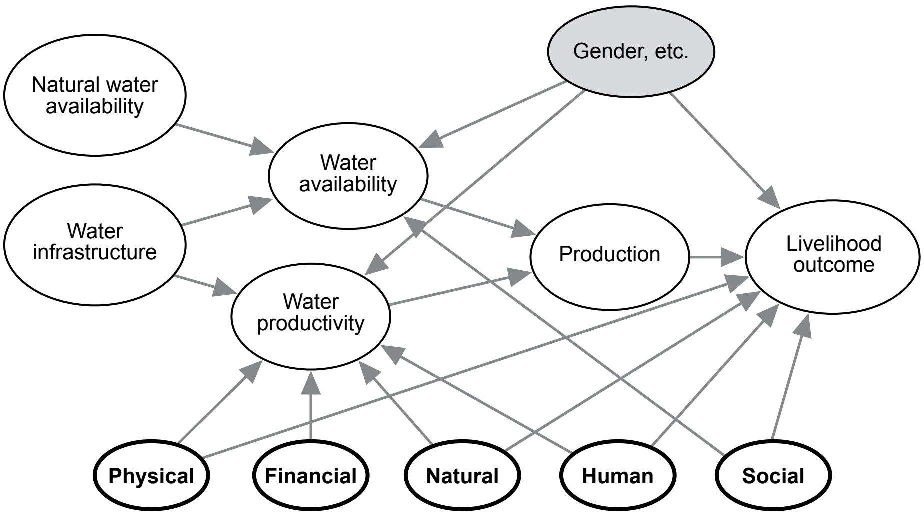

Figure 6: The basic conceptual framework

The key informant interview with a farmer in Khon Kaen, referred to here as 'KK', provided some interesting and somewhat unusual examples of categorical inequality. The farmers in Khon Kaen had been relocated by the government when their farms were flooded after construction of the Ubonrattana Dam (see Figure 1). Previously they had grown lowland rice, but in the uplands of Khon Kaen that was not possible. Making use of their knowledge of farming—their human capital—they cut down the forest that was there and planted cassava. However, because the nutrients in the soil were exhausted, they got only one year's harvest from the cassava. After that the village did very poorly, with many of the men turning to gambling and drinking. A local doctor tried to encourage the farmers to use integrated farming methods to revitalize the soil and increase their income, but the farmers were dismissive of the doctor's advice because he was not a farmer. That is, there existed, as in Tilly's framework, a categorical distinction across a socially recognized dividing line that made it difficult to introduce innovations. Eventually the doctor managed to persuade one of the women to try his techniques. She began using them on the farm and it was highly successful. It was not clear from the interview how much resistance she faced as a woman, but the situation was unusual as the men were not seriously working the land. In any case, KK is not limited by her gender now: she is running the farm and a training centre and is the village headman.

Variability, diversity, and uncertainty

In the next section (Section 3), the conceptual framework elaborated above will be transformed into a quantitative model, with each node potentially becoming a variable. Before doing so it is worth pausing for a moment to consider the links shown in Figure 6. In many models the values for variables and the relationships between them are sharp—that is, every variable has a single value at each moment in time. Examining the nodes, it is clear that many of them are better characterized as ranges, or distributions, of values—that is, as fuzzy variables—rather than as single values. There are three sources of fuzziness in the framework: variability, diversity, and uncertainty.

Variability

Variability is change over time. Models with single values can capture variability by having values change over time, and while that is an option for this model as well, when time-series data are unavailable it is useful to represent variability with fuzzy variables. For example, one source of variability in the framework in Figure 6 is natural water availability. Changes in rainfall and river flow are often monitored, so that data over time exist for this variable. Other variables in the framework in Figure 6 are typically informed from a study of a single site, and reflect both the long-term accumulation of adaptive responses to variability (Anderson, 2003; Csermely, 2006), as well as the specific responses to the water availability in that year (Bharwani *et al.*, 2008; Moench, 2005). The long-term structure can be thought of as responding to the long-term pattern, while the specific behaviour is a response to conditions at a particular moment in time.

Variability of natural water availability can be illustrated by patterns of rainfall over a period of time. The distinction between rainfall over time and as a distribution can be seen in Figure 7. In the figure, monthly precipitation data for the catchment for the Chi River in Northeast Thailand, 1995–2005, is shown both as a time series and as a distribution. The distribution shows a characteristic 'mixed gamma' distribution shape (Thom, 1968), in which there is a relatively high probability of a very low rainfall and, if there is rainfall, the distribution is skewed, with a long tail toward high values.

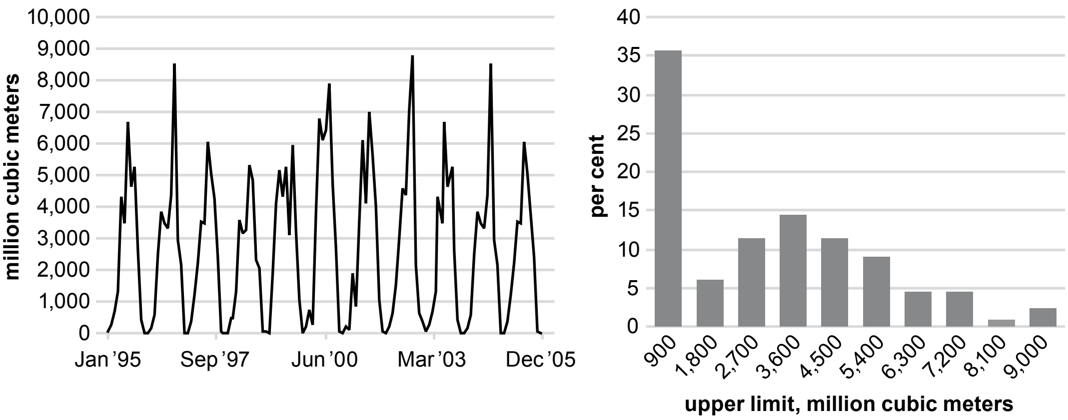

Figure 7: Rainfall in the Chi River catchment over time and as a distribution across monthly values 1995–2005

Where data are available it is best to try to elucidate both the long-term response to variable rainfall as well as specific responses in particularly wet or dry years. However, data often do not permit such an analysis, and in such cases distributions of time-varying variables can be used.

Diversity

In the course of responding to both externally and internally varying circumstances, communities and households develop adaptive responses (Csermely, 2006). This leads to a variety of potential responses that become embedded in the structure of the community or household. That is, variability in the environment drives diversity within the community or household. Communities, households, and societies have many characteristics of complex adaptive systems (CAS), defined as systems composed of a set of interacting components whose interactions may be complex and whose components are diverse or have a capacity for constructing adaptive behaviours (Norberg & Cumming, 2008). It has been found that diversity helps to make complex adaptive systems more resilient (Norberg, Wilson, Walker, & Ostrom, 2008). For this reason, the fuzziness of the system due to diversity is a key contributing factor to the resilience of communities and households, and it is important to capture it when representing the system as a model.

Uncertainty

The final source of fuzziness in the model is uncertainty. In theory, it is impossible to ever know everything about the state of a system (Monod, 1971). In practice, it is not necessary to resort to theory. To know everything relevant to the livelihoods of an agricultural community would be both impractical and intrusive. In the best of situations what is known is a large set of indicators of livelihood assets, strategies, and outcomes. Indicators are an imperfect window on reality, so that even when sharp values are available for indicators, the quantities that they indicate should in most cases be best expressed as fuzzy values (Ascough, Maier, Ravalico, & Strudley, 2008).

Applying the framework

The framework elaborated in this section is intended to be used as a guide, rather than as a prescription. At the same time, it is designed so that it can be translated relatively directly into a computer model. Subsequent sections of this report will describe how it can be used as the basis of a Bayesian network model. As an illustration of the non-prescriptive nature of the framework, the application in this report will depart from it in some ways in order to make the most effective use of the available data.

3 BAYESIAN BELIEF NETWORKS, LIVELIHOODS, AND NATURAL RESOURCES

As argued at the end of the previous section, there are reasons to use a fuzzy modelling approach for the framework in Figure 6 rather than a more standard modelling approach in which variables have well-defined, single values. For this reason and others the project documented in this report chose a Bayesian network (BN) modelling approach. In a BN, variables are probability distributions, such as the distribution of monthly rainfall in the right-hand side of Figure 7.

Bayesian networks have several appealing features that make them useful for policy modelling of livelihoods and natural resources (Cain, 2001; MERIT, 2005). While building BNs requires some modelling expertise, it is less demanding than other modelling approaches, permitting a broader use of the tool. Models based on BNs can be represented graphically—the network in Figure 6 was constructed using a Bayesian network software tool—and they can represent simple cause-effect relationships (Cain, 2001; Pearl, 2001). For these reasons, BNs can in many cases be understood by people who are not modellers, but who have some knowledge of the system being modelled. Bayesian networks can also capture qualitative information and incomplete information. Because of these appealing features, BNs have been used to study the impact of interventions on forest-derived livelihoods (Newton *et al*., 2006), on fisheries (Baran & Cain, 2001), and for water resources management (Cain, 2001; MERIT, 2005), among other applications.

Beyond these practical considerations, Bayesian methods in general, and BNs in particular, offer a conceptually coherent way to treat uncertainty (Ascough *et al*., 2008). Bayesian statistical methods can be contrasted with the conventional 'frequentist' approach to statistics. In the frequentist interpretation, the values in a particular experiment are conceptualized as a sample drawn from a theoretically infinitely-repeated set of experiments. This conceptual starting point has proven to be very fruitful, and has given rise to an elegant and useful body of theory and practice. However, few experiments fit the theoretical ideal, and nearly all policy-relevant analyses lie at the opposite extreme of the ideal, characterized as they are by sparse data, non-repeated experiments, and a fluctuating policy, economic, social, and natural environment. Bayesian statistical methods urge the analyst to make use of whatever additional information she has available to her to construct a 'prior' set of probabilities. The prior probability distribution is then updated by observation, in a coherent way, to produce a 'posterior' probability distribution. The essential aspects of the updating procedure are described below.

The general approach to assessment using the framework developed in Section 2 is to elicit the main components of the network in a participatory exercise, represent it as a Bayesian network model, and then use Bayesian statistical reasoning to interpret indicators—that is, the available data—as evidence of the state of the variables in the network. The network in Figure 6 is of the right form for a Bayesian network: the requirements are that each variable, or node, is linked by a directed arrow, and there are no closed loops: that is, it is not possible to return to the same node by following a trail of arrows from one node to the next.[1]

Bayesian networks

Bayesian networks are models that relate variables to one another using conditional probabilities (Jensen, 1996; Pearl, 2001). The variables themselves are probability distributions rather than fixed values; that is, as discussed above, they use fuzzy rather than sharp values.

A conditional probability is the probability of some event, *Y*, happening, given that some other event, *X*, has happened. This is written

$$P(Y|X) = \text{the probability of Y given X} \tag{1}$$

It is a standard result in probability that the total probability of *Y* occurring is equal to the sum of the conditional probabilities for each possible event *X*, multiplied the probability of *X* occurring, and summed over all *X*. That is,

$$P(Y)=\sum_X P(Y|X)P(X) \tag{2}$$

This result can be represented graphically in a Bayesian network. Suppose that we are interested in testing for a water contaminant with a test that is not perfectly accurate. In that case, the events *Y* that we might be interested in are 'test is positive' and 'test is negative'. The events *X* of interest might be, 'water is polluted' or 'water is not polluted'. This can be represented as a Bayesian network as shown in Figure 8.

1 This means that Bayesian networks cannot have feedback loops. However, it is possible to create Bayesian models with feedback loops by taking the output at one time step and using it as the input for the next time step. This creates a closed feedback loop with a delay.

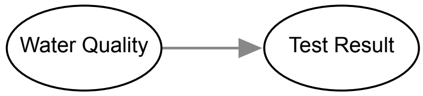

Figure 8: Example Bayesian network

The network in Figure 8 has a link from one variable, 'Water Quality', to another variable, 'Test Result'. The conditional probabilities (Equation 1) would answer the question, "What is the probability of a positive test, given that the water quality is actually poor?" This is the sort of information that comes out of using the test in trials, and in this way, the network explicitly captures uncertainty about the test. The set of conditional probabilities is called the conditional probability table, or CPT.

Bayes' rule[2]

The direct use of the water quality testing network allows us to infer how often we will get a positive test, *given* information about water quality. However, this is the reverse of what we actually want, which is to do a backwards inference: *given* a particular test result, as well as prior information about the probability of a polluted source, we wish to infer the probability that any particular source is polluted. This result can be calculated using a result derived from the mathematician the Reverend Thomas Bayes (Bayes, 1763). He started with the observation that the probability that events X and Y both occur, $P(X, Y)$, can be written in two equivalent ways:

$$P(X,Y) = P(Y|X)P(X) = P(X|Y)P(Y) \tag{3}$$

That is, the probability that both X and Y occur can be thought of as the probability that Y occurs, given X, multiplied by the probability that X occurs, or vice-versa. Rearranging this equation slightly gives Bayes' rule,

$$P(X|Y)=\frac{P(Y|X)}{P(Y)}P(X) \tag{4}$$

Combining this equation with Equation (2) gives a particularly useful form of Bayes' rule,

$$P(X|Y)=\frac{P(Y|X)}{\sum_{X'} P(Y|X')P(X')}P(X) \tag{5}$$

Equation (5) says that using our conditional probability for Y given X—in our example, the results of field testing the diagnostic water quality test—with a prior belief about the probability that X occurs—for example, the probability that any given water body is contaminated—we can estimate what we want to know, that is, the probability that the water body is contaminated given the result of the test.

To make the abstract formula in Equation (5) more concrete, suppose that water quality testing has been carried out using a slow but very precise method at 100 sites, and 10 sites have tested positive for the pollutant, with possible public health risks. From these pilot results, it is expected that of the 1,000 sites in the country, around 10 per cent, or 100, of them are contaminated. From a public health perspective it is necessary to establish with some degree of certainty which ones actually are polluted.

2 This text is adapted from the author's contribution to the Mekong Basin Focal Project final report (Kirby *et al*., 2008).

In this simple example, assuming that no readily observable factors indicate that one site is more likely to be polluted than another, public health concerns can only be fully addressed by testing all 1,000 sites. However, the same slow and precise method that was used to establish the proportion of sites that are polluted cannot be used to identify all of the sites where action needs to be taken because it would be prohibitively expensive to do so.

To address this problem, a promising new test for the pollutant is put under trial. The new test is quick and inexpensive, but not as accurate as the original test. To establish the reliability of the new test, a total of 100 trials are carried out on sites known to be polluted, and a further 100 trials are applied on sites that are known to not be polluted. The results (the conditional probability table, or CPT), are shown in Table 1.

Table 1: CPT for water quality test

Actual State	**Test Results**		
	Positive	*Negative*	*Total Tests*
Polluted	95	5	100
Not Polluted	10	90	100

From the results in Table 1, it is estimated that the false positive rate is 10 per cent, and the false negative rate is five per cent. Given the public health risk, it is felt that a higher false positive rate than false negative rate is a good property of the test, and it is prepared for use in the field. However, this conclusion does not use the *prior* information that around 10 per cent of sites are expected to be polluted. With Bayes' rule, this information can be used to refine the assessment of the test. Suppose that a positive result is obtained. What is the probability that the site is polluted? To answer this, we turn the question around, and ask, of all of the 1,000 sites, for how many can a positive test be expected? As shown in the table above, if the site is polluted, then the probability of a positive result is 95 per cent, while if the site is not polluted, then the probability is 10 per cent. But it is expected that 100 sites will be polluted, while 900 will not, so the number of positive tests is expected to be

$$\text{Number of positive tests} = 95\% \times 100 + 10\% \times 900 = 185 \tag{6}$$

This is nearly twice the number of polluted sites. The number of sites that are both polluted and that give a positive result is expected to be 95 per cent × 100, or 95. Therefore, the probability that a positive result actually indicates a polluted site is

$$\text{Probability that positive indicators means polluted} = 95/185 = 51\% \tag{7}$$

This is a much less encouraging result than the laboratory results for the quick and inexpensive test shown in the table seemed to suggest. Only about one-half of the time will a positive test actually correspond to the presence of the pollutant.

This result can be obtained more quickly using Bayesian network software. Using the GeNIe Bayesian network software, the problem can be set up as shown in Figure 9.[3]

3 GeNIe may be downloaded at no charge from http://genie.sis.pitt.edu/.

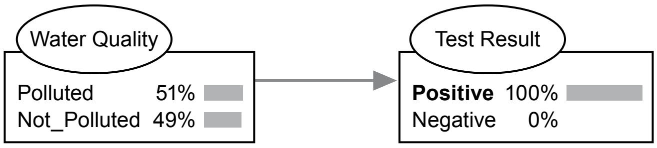

Figure 9: Water quality network in GeNIe

By setting evidence within GeNIe that the test result was positive, it is possible to infer that there is a roughly 50-50 chance that the site is polluted. This is the same result that we obtained in Equations (6) and (7). Note that the weakness of the link between the test and the actual water quality is due to the prior distribution—polluted water bodies are relatively rare (one in ten).

A Bayesian interpretation of indicators

In the water quality example given above, the variable of interest is the water quality, but the data available is the test result. That is, the test result is an indicator of the variable of interest. Typically, in a modelling exercise, the test result would be used as a proxy variable for water quality. However, Bayes' rule and the Bayesian network presentation of the problem suggests another way to use the indicator; that is, to view the indicator as indirect evidence for the underlying state of the system and provide a fuzzy, probabilistic, link between the two. The potential value of this approach can be seen in the example above, in which an apparently robust test (Table 1) is found to be only weakly connected to the variable of interest (Figure 9) within a Bayesian framework. Using the test result directly as a proxy for water quality would have implied a stronger relationship than actually exists. Thus, a Bayesian approach can help to guard against putting too much faith in indicators as proxies for the variables in a model or conceptual framework by reflecting the uncertainty inherent in indicators. This is a 'latent variable' method, in which some of the variables in a model—the latent variables—are only indirectly observed. [See Krishnakumar & Ballon (2008) for a non-Bayesian latent variable technique for determining household capabilities within a livelihoods framework.]

In studies of livelihoods, evidence is typically drawn from case studies or household survey data. For example, the level of education, assessed from a household survey, is commonly used as a proxy for human capital. By analogy with the water quality network (Figure 8), within a Bayesian framework, this relationship can be represented by drawing an arrow from the variable of interest to the indicator, as shown in Figure 10.

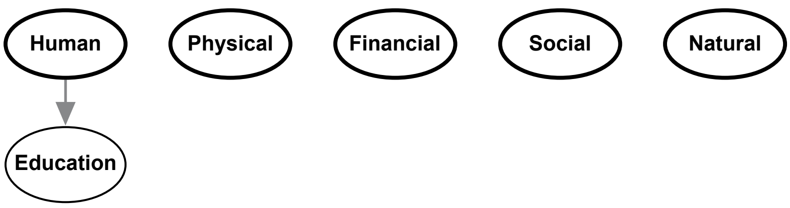

Figure 10: Linking livelihood assets to indicators

The reader might wonder why the arrow in Figure 10 goes from the asset category to the indicator or, more generally, from the latent variable to the observed variable. The reason is that it makes causal sense: the indicator takes on the value it does as a consequence of the value of the latent variable, and not the other way around, as in the water quality example above. Reflecting upon Figure 10, it becomes clear that while the level of education is evidence of human capital, it might also be evidence of financial capital and perhaps also of social capital. Moreover, the same capitals can provide evidence for multiple indicators. That is, there may be more than one link, as shown in Figure 11.

We argue that this is preferable to the standard approach of associating each indicator with a single category (Newton *et al*., 2006; Sullivan, 2002), as the restriction to a one-to-one relationship is unrealistic and artificial, driven more by the formalism of the indicator's construction than by the relationship of the indicator to the framework.

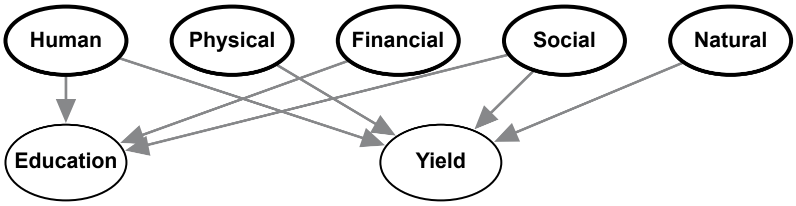

Figure 11: Linking multiple livelihoods to a single indicator

Data and elicitation

The network structure, as in Figure 11, is part of the specification of a Bayesian network. The network structure has a convenient and comprehensible graphical representation that can be used very effectively in stakeholder settings (Cain, 2001). The rest of the Bayesian network model is a set of conditional probability tables (CPTs), which express relationships between variables. The CPTs are more challenging to fill in than the network structure, both because it is difficult for people to think in terms of probabilities (O'Hagan *et al*., 2006) and because the CPTs can easily become rather large.

There are essentially two ways to fill in the CPTs: from data, and using expert elicitation. Using data is in most cases straightforward but tedious, and should generally be carried out without the stakeholders present. However, it is rare when considering livelihood strategies to have enough data to fully specify a network, and elicitation is almost always necessary. The more complex the model is, the more the modeller must rely on elicitation to fill in the CPTs. Figure 12 illustrates this trade-off schematically: complex models are shown as having a high proportion of elicitation relative to data, while simple models are informed more by data than elicitation. In nearly every case, a combination of elicitation and data is required, thus the line does not extend all the way to the axes.

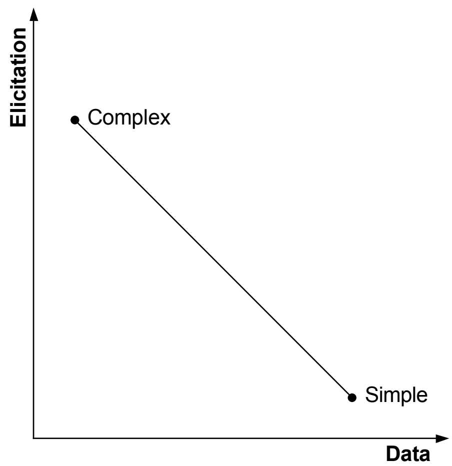

Figure 12: Trading off elicitation *vs*. observation in Bayesian networks

Expert elicitation should happen with stakeholders, but filling in the CPTs can be time-consuming and difficult, as the expert must estimate conditional probabilities for a large set of combinations—often hundreds or thousands—of events. Direct elicitation of probabilities is the most common approach in the literature, but it has many problems, because people have difficulty thinking in probabilistic terms (Fraser, Smith, & Smith, 1992; O'Hagan *et al*., 2006; Tversky & Kahneman, 1974). As an alternative to direct elicitation, it is often more useful to use a simplified probability table that has only a few parameters. A spreadsheet or special-purpose software tool then applies an algorithm that transforms the parameters into a CPT (Cain, 2001; Kemp-Benedict, 2008).

For the study described in this report, an elicitation exercise with experts was not possible, due to budget constraints. Instead, the model is primarily informed by data. We make use of an elicitation algorithm, but apply a simple, routine set of assumptions to generate an initial set of CPTs. Subsequently, the Bayesian network is trained with data, a process that refines the CPTs. The next section provides further details of the procedure.

4 MODEL-BUILDING PROCESS

As explained in previous sections, process described in this report results in a Bayesian network that is linked to indicators that provide evidence about the variables in the model. As with any Bayesian model, the variables themselves are probability distributions rather than fixed values, and the network has two components: a network structure and a set of conditional probability tables. In the next section the process is carried out using data for Si Sa Ket province in Northeast Thailand.

Eliciting the network structure

The first step in building a model for a particular area is to elicit the network itself. The conceptual framework developed in Section 2 forms the basis for the network structure. However, an elicitation should start with semi-structured interviews about challenges and strategies. The interviews aim to discover

1. what water resources are available and the ways in which people make use of them,
2. the assets that people have at their disposal and how they use them to support their livelihood strategies,
3. variability in water supply and how it affects what people do and the outcomes they experience,
4. the external influences—such as political influences and prices—that people respond to,
5. how categorical inequality affects people's ability to translate assets and outputs into livelihood outcomes,
6. how things have changed over time for themselves and the community,
7. what people think is necessary to have or to do in order to move out of poverty.

With the answers to these interview questions, the network can be created and then presented to the informants. The starting point is the Sustainable Livelihoods asset categories. The assets identified by informants (or available from surveys or other data) are then linked to the asset categories as evidence. The answers to the remaining questions help to fill in the rest of the network. A variety of standard case-study or household survey data may be used to provide evidence about the variables in the model, while the specific links will depend on the situation.

Eliciting the probability distributions

In the procedure described in this report, CPTs are generated indirectly using an algorithm that requires relatively few parameters compared to the number of independent values in the CPT. The algorithm is based on the 'likelihood' method described in Kemp-Benedict (2008), and is presented in this report through an example.

Consider the sub-network shown below in Figure 13. In the sub-network, the indicator 'rice_disposal' is treated as evidence of financial, social, and human capital. In the Si Sa Ket application presented in the next section, the rice_disposal indicator was collected in field interviews and had the possible values 'mostly own use', 'mostly sold', and 'no answer'. Across the survey data set, the frequency of responses was

- mostly own use, 57 per cent;
- mostly sold, 38 per cent;
- no answer, 5 per cent.

The states of the capitals—financial, social, and human—are not directly observable. Instead, the indicators provide indirect evidence for them. They are given three possible levels: low (L), medium (M), and high (H). The initial distributions are assigned in the ratios L : M : H :: 1 : 2 : 1. That is, across the sample, 25 per cent of households are considered to be 'low', 50 per cent 'medium', and 25 per cent 'high'. These assignments can be taken to define what is meant by 'low', 'medium', and 'high'.

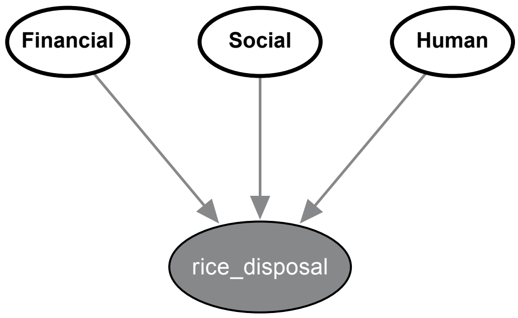

Figure 13: Sub-network for the variable 'rice disposal'

To calculate rice_disposal (the child node) from the capitals (the parent nodes), it is necessary to assign a probability for each of the possible child node values for each of the possible combinations of parent node values. The table that must be filled in is shown in Table 2. Each row must sum to 100 per cent, thus the number of independent values is (3 – 1) × 33 = 54. That is a lot of values to elicit, and for this reason it is useful to use an algorithm to generate the CPT from a few input assumptions.

Table 2: Required entries for a conditional probability table

Human	Social	Financial	Mostly own use	Mostly sold	No answer
L	L	L			
L	L	M			
L	L	H			
L	M	L			
L	M	M			
L	M	H			
L	H	L			
L	H	M			
L	H	H			

Human	Social	Financial	Mostly own use	Mostly sold	No answer
M	L	L			
M	L	M			
M	L	H			
M	M	L			
M	M	M			
M	M	H			
M	H	L			
M	H	M			
M	H	H			
H	L	L			
H	L	M			
H	L	H			
H	M	L			
H	M	M			
H	M	H			
H	H	L			
H	H	M			
H	H	H			

The procedure used for this project uses Bayesian statistical reasoning to generate the CPT. First, it is assumed that a prior distribution is available for the child node. The prior distribution is the probability distribution that is expected to apply in the absence of any further evidence, as given by the values of the parent nodes. In the example, the prior distribution is given by the survey data. That is, 'mostly own use' is expected to be true 57 per cent of the time for a given household, 'mostly sold' 38 per cent of the time, and 'no answer' (here interpreted as 'not applicable') five per cent of the time. The CPT is then calculated using Bayes' rule:

$$P(c \mid p_{1,} p_{2,} \ldots, p_N) \propto P(p_{1,} p_{2,} \ldots, p_N \mid c) P(c) \tag{8}$$

This is equivalent to Equations (4) and (5), because the denominator in Equations (4) and (5) is completely determined by the need for the probabilities to sum to 100 per cent; for this reason, it is sufficient to say that the left-hand side is proportional to ($\propto$) the right-hand side. In this proportionality, c represents a possible value of the child node, and $p_1, p_2, \ldots, p_N$ represent the possible values of the parent nodes. In the example shown in Figure 13, the number of parent nodes, N, is three. The probability distribution $P(c)$ is the prior probability, and the distribution $P(c|p_1, p_2, \ldots, p_N)$ is the conditional probability table.

The algorithm provides a way to build up the probability distribution $P(p_1, p_2, \ldots, p_N |c)$, also known as the likelihood $L(c|p_1, p_2, \ldots, p_N)$,

$$L(c \mid p_{1,} p_{2,} \ldots, p_N) \equiv P(p_{1,} p_{2,} \ldots, p_N \mid c) \tag{9}$$

The likelihood asks the question, suppose that you observe a particular value for the child node, *c*. What probability would you assign to different combinations of the parent nodes? While the question may sound equivalent to asking what the probability of *c* would be given the values of the parent nodes, that is not the case. The difference lies in the prior distribution, as shown in Equation (8). By focusing on the likelihood, rather than the probability, the elicitation procedure removes the difficult challenge of keeping the relative frequency of the occurrence of different values of the child node in mind. Instead, it keeps the focus on the more comprehensible elicitation question of how much influence the different parent nodes might have on the possible outcomes for the child node. In this way the procedure follows the recommendation of O'Hagan that statistical theory be used to ease the burden of elicitation (O'Hagan *et al*., 2006).

In addition to using Bayes' rule [Equation (8)], the algorithm uses a simplified expression for the likelihood. The likelihood is given as

$$L(c \mid p_1, p_2, \ldots, p_N) \propto b^{r_c\left(s_{p_1}+s_{p_2}+\cdots+s_{p_N}\right)} \tag{10}$$

With this formulation, the elicitation procedure is accomplished by providing the following information:

1. the base, *b*,
2. a weighting factor for each value of the child node, r_c,
3. a weighting factor for each value of the parent nodes, s_{p_1}, s_{p_s}, …, s_{p_N}.

For the example in Figure 13 and Table 2, the number of elicited values is 1 + 3 + 3 × 3 = 13. This is many fewer than the 54 independent probability values in the CPT. Moreover, the requested values can make more sense to the participants in an elicitation exercise than can the raw probabilities. The elicitation procedure is essentially asking for influence weights rather than probabilities.

To assist in using the algorithm, the Stockholm Environment Institute developed a software tool, the Bayes Table Generator.[4] A screenshot of the tool is shown in Figure 14 with values for the example in Figure 13. As seen in the screenshot, the base is set to two, and the child value 'no answer' or 'NA' is given a weight of zero. The values 'mostly own use' and 'mostly sold' are given the values -1 and 1, respectively. For each of the parent nodes, the weights are -1 for low (L), 0 for medium (M), and 1 for high (H)—as they all have the same set of values, each of the parent nodes is thought to be equally significant. Low values for the capitals are associated with more own use, because 'mostly own use' and 'L' are both negative, while the positive value for 'mostly sold' and 'H' reflect an assumption that higher levels of the capitals are associated with market production.

The table shown in the bottom panel of the screenshot in Figure 14 can be displayed either as numbers or as a colour map (as shown in the figure), with the darker colours indicating a higher probability. The influence of the prior distribution can be clearly seen in the colour map, where 'mostly own use' in most cases has a higher probability than 'mostly sold', as in the prior. However, when the capitals are high, then 'mostly sold' is more probable than 'mostly own use'.

The Bayes Table Generator and the algorithm it employs is an elicitation tool, and it is intended to be used interactively, rather than mechanically. During an elicitation activity, the probability distribution generated using the Bayes Table Generator tool can be reviewed, and the input assumptions adjusted, if the probabilities seem unreasonable. Also, once the generated CPT is entered into GeNIe or other Bayesian network software, individual probabilities can be adjusted by hand. Finally, after the CPT generated by the Bayes Table Generator is entered into the Bayesian network, the CPTs can subsequently be modified by training the network with data.

4 The Bayes Table Generator program, along with two example files, can be downloaded from ftp://sei-us.org/BayesTableGenerator. Note that while the software is usable and robust enough for technical users in practical applications, it is not production quality and is likely to be frustrating for non-technical users.

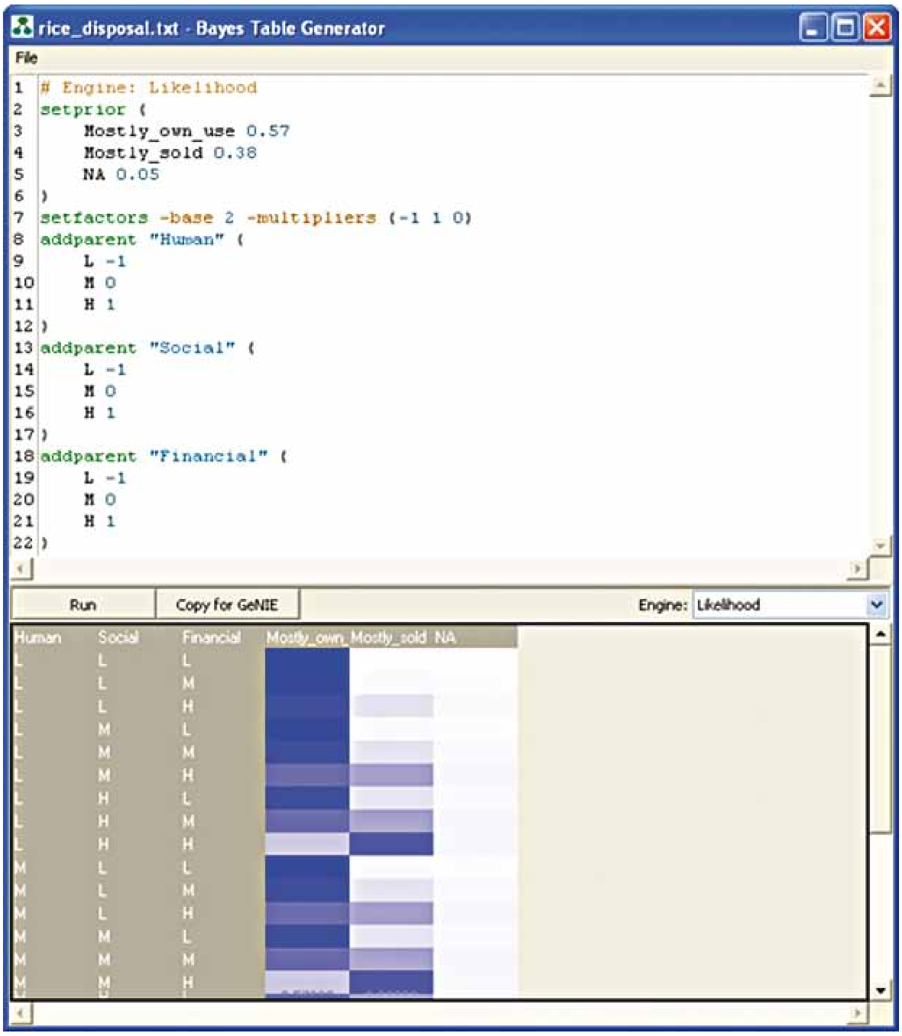

Figure 14: The Bayes Table Generator for creating CPTs

5 NETWORK FOR SI SA KET PROVINCE

The framework and procedures described above were applied to a data set collected for five villages in Si Sa Ket province, Northeast Thailand (SEI, 2008). The data were collected under the Mekong Basin Focal Project (BFP), part of the CGIAR's Challenge Program on Water and Food (CPWF). Si Sa Ket province was identified as a water poverty hot spot by the BFP team. The area, located in the Chi-Mun river basin, is mostly agricultural and poor, and experiences long dry periods (as shown in Figure 7, Section 2).

Model construction

For the analysis described in this report, data for all five villages was pooled into a single data set. This provided a larger training data set, but at the expense of losing detail about the real differences that exist between the villages. Of the full set of data, a subset was used to provide evidence for the network. Variables were chosen based on their relevance to the framework shown in Figure 6. Of the set of relevant variables, some were excluded either because they were redundant with other variables or because they displayed insufficient variability between households. The network structure used for the study is shown in Figure 15.

The network in Figure 15 contains a simplified version of the conceptual framework (Figure 6). The livelihood assets are collected into an aggregate node, 'Non ag water assets', while assets related to agricultural water use are reflected by the node 'Water infrastructure'. Categorical variables are used as evidence for a node, 'Categorical access'. Survey data are used as evidence for the variables in the model.

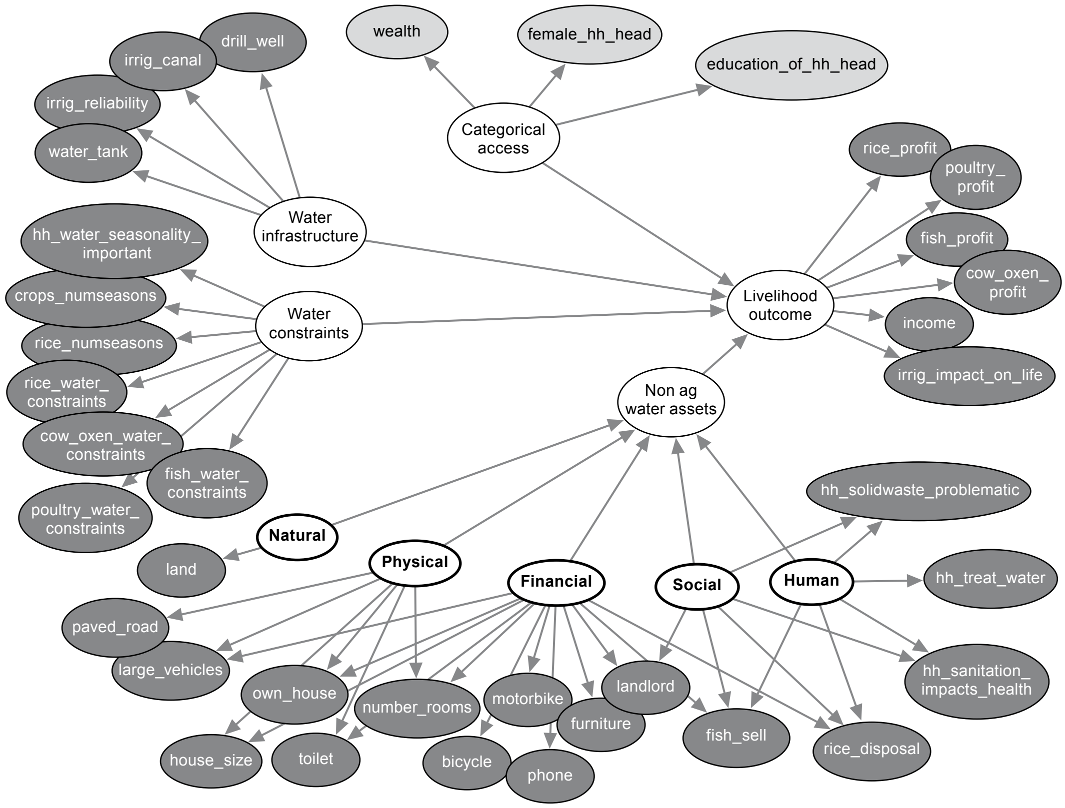

Figure 15: Bayesian network structure for Si Sa Ket province

The CPTs in the model were constructed using a uniform approach that made use of the elicitation algorithm described in the previous section and illustrated in Figure 14.

1. For each survey variable, the prior distribution was set to the distribution of responses across the sample.
2. For model variables, possible values were low (L), medium (M), and high (H), and the prior distribution was set to L: 25 per cent, M: 50 per cent, H: 25 per cent.
3. Weights for possible child node values were set within the range -1 to 1, and were separated by equal steps. (Thus, for example, a set of five responses running from a poor outcome to a good one would be given weights of -1.0, -0.5, 0.0, 0.5, 1.0.) Values recorded as 'NA' were given a weight of zero.

Following the assignment of the initial CPTs, the network was trained with data in GeNIe. GeNIe uses the expectation maximization algorithm (Bilmes, 1998), and offers two options—randomisation of the CPTs before running the algorithm and (if not randomised) assigning some weight to elicited values. For the application described in this report, the CPTs are given initial values, so the CPTs were not randomised. However, the initial values were given no weight, so that the data would have the greatest possible influence in affecting the model parameters.

Running the model

A Bayesian network model is run by first setting evidence for certain nodes—that is, specifying their values—and then calculating the resulting probabilities for the other nodes, given the evidence. The Bayesian network software calculates the values for nodes without evidence either directly, using conditional probability calculus [Equation (2)] or indirectly, using Bayesian inference [Equation (5)].

The model for Si Sa Ket province was built and run in GeNIe. GeNIe allows groups of nodes to be gathered together into 'sub-models', which can be used to clean up the display. Also, selected nodes can be shown as bar graphs, so that the probability distributions can be seen clearly. These two features were used to reorganize the network as shown in Figure 16. Most of the nodes from Figure 15 are hidden in the sub-models 'Water infrastructure and constraints', 'Other assets', and 'Evidence of livelihood outcomes'. Certain key variables are highlighted to explore their responses to setting evidence.

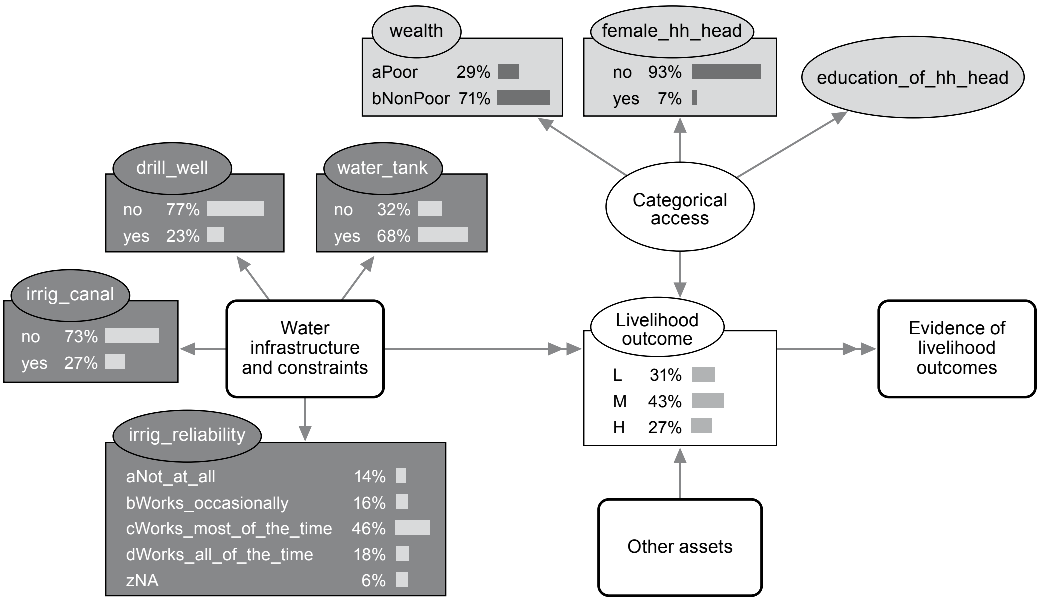

Figure 16: Network for Si Sa Ket province reorganized for interactive use

As an example of running the model, we ask the following question:

> What water infrastructure makes the most difference to poor households in achieving positive livelihood outcomes?

To answer the question, we carry out several runs of the model. In each of the runs, we set as evidence that 'wealth' is 'aPoor'. Between runs we set the evidence for different types of infrastructure and view the resulting implications for livelihood outcomes. The results are shown in Table 3. According to the model, and as seen in the table, there is little comparative benefit to having access to a water tank compared to a drill well or reliable irrigation. If irrigation is not reliable, then it is a less desirable option than either a water tank or a drill well, although it is still better than having no water infrastructure at all. These results are qualitatively consistent with the findings of the field study, in which large scale irrigation projects were frequently unsuccessful, because they were unreliable, while locally-originated water supply schemes (including irrigation schemes) were more successful (SEI, 2008). The results suggest that installing relatively inexpensive water tanks could be a cost-effective way to improve the livelihoods of poor households that do not already have a water tank installed, although it could also mean that poor households with better livelihood outcomes can afford water tanks—further investigation would be required to confirm the direction of causality.

Table 3: Results from running the model

		Irrigation		Livelihood outcome (per cent)		
Water tank	Drill well	Access	Reliability	Low	Medium	High
no	no	no	—	45	39	15
yes	no	no	—	37	42	20
no	yes	no	—	38	42	20
no	no	yes	occasional	41	42	17
no	no	yes	all of the time	37	43	21

Reflections on the model

The model for Si Sa Ket province illustrates the major features of the Sustainable Livelihoods-based Bayesian modelling approach that this report advocates. Some of the more important points to note are the following:

- the Bayesian network uses 'fuzzy' variables that reflect the variability, diversity, and uncertainty in livelihood systems;
- survey data are used relatively directly, as the qualitative nature of the survey response categories works well within the Bayesian framework;
- Bayesian statistical reasoning allows indicators to be used as evidence for model variables, which this report has argued is more theoretically coherent than using indicators directly as proxies for model variables;
- the model results are expressed in terms of an underlying conceptual model that is based on the Sustainable Livelihoods framework.

By the end of the project, the model was relatively easy to develop compared to other modelling approaches. The most time-consuming part of the modelling effort was in transforming the survey data into the tabular format required by GeNIe. Also, the data were first explored using non-Bayesian statistical techniques within the R statistical software system. The techniques used at the end of the project, and described in this report, were the product of extensive development over the course of the project, during which a number of models of varying complexity were created using a variety of techniques. Had the final modelling approach been available at the start of the study it would probably have been feasible to have a full elicitation workshop. As the final modelling approach was not available, and required considerable theoretical and practical development, an elicitation workshop was not possible. For this reason, the model produced for this project should be used cautiously, as it has not been evaluated by knowledgeable experts and has not benefited from expert elicitation. This report recommends expert input throughout the model development process.

The reader might be struck by the seemingly weak response to the infrastructure interventions shown in Table 3. This is a general feature of Bayesian models—the interposition of fuzzy relationships between probabilistic variables leads to relatively weak responses. The authors of this report see this as an advantage of the Bayesian approach, rather than a failing. In reality, communities often do respond weakly to interventions in the short term. Rather than having an immediate, dramatic effect, many interventions instead change the relative chances of a positive *vs*. a negative outcome in any given year. Over time, the greater frequency of positive outcomes leads gradually to improved livelihoods. Indeed, part of the reason why the response is weakened is because of the diverse livelihood strategies that households put in place to buffer shocks: adaptive diversity blunts the effects of policy interventions.

6 CONCLUSION

This report presents a conceptual framework and modelling technique to assess the possible effect of interventions on water-related poverty. The approach takes into account the complexity of livelihood systems and the variety of strategies that they represent. This is accomplished by using a Bayesian network model to represent a community, household, or group's livelihood strategies. The Bayesian network explicitly captures variability, diversity, and uncertainty by using probabilities, rather than fixed numbers, for variables.

While the approach is based on the well-established Sustainable Livelihoods framework, the report recommends that elicitation proceed through semi-structured interviews. The results of the interviews provide insight that informs a Bayesian network. The network should reflect the information collected during the interviews while also making use of the conceptual framework described in this report.

Bayesian networks have a convenient graphical representation, and work well with qualitative and categorical data, such as is collected in field surveys. Also, the network itself is relatively straightforward to construct using Bayesian network software such as the GeNIe software used for this project. The challenge lies in constructing the conditional probability tables (CPTs) that link the variables in the model. Over the course of the project a convenient elicitation algorithm and accompanying software tool were developed that make this task considerably simpler.

Near the end of the project, an interim version of the approach was presented at three universities in Northeast Thailand: Khon Kaen University, Mahasarakham University, and Ubon Ratchathani University. The project team asked university researchers for critical feedback on the approach. For the most part the feedback was positive, with the researchers supporting the goals of the project to make more direct use of survey data and to incorporate qualitative data and elicited input. Particular attention was paid to the possibility of using Bayesian learning to include the insights of local actors into a model in a statistically consistent way.

There were also critical comments. Researchers at Ubon Ratchathani University, who have used Companion Modelling techniques (Barretau & others, 2003) to construct agent-based models in a field setting were very doubtful that the Bayesian model could be used in a similar manner. The authors of this report think that this is a valid point. Indeed, others who have carried out Bayesian modelling exercise emphasize that the method is not easy, and requires some expertise (Cain, 2001; Lynam, de Jong, Sheil, Kusumanto, & Evans, 2007). Other concerns were raised about the subjective elements in the modelling exercise. This is understandable, but a degree of subjectivity is unavoidable. Policy-relevant models of complex socio-ecological systems are plagued by uncertainty about present conditions, interrelationships between system components, and future influences. The authors of this report believe that it is better to be explicit about the subjective elements within such models, and the approach we advocate is to build models in part through expert elicitation.

From the experiences of the project team with the methods described in this report, as well as the feedback from researchers at universities in Northeast Thailand, the authors draw the following conclusions:

- Bayesian network models offer a theoretically coherent way to deal with incomplete information, qualitative data, and indicators.
- The Sustainable Livelihoods framework is a useful, asset-based poverty framework that can be used to inform a modelling exercise, especially when using a fuzzy modelling approach.
- The analytically challenging task of representing institutions can be simplified by focusing on the effects of institutions on livelihood outcomes, especially as institutions reflect socially-recognized categorical inequality. This approach also introduces a capabilities-based view of poverty into the conceptual framework that complements the asset-based view of the Sustainable Livelihoods framework.

- Bayesian networks consist of two components: the network structure and the conditional probability tables (CPTs). Of the two, the network structure is more amenable to development in a participatory workshop. However, elicitation techniques and algorithms can be developed that greatly simplify the task of creating CPTs, making it more feasible to construct CPTs in the course of a workshop.
- Bayesian network models are suitable for participatory engagement with technically knowledgeable people who are not modellers. However, Bayesian networks are not suitable for engagement with informants in the field, where agent-based models and role-playing simulations appear to be preferable. It is possible that with a specially-designed interface a Bayesian modelling system could be developed for use with informants in the field.
- Bayesian network models can be used to relatively quickly construct a working model for assessing the impact of interventions that makes use of survey data, statistical data sets, and expert elicitation. For this reason, Bayesian network models appear to be well suited for early assessment of the possible impacts of interventions.
- Going beyond the applications explored in this report, Bayesian techniques may be useful for combining the insights from data, policy makers, technical experts, and farmers in a statistically consistent way. An interactive Bayesian learning system in which the relative weights of data and the assessments of different stakeholders can be easily changed, could be a powerful tool for representing farmer interests. An interface targeted to the intended users could be developed that would allow people to interact comfortably with the model. To ensure transparency, it would also be possible for interested users to access the underlying model if they wish.

REFERENCES

Anderson, J. R. (2003). Risk in rural development: challenges for managers and policy makers. *Agricultural Systems, 75*(2–3), 161–197.

Ascough, J. C., II, Maier, H. R., Ravalico, J. K., & Strudley, M. W. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological Modelling, 219*(3–4), 383–399.

Baran, E., & Cain, J. (2001). Ecological and Modelling Approach to Flood-Fish Relationships in the Mekong River Basin. In H. L. Koh & Y. Aby Hasan (Eds.), *National Workshop on Ecological and Environmental Modelling* (pp. 20–27). Universiti Sains Malaysia, Penang, Malaysia.

Barretau, O., & others. (2003). Our Companion Modelling Approach. *Journal of Artificial Societies and Social Simulation, 6*(2).

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances, by the late Rev. Mr. Bayes, Communicated by Mr. Price, in a latter to John Canton, M.A. and F.R.S. *Philosophical Transactions of the Royal Society of London, 53*, 370–418.

Bebbington, A. (1999). Capitals and Capabilities: A Framework for Analyzing Peasant Viability, Rural Livelihoods and Poverty. *World Development, 27*(12), 2021–2044.

Bharwani, S. (2006). Understanding Complex Behavior and Decision Making Using Ethnographic Knowledge Elicitation Tools (KnETs). *Social Science Computer Review, 24*(1), 78–105.

Bharwani, S., Downing, T., Haase, D., Pahl-Wostl, C., Taylor, A., Shale, M., *et al*. (2008). *Dynamic Vulnerability: Learning from NeWater Case Studies* (NeWater Deliverable No. D 2.1.2). Oxford: Stockholm Environment Institute.

Bilmes, J. A. (1998). *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models* (No. TR-97-021). Berkeley, CA: International Computer Science Institute and Univ. Calif. Berkeley.

Black, M., & Hall, A. (2004). Pro-Poor Water Governance. In *Water and Poverty: The Themes* (pp. 11–20): Asian Development Bank.

Cain, J. (2001). *Planning improvements in natural resources management*. Wallingford, UK: Institute of Hydrology.

Carloni, A. S., & Crowley, E. (2005). *Rapid Guide for Missions: Analysing Local Institutions and Livelihoods* (Institutions for Rural Development No. 1). Rome: Food and Agriculture Organization of the UN.

Carter, M. R., & Barrett, C. B. (2006). The Economics of Poverty Traps and Persistent Poverty: An Asset-based Approach. *Journal of Development Studies, 42*(2), 178–199.

Cook, S., & Gichuki, F. (2006). *Analyzing Water Poverty: Water, Agriculture and Poverty in Basins* (Basin Focal Project Working Paper No. 3): CGIAR Challenge Program on Water and Food.

Csermely, P. (2006). *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks*. Berlin: Springer.

DFID. (1999). *Sustainable Livelihoods Guidance Sheets*. London: UK Department for International Development.

Easterly, W. (2006). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: The Penguin Press.

Fraser, J. M., Smith, P. J., & Smith, J. W. (1992). A catalog of errors. *International Journal of Man-Machine Studies, 37*(3), 265–307.

Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. New York: Springer-Verlag New York, Inc.

Kemp-Benedict, E. (2008). *Elicitation Techniques for Bayesian Network Models* (Working Paper No. WP-US-0804). Somerville, MA: Stockholm Environment Institute – US.

Kirby, M., Krittasudthacheewa, C., Mainuddin, M., Kemp-Benedict, E., Swartz, C., & de la Rosa, E. (2008). *Mekong Basin Focal Project: Final Report*: CSIRO and Stockholm Environment Institute.

Krishnakumar, J., & Ballon, P. (2008). Estimating Basic Capabilities: A Structural Equation Model Applied to Bolivia. *World Development, 36*(6), 992–1010.

Lynam, T., de Jong, W., Sheil, D., Kusumanto, T., & Evans, K. (2007). A Review of Tools for Incorporating Community Knowledge, Preferences, and Values into Decision Making in Natural Resources Management. *Ecology and Society, 12*(1), 5. [online] URL: http://www.ecologyandsociety.org/vol12/iss11/art15/.

MERIT. (2005). *Guidelines for the Use of Bayesian Networks as a Participatory Tool for Water Resource Management: Based on the Results of the MERIT Project*. Wallingford, UK: Centre for Ecology and Hydrology for the Management of the Environment and Resources using Integrated Techniques (MERIT) Project.

Moench, M. (2005). *Water, Climatic Variability and Livelihood Resilience: Concepts, Field Insights and Policy Implications* (The Resilience and Freshwater Initiative Policy Paper No. II). Stockholm: Swedish Water House.

Molden, D., Oweis, T. Y., Steduto, P., Kijne, J. W., Hanjra, M. A., & Bindraban, P. S. (2007). Pathways for Increasing Agricultural Water Productivity. In *Water for Food, Water for Life: Comprehensive Assessment of Water Management in Agriculture*. London: Earthscan.

Monod, J. (1971). *Chance and Necessity; An Essay on the Natural Philosophy of Modern Biology. Translated from the French by Austryn Wainhouse*. New York: Knopf.

Newton, A. C., Marhsall, E., Schreckenberg, K., Golicher, D., te Velde, D. W., Edouard, F., *et al*. (2006). Use of a Bayesian Belief Network to Predict the Impacts of Commercializing Non-timber Forest Products on Livelihoods. *Ecology and Society, 11*(2), 24 [online] http://www.ecologyandsociety.org/vol11/iss22/art24.

Nicol, A. (2000). *Adapting a sustainable livelihoods approach to water projects: Implications for policy and practice* (No. 133): Overseas Development Institute (ODI).

Norberg, J., & Cumming, G. S. (2008). Introduction. In J. Norberg & G. S. Cumming (Eds.), *Complexity Theory for a Sustainable Future*. New York: Columbia University Press.

Norberg, J., Wilson, J., Walker, B., & Ostrom, E. (2008). Diversity and Resilience of Social-Ecological Systems. In J. Norberg & G. S. Cumming (Eds.), *Complexity Theory for a Sustainable Future*. New York: Columbia University Press.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., *et al*. (2006). *Uncertain Judgements: Eliciting Experts*' Probabilities. Chichester, UK: John Wiley & Sons, Ltd.

Pearl, J. (2001). Causal Inference in Statistics: A Gentle Introduction, *Computing Science and Statistics, Proceedings of Interface* '*01 33*.

Saravanan, V. S. (2008). A systems approach to unravel complex water management institutions. *Ecological Complexity, 5*(3), 202–215.

Scoones, I. (1998). *Sustainable Rural Livelihoods: A Framework for Analysis* (IDS Working Paper No. 72). Brighton, UK: Institute for Development Studies.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, Connecticut: Yale University Press.

SEI. (2008). *Water Poverty and Livelihoods, Si Sa Ket Province, Northeastern Thailand* (Background report for the Mekong Basin Focal Project). Bangkok, Thailand: Stockholm Environment Institute.

Sen, A. (1999). *Development as Freedom*. New York: Anchor Books.

Sullivan, C. (2002). Calculating a Water Poverty Index. *World Development, 30*(7), 1195–1210.

Sullivan, C., & Meigh, J. (2003). Considering the Water Poverty Index in the Context of Poverty Alleviation. *Water Policy, 5*, 513–528.

Thom, H. C. S. (1968). Approximate Convoluton of the Gamma and Mixed Gamma Distributions. *Monthly Weather Review, 96*(12), 883–886.

Tilly, C. (1998). *Durable Inequality*. Berkeley, California: University of California Press.

Tomaskovic-Devey, D., Avent-Holt, D., Zimmer, C., & Harding, S. (2009). The categorical generation of organizational inequality: A comparative test of Tilly's durable inequality. *Research in Social Stratification and Mobility, 27*(3), 128–142.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131.

SEI - Africa
Institute of Resource Assessment
University of Dar es Salaam
P.O. Box 35097, Dar es Salaam
Tanzania
Tel: +255-(0)766079061

SEI - Asia
15th Floor Witthyakit Building
254 Chulalongkorn University
Chulalongkorn Soi 64
Phyathai Road Pathumwan
Bangkok 10330
Thailand
Tel: +(66) 22514415

SEI - Oxford
Suite 193
266 Banbury Road,
Oxford, OX2 7DL
UK
Tel: +44 1865 426316

SEI - Stockholm
Kräftriket 2B
SE -106 91 Stockholm
Sweden
Tel: +46 8 674 7070

SEI - Tallinn
Lai 34, Box 160
EE -10502, Tallinn
Estonia
Tel: +372 6 276 100

SEI - U.S.
11 Curtis Avenue
Somerville, MA 02144
USA
Tel: +1 617 627-3786

SEI - York
University of York
Heslington
York YO10 5DD
UK
Tel: +44 1904 43 2897

The Stockholm Environment Institute

SEI is an independent, international research institute. It has been engaged in environment and development issues at local, national, regional and global policy levels for more than a quarter of a century. SEI supports decision making for sustainable development by bridging science and policy.

sei-international.org