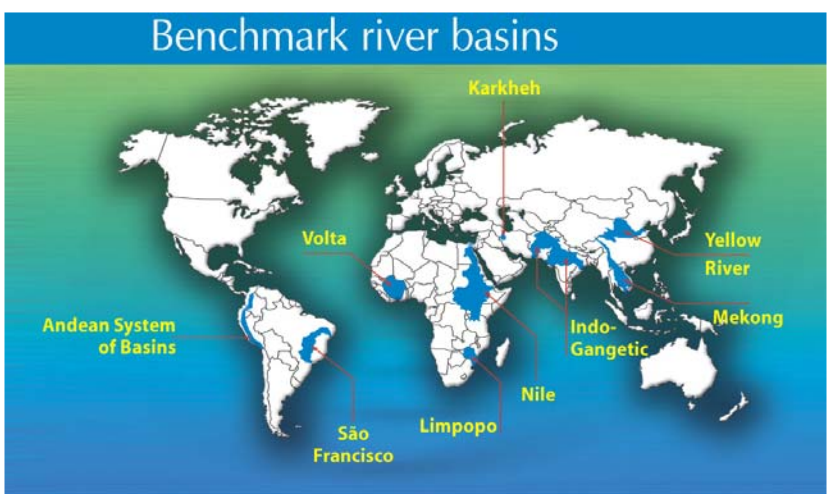
WATER POVERTY MAPPING IN THE VOLTA BASIN: Looking for linkages between water and poverty

Basin Focal Project

Workshop Report

Accra, Ghana 3-8 March, 2007.

WORKING WITH PARTNERS TO ENHANCE AGRICULTURAL WATER PRODUCTIVITY SUSTAINABLY IN BENCHMARK RIVER BASINS



DISCLAIMER

This is an internal report on a Basin Focal Project workshop, **Water Poverty Mapping in the Volta Basin: Looking for linkages between water and poverty** held in Accra, Ghana 3-8 March, 2007. It may in due course become a working paper and be published formally by the Challenge Program on Water and Food. This report contains less than fully polished material. Some of the works may not be properly referenced. The purpose is to share information and concepts early to support development of more rigorous methods of analysis.

The findings, interpretations, and conclusions expressed here are those of the author(s) and do not necessarily reflect the views of the Challenge Program.

Comments and additional inputs that could contribute to improving the quality of this work are highly welcomed.

They should be sent to the corresponding author:

Simon Cook – s.cook@cgiar.org

This document will be updated as more data and information becomes available. For the latest update please visit our website

www.waterforfood.org

WATER POVERTY MAPPING IN THE VOLTA BASIN: Looking for linkages between water and poverty

Workshop report

Accra, Ghana 3-8March, 2007.

Simon Cook[1], Jorge Rubiano[2], Caroline Sullivan[3] , Winston Andah[4], Felix Ashante[5], Jeremy Wallace[6], Isabelle Terrasson[7], Aude Nikiema[8] , Eric Kemp-Benedict[9], Ignacio Tourino[10], and Gerald A. B. Yiran[11]

---

[1] CPWF-BFP, IWMI, Colombo, Sri Lanka. Corresponding author. s.cook@cgiar.org

[2] UNAL, Palmira and CIAT, Cali, Colombia. jerubianom@palmira.unal.edu.co

[3] Oxford University Centre for the Environment, Oxford, UK caroline.sullivan@ouce.ox.ac.uk

[4] Ghana Water Research Institute, CSIR, Accra, Ghana. weiandah@africaonline.com.gh

[5] Institute of Statistical, Social and Economic Research, University of Ghana, Accra, Ghana. fasante@ug.edu.gh

[6] Commonwealth Scientific and Industrial Research Organisation CSIRO, Perth, Australia. Jeremy.Wallace@csiro.au

[7] Institut De Recherche Pour Le Developpement IRD, Montpellier, France. terrasso@msem.univ-montp2.fr

[8] CPWF-BFP, IRD, Burkina Faso. nikiaude@yahoo.fr

[9] Stockholm Environment Institute SEI, Boston, USA. erickb@sei-us.org

[10] IWMI, Accra, Ghana. nacho.taurino@yahoo.fr

[11] Ghana Water Research Institute, CSIR, Accra, Ghana. Yiranab@yahoo.com

Executive Summary

The purpose of the workshop was to explore a method of analysing the linkages between poverty and agricultural water management in the Volta Basin, using the statistical power of Bayesian Networks (BNs). BNs have been used in a wide range of diagnostic problems in 'real-life' systems, that is, those that are complex and vague. They offer the promise of an ability to identify causal relationships that define water-related poverty. A further important advantage of BNs is that the algorithm is 'expert-friendly' and hence it is amenable to a process of joint discovery amongst analysts and domain specialists. A third advantage is that BNs are tolerant of uncertainties in the content and structure of analytical models, which is a critical attribute when dealing with the mixed data that are typical for large river basins. A key objective of the workshop was to explore if and how BNs could be coupled to the concept of the Water Poverty Index (WPI). This is important because it would enable users to benefit from the progressive accumulation of knowledge that is implicit in the WPI while analyzing water poverty in specific, concrete terms.

The workshop progressed from a description of the state of knowledge about poverty, agriculture and water, through the development of the WPI, tutorials on BNs and their prior application to large-scale analysis in Ecuador and Australia. This progressed to analysis of poverty data from Ghana and Burkina Faso, leading to an initial phase of BN analysis, discussion, interpretation and follow-up modelling using GENIE software. The final stage of mapping BNs was not completed because of technical difficulties with data preparation. Similarly, the combination of models from Ghana and Burkina Faso into a single assessment of poverty for the Volta was completed subsequent to the workshop.

The workshop accomplished much in terms of the analytical method and the process of using it. The BN methods proved robust and intuitive to experts. They produced models that were easy for specialists from Ghana and Burkina Faso to interpret, These models demonstrated a method of progressive improvement in the causal associations between poverty measures and physical and social attributes. With more time, additional input data could have been obtained to enable even more detailed models. This is a process of continual learning, using the WPI as a conceptual framework, with understanding of relationships provided by BNs.

Two main conclusions from the data seem to be: *first* that relationships between measurements of water management and poverty indeed are discernible, but that many complexities become evident as a result of the way people organize themselves around natural resources; and *second* that while water poverty does not exist as a strictly measurable attribute, through modelling it can be determined in relation to those factors that are measurable, and hence can be used to explain the effect of factors such as land ownership, water productivity and institutions.

The next steps are (1) to clarify key messages for stakeholders; (2) to complete the trial mapping of BNs as a guide to interpretation of WPI and (3) when additional data are available, re-work the BNs.

Introduction: Water, agriculture and poverty

A major challenge to addressing poverty throughout the world is both to locate where it is most concentrated and to identify its causes. Though empirical evidence is scarce on the exact nature of the relationship between water and poverty, it is widely recognised that water is both an essential part of livelihood systems, and an important component of agricultural production. People's largest use of water is for agriculture so that a key issue is to develop ways to increase agriculatural water productivity. To address this specifically, The Challenge Programme on Water and Food addresses this issue specifically with the key objective of "keeping the level of water use constant, while doubling food output by 2015" (CPWF 2002)

Water, agriculture and people's well-being are three major aspects of complex and highly variable rural livelihood systems. We are interested in water and agriculture inasmuch as they not only influence people's livelihoods, but especially how they can be modified to improve well-being. We know that people, both individually and collectively, adopt highly intricate strategies in agricultural systems to use water in ways that are often difficult to assess. The important issue is to analyze these strategies to arrive at the relation between water, agriculture and poverty.

Analyzing the link between water and poverty

In common with other complex systems, analysis of agricultural systems faces a difficult compromise between completeness and accuracy. On the one hand, analysis may use a complete model that attempts to include all factors the analysts perceive to be important. However, in choosing the model to use, this methodology is likely to pre-define the outcomes. This approach is typified by the Water Poverty Index (WPI) and its variates (Sullivan *et al.* 2001), where all factors believed to be important are included in the model. The relationship between various components of the model (i.e. the model structure) is imposed by the model architect in a way that is independent of the data. Assessment is knowledge-driven rather than evidence-driven so that causality between factors can only be inferred.

The second approach is typified by regression analysis of Farrow *et al.* (2005). Here, the analysis makes no attempt to include all aspects known to be important and focuses instead on data-driven evidence of the associations between poverty and a range of other attributes. While the results are specific, the model is incomplete, since only those factors for which data are available are analysed. No judgement is made about the overall structure of the model so that factors that might be considered important by the WPI method are omitted.

What we sought was an analytical method that incorporates the advantages of both approaches. The analytical model should be flexible enough to accommodate information that comes from a wide range of sources, such as in the WPI, yet be sensitive enough to enable causality to be established between specific factors. This should be a progressive process of model development, in which analysis of data occurs in parallel with knowledge-building.

To achieve this objective we proposed to use Bayesian Networks (BNs). These have been used in many complex, practical problems to determine causality amongst multiple, inter-related attributes (see Heckerman et al, 1996). Increasingly, BN techniques are being used to deliver knowledge-based systems to solve real world problems. Belief networks, in which related techniques are used to deal with reasoning under uncertainty, use Bayesian methodology and are particularly useful for diagnostic applications (see, for example, http://www.aiai.ed.ac.uk/links/bn.html).

BN methods provide "a formalism for reasoning under conditions of uncertainty, with degrees of belief coded as numerical parameters, which are then combined according to rules of probability theory". Bayesian networks are visualized as diagrams that organize knowledge as key variables that are mapped according to the cause-and-effect relationships amongst them. The degree to which one variable is likely to affect the other is determined by set of basic algorithms known as Bayes' theorem, and the whole network taken to represent what is known about a particular system.

A major advantage of BN is the ability to incorporate expert knowledge explicitly in the model. Most people find BNs easy to construct and interpret (Hackerman *et al.* 1996). (Pearl, 1986) shows that while some risks of conditional dependency can occur for ill-constructed models, human reasoning can normally understand the so-called triangular relationships in which one variable may in reality be expressing dependence through, rather than with, another. This conveys important advantages to the use of the results from BNs in a learning process. BNs can accommodate a range of data types, and can be data-driven or mediated by expert knowledge.

The purpose of the exercise undertaken at the workshop was to combine the strengths of both approaches: that is to use the inclusive, knowledge-based structure of the WPI with data-driven analysis of BNs. We used the structure of the WPI as a basis to understand the relationships between variables, but allowed the analysis to be driven by data from Ghana and Burkina Faso to develop progressively detailed insights into the associations of water with poverty.

Linking Bayesian approaches to theoretical frameworks

The purpose of analysis of water poverty within an area is twofold:

1) To infer the causes of poverty through comparison with other data. Here, we are particularly interested in the effects of measurable agricultural system attributes on poverty;

2) From the model created in (1), to estimate the distribution of water-associated poverty for places where it is not observed directly.

Bayes' Theorem was published over 200 years ago and many books and papers explain how it is derived and used. For further details, see the reading list (last section). Bayes' Theorem states simply that the probability that a particular event will occur can be estimated if some other event with which it has a known association has taken place. It thus uses prior knowledge about events. A simple example might be that of the probability of getting a ticket for any traffic violation, not just speeding. But we know that if we exceed the speed limit on the highway, there is a strong chance of being ticketed. Moreover, the probability that we shall get a ticket increases the more we speed. So, the chance of getting a ticket depends on the speed at which we drive even though if we do not speed, there is still a chance we might get a ticket for some other violation, a parking offence, perhaps. Thus

IF (Speed) THEN (Ticket)

Clearly this statement is not always true (we might not get caught), but is probably true in many, if not most, circumstances. And the stronger the relationship (the higher the speed, the more likely we are to get a ticket), the truer the statement becomes.

We can apply the same logic to the relation between water and poverty. Take the statement:

Water scarcity causes poverty

Data presented by Farrow *et al* (2003) for Ecuador (reworked by here in a Bayesian format), shows that indeed poverty is deepest in those areas that are most affected by drought. The rule that results from this analysis may be written:

IF (Water scarcity) THEN (Poverty)

The probability of poverty occurring may then be written:

$$p(POV \mid Drought) = \frac{p(POV, Drought)}{p(Drought)}$$

$$p(Drought \mid POV) = \frac{p(POV, Drought)}{p(POV)}$$

The statement is conditioned by the data. The stronger the association of poverty with drought (right-hand side terms), the more definitive statements can be about the probability of poverty, given that drought has occurred (LHS).

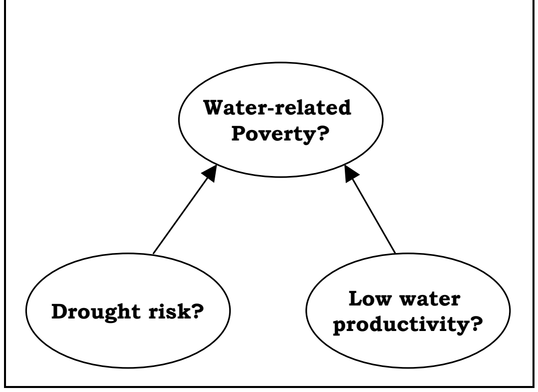
However, this is an oversimplification of the dynamics of poverty and unlikely to explain very much. Castillo *et al* (2007)[12] state that people's water endowment is less influential than their ability to use it. Hence we move to a more complex model that includes both the effects of drought (D) and water productivity (WP):

$$p(POV \mid D, WP) = \frac{p(POV, D, WP)}{p(D, WP)}$$

The three attributes POV, WP and D can be shown as nodes in a simple 3-point network, linked by arcs. The degree of causality between them is quantified from statements such as those above, which comprise a BN. Each variable added to the network model increases its

[12] Check the reference: Comprehensive Assessment

Figure 1. A 3-node BN



inclusiveness and complexity but the basic logic remains that pairs, triplets and so on are linked to the degree that the data or knowledge supports their association. The association can be positive (e.g. poverty is more often found in droughty areas) or negative (e.g. poverty rarely occurs where rainfall is well distributed).

Applying Bayesian Networks to the Water Poverty Index

One of the objectives of the workshop was to evaluate the links between water and poverty, using the structure of the WPI as the underlying framework of analysis. This enables the problem to be evaluated in a consistent way and has potential for generic application of the approach. The components of the WPI (Figure 2) highlight the point that, to some extent at least, its WPI's strengths are also its weaknesses. It is with this in mind that more exploration of the links between the components was considered worthwhile.

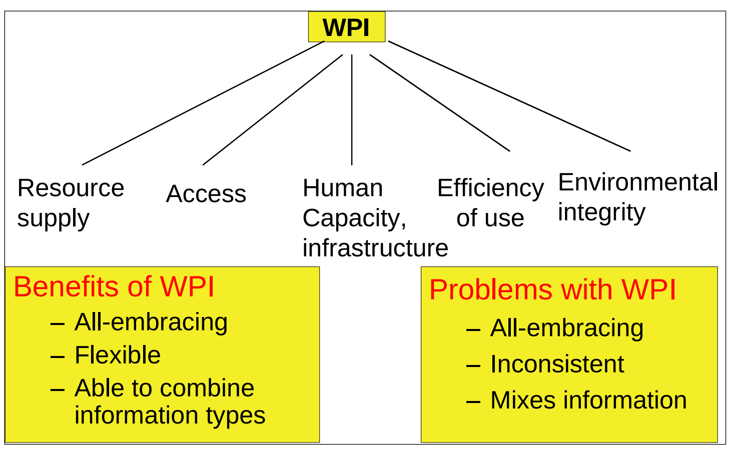


Figure 2. Characteristics and problems of the Water Poverty Index (WPI).

Example: Mapping water poverty from a case study in Ecuador

Farrow *et al.* (2005) provided detailed regression anlaysis of poverty data from Ecuador. The high resolution of the data enabled comparison against data within the original socio-economic data-set, but also against additional biophysical data that were extracted from other geographical surfaces, such as climate, land suitability for cropping, distance to markets, and elevation.

This dataset was re-analyzed to produce a simple BN model. The model shows the relation between the severity of food poverty and three variables related to the agricultural system that were shown to be influential:

- mn_dry: Mean number of consecutive dry months (months);
- PorIndig: Percentage of the population classed as indigenous (%); and
- AgrWF: Percentage of the workforce in agriculture (%).

Mn_dry represents an important resource factor (WPI factor 1). Another variable that might be considered important could be gross water availability (m^3/cap/yr). Availability of irrigation provision was important, but is not included here because we believed that its linkage with poverty was more complex we could model in this exercise.

PorIndig, the % of the population classified as indigenous, is representative of a range of poorly-defined institutional and resource factors that influence well-being. In the WPI it is included as one factor of Capacity.

AgrWF is another Capacity factor. Here it characterizes the reduction of poverty that is associated with salaried agricultural employment .

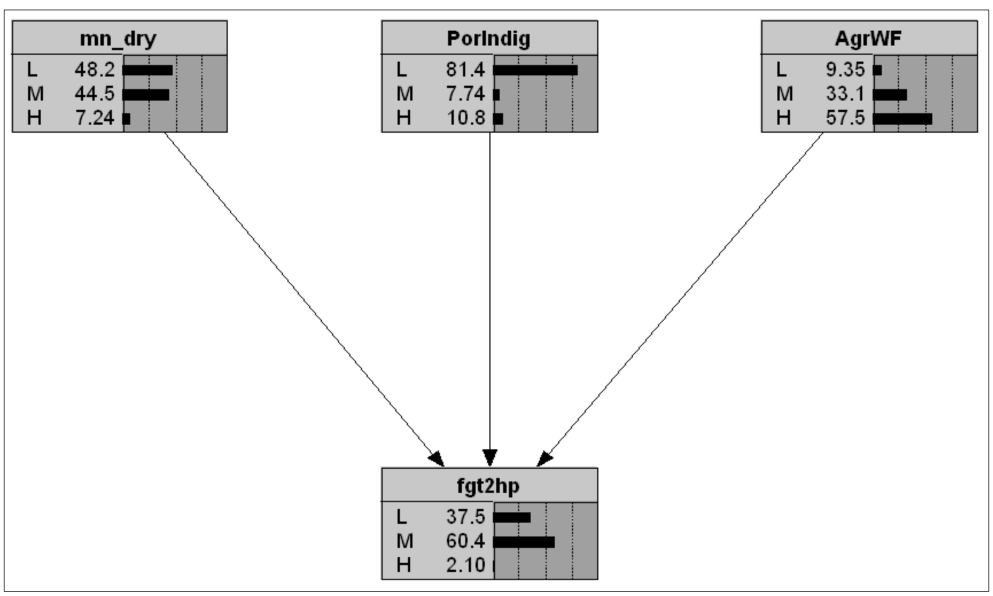


Figure 3. BN with prior probabilities.

Maps of these variables were then combined according to BN to produce a map of 'water food poverty', that is, the poverty that appears to be related to the above three variables of drought, % indigenous workforce and percentage of workforce in agriculture. This used the

Expector mapping method (Corner *et al.,* 1996)[13]. Map 1 shows the distribution of poverty related to these four factors. The least-food poor are those in the blue areas in and immediately around Quito, the capital city, in the more prosperous agricultural areas around the Gulf of Guayaquil and around the provincial centres of Manta and Portoviejo. Some poverty is related to drought, but clearly not always as poverty is high in Amazonia, where rainfall is high, but where food poverty is related limitations of the agricultural system other than drought.

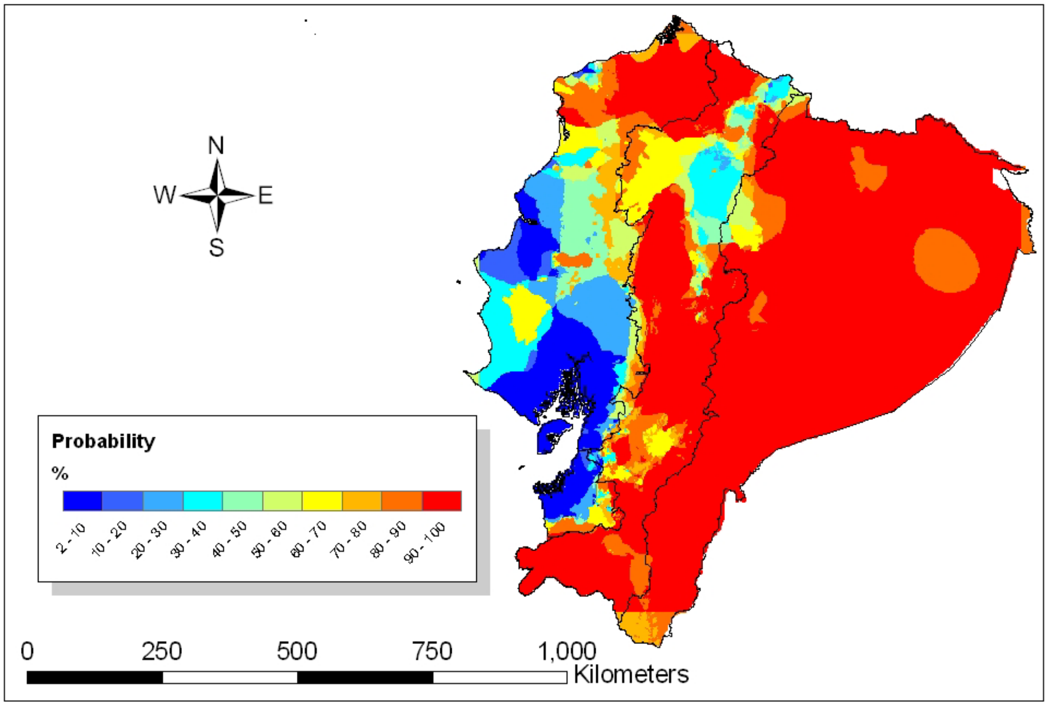


Figure 4. Probabilities that the severity of food poverty using the higher threshold (variable fgt_2h) exceeds 15 percent in Ecuador. The predictions were made using Expector with training variables mn_dry, mn_ap, agrWF and sal_agr as defined in the text.

WORKSHOP CASE STUDY

Building a BN for the Volta

We constructed a BN analysis for Ghana and Burkina Faso, which included the following steps.

Before the workshop, we:

Step 1. Identified basic problems of poverty, water and agriculture through consultation with stakeholders,. Assembled as much background knowledge as possible to help understand the results of subsequent analysis.

[13] Bonham-Carter (1994) provides a similar method of mapping that is downloadable.

Step 2. Assembled country data using available survey data. Specialists from Ghana and Burkina Faso have access to large survey data-sets. Biophysical data were supplemented from alternative sources.

Step 3. Carried out exploratory data analysis. A tedious but necessary process of ensuring that data are in a format suitable for analysis, and that obvious errors/outliers are identified before undertaking further analysis.

At the workshop, we then:

Step 4. Constructed preliminary BNs for discussion and analysis. An essential part of the knowledge development process was to ensure that domain experts were engaged fully in the modelling. The initial round of BN modelling highlighted some causal associations that demanded more detailed analysis.

Step 5. Identify where more data were needed so that the BNs might be refined and the results interpreted in consultation with experts. Progressive development of BNs should include additional data, identified after the initial analysis. For example, further explanation of WPr in grain-growing areas could be provided by re-examining the climate and soil data to provide more accurate assessment of expected yields.

Step 6. Mapping the outcomes of the BN model. [Where possible]. The intention was to use EXPECTOR software to map the principal BNs for Ghana and Burkina Faso. Preparation of input GRID data sets proved too difficult to complete during the workshop.

Step 7. Use the analysis and maps to describe causation –where possible – and as a basis for discussion about specific problems, and opportunities for change.

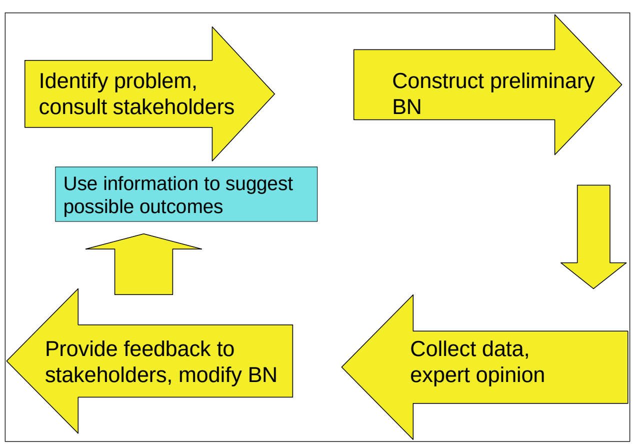


Figure 5. Building a Bayesian Network

Step 1. Identify the Problem. Assemble Background Information

The Volta River Basin

The Volta river basin is shared between Ghana and Burkina Faso,with additional small areas in Togo, Cote d'Ivoire, Benin, and Mali [14] (http://www.fao.org/docrep/W4347E/w4347e0u.htm). The basin covers about 400,000 km^2. The basin covers about 75% of the total area of Ghana, and the country has dammed the river to create Volta lake and to generate electricity. The lake is a primary resource for inland fisheries and for transportation in Ghana. Rainfall varies from about 500mm in the north of the basin to 1200 mm in the south. In the north rainfall has a unimodal distribution with an increasingly un reliable rainy season May-September. In the south rainfall has a bimodal distribution with long rains April – July and short rains September-November. The basin is characterised by poor soils and primarily supports savannah vegetation.

In the Ghanaian part of the Volta basin, 80% of the population is rural, with 20% urban. Poverty is widespread, and its distribution in Ghana is shown in Figure 7. All the regions shown in Figure 1 (except central and western regions) fall within the Volta Basin.

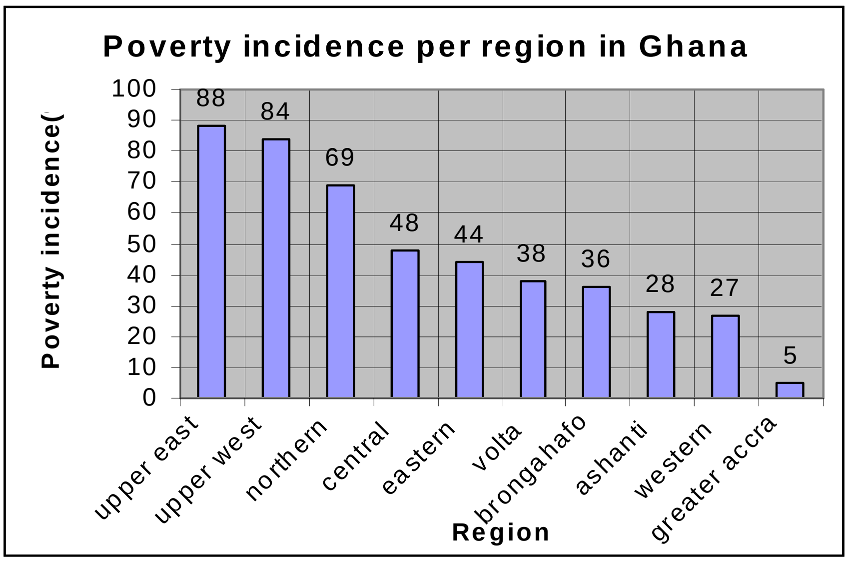


Figure 7. Poverty distribution in Ghana

Four main causes of poverty have been identified in the Ghana part of the Volta basin (Assante 2006, Assante, 2007). These are:

- Type of available economic activities (fishing, agriculture);
- Degree of water insecurity (poor access, health impacts, loss of labour);
- Environmental impact of upstream catchment activities (dams, farming practices); and
- Water related diseases (malaria, guinea worm, etc.)

In Burkina Faso, according to the last census (1996), 84% of the population live in rural areas. In that country, poverty is primarily a rural phenomenon, with 52% of rural people living below the poverty line in 2003 (fixed at 82 672 CFA, around US$160). The problem is

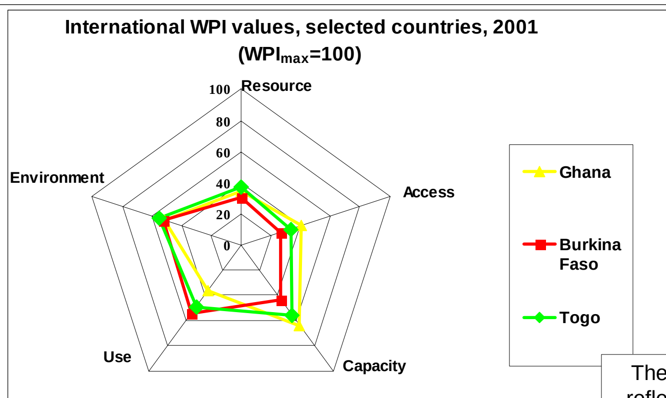
[14] http://www.fao.org/docrep/W4347E/w4347e0u.htm

worst in the central area of the country where access to primary social services (education, health, sanitary equipment) is poor. In addition, agriculture based on cereals is characterized by low productivity, largely due to poor quality soils. Migration from the central region to western and eastern regions has been important.

Although access to water has been improved through an increase in the number of boreholes and pumps, the number of rural households having access to safe water from bore holes, pumps or piped water taps is only 53%, compared to 83.3% in urban area. Amongst the most extreme rural poor only 51% have access to safe water, while amongst the least poor, 73% have safe water.

Primary school enrolment is about 52% in rural areas, but the adult literacy rate is only 26%. Despite the increases in health facilities across the country, infectious diseases are still the main cause of death amongst poor people.

The WPI in its non-BN form provides a useful assessment at national level (Figure 8).



	Resource	Access	Capacity	Use	Environ ment	WPI
Ghana	34.61	40.60	63.33	36.23	51.88	45.33
Burkina Faso	30.69	26.76	42.77	54.55	52.60	41.47
Togo	36.96	33.13	55.48	49.00	55.20	45.95

These figures reflect national data only, and not just the part of each country in the basin

Figure 8. National level application of the WPI framework in the Volta Basin.

To apply this framework to the data for the Volta basin, a number of specific variables had to be selected from the data. At the workshop there was much discussion on this topic. The variables that were suggested from within the available data for the Volta basin are shown in Table 1.

Table 1. Possible variables used to model the Water Poverty Index in the Volta basin.

Resource Lt per capita per day	**Access**	**Capacity**	**Use**	**Environment**
Rainfall	% pop with access to improved water supply	Adult literacy	Cereal yield	Land degradation
Groundwater	% pop with access to sanitation	% unemployed	Maize yield	Population density
Water quality	boreholes	% underemployed	Water productivity of cereals	Livestock density
Surface water in lakes etc?	Number of small dams (volume per capita?)	% landless	Water productivity of maize	Erosion?
Variability of rainfall	Access to irrigation?	% underweight children	Water for livestock?	NDVI[1] data?
Seasonal variation	Piped water	Size of land holdings	Industrial use?	
	Gender?	Wealth holdings		
	Proximity to river or lake?	Number of cattle per capita		

[1]Normalised difference vegetation index, which is a simple numerical indicator used to analyze remote sensing measurements to assess whether the observed target contains live green vegetation.

Step 2. Data exploration and representation of the linkages between water and poverty

In order to analyse the linkages between water and poverty, data from the Volta basin were examined, using the structure of the Water Poverty Index as a conceptual framework. Within the Volta basin, a variety of data sources were consulted from both Ghana and Burkina Faso. These data were generated from government surveys in both countries, including:

- *Core Welfare Indicators Questionnaire Survey* (CWIQ, 2003). Report of the Ghana Statistical Service (GSS). The data from this survey are from a two-stage national sample of households aimed at generating welfare indicators at the district level. The first stage was a random, systematic sample of 27 enumeration areas drawn from each district. In the second stage, all households within each selected enumeration area were listed, and 15 households were selected systematically from each, yielding a total of 49,005 households nationwide in 121 districts of Ghana. Out of the 121 districts, 62 fall within the Volta Basin.
- *Ghana Census Based Poverty Map, District and Sub District Levels* (2005). This report was commissioned by the World Bank and UK Department for International Development, and provides detailed information on the distribution of poverty across the districts of Ghana. These data have been used as a basis of a number of the calculations we carried out at the workshop.

- *Ghana Living Standards Survey, 4th round* (GLSS 4, 1998/99). This report has been used to provide a number of variables relative to poverty in the Volta basin.
- *GSS Housing and population census* (GSS ISSER and University of Ghana, 2000).
- *Enquêtes Burkinabè sur les Conditions de Vie des Ménages* (INSD, 2003).
- *La pauvreté au Burkina Faso* (INSD, 2003).

In addition to these pre-existing datasets, other data were used from projects carried out for the Challenge Progarm itself. Information on Ghana is provided in the report *Analysis of Water Related Poverty in the Volta Basin of Ghana* (Asante 2007), while for Burkino Faso it is provided from the CP report *Pauvreté et pauvreté hydrique au Burkina Faso* (Nikiema and Dipama 2007).

Step 3. Exploratory Data Analaysis

Data were manipulated in Excel and the S+ statistical packages until they were in a format suitable for input to the BN software NETICA and GeNIe[15].

Data were also generated from map layers for variables not represented in the original data sets. Coordinates were generated for the centroid of each administative polygon and were used to extract the data value for climate, crop productivity, water productivity, soil or terrain variables, where these did not exist in the original database. Exploratory analysis, mainly visual inspection of distributions, simple correlations and outliers was used to help determine sensible class limits for subsequent BN analysis (both software packages require that the data be divided into a limited number of discrete classes).

Step 4. Construct preliminary BNs for discussion/ analysis.

Figure 9 below shows a simple BN for data from Ghana. Po is a poverty headcount measure (from Coulombe 2005); NBDRY_MONT is a crude measure of drought. We use it to illustrate the process of developing understanding. The implication from this preliminary BN is that poverty is caused by drought, and that drought also exacerbates food need. However, food need does not appear to relate to poverty, nor landlessness. Finally, the most unusual observation from this preliminary network suggests that a strong relationship exists between poverty and landlessness. Closer inspection of probability distributions, however, suggests that areas with most poor people are also those with least proportion of landless. The situation seems to be more complex than this simple explanation.

[15] NETICA, widely used Bayesian Network software, is avaiable from http://www.norsys.com/. GeNIe is a graphical interface to the SMILE Bayesian inference engine. http://genie.sis.pitt.edu/about.html.

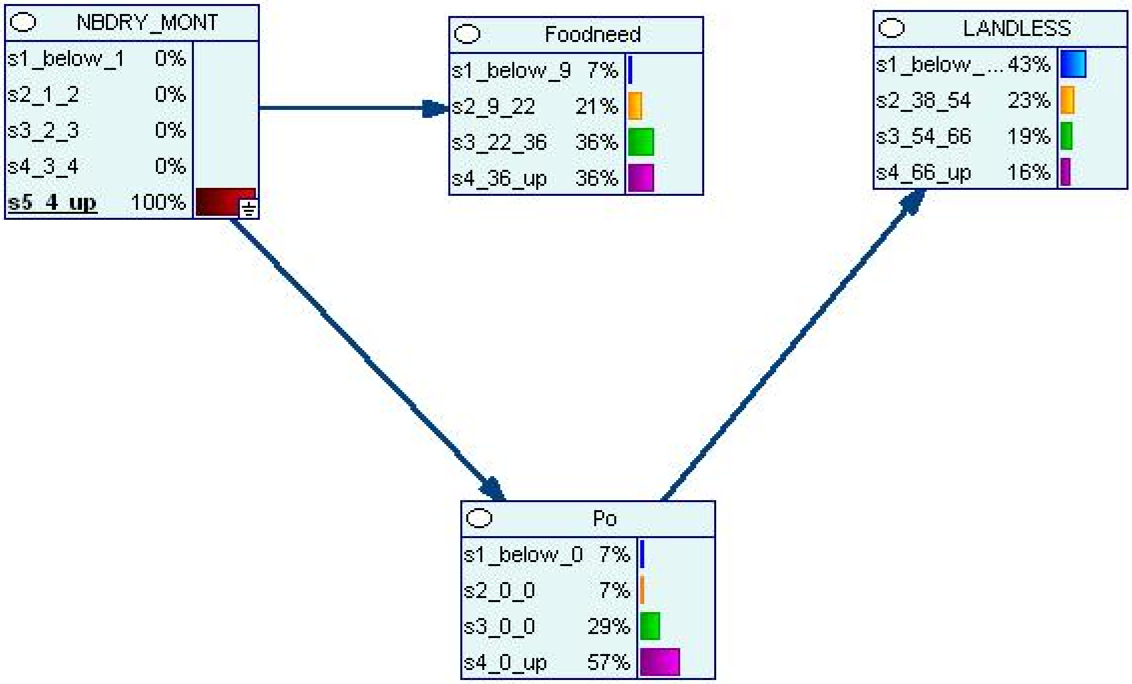


Figure 9. Preliminary Bayesian Network to describe the relation of poverty with potential causal variables in Ghana.

Each box, or node, represents an attribute that is associated, to the order and degree indicated by the arrows, with other attributes. Each attribute is discretised into classes, that can be interrogated as conditional probabilities (e.g. the p(Foodneed<9|ndrymonths>4)). . Other variables were sought from the data to help explore the relationship of agricultural factors with poverty, the list of data included from the surveys is listed in the Appendix.

Step 5. Collect more data, refine BNs and interpret results with experts.

Associations between variables become clearer as the network is grown by adding more variables. For example, the data from household surveys contains only those factors included on the questionnaire. To add other variable, we sought additional data for the same locations from GIS layers in the same way that we obtained data for drought. These variables included rainfall, access to nearest town, water productivity (kg/m^3) and crop productivity (t/ha).

BN for poverty in Ghana.

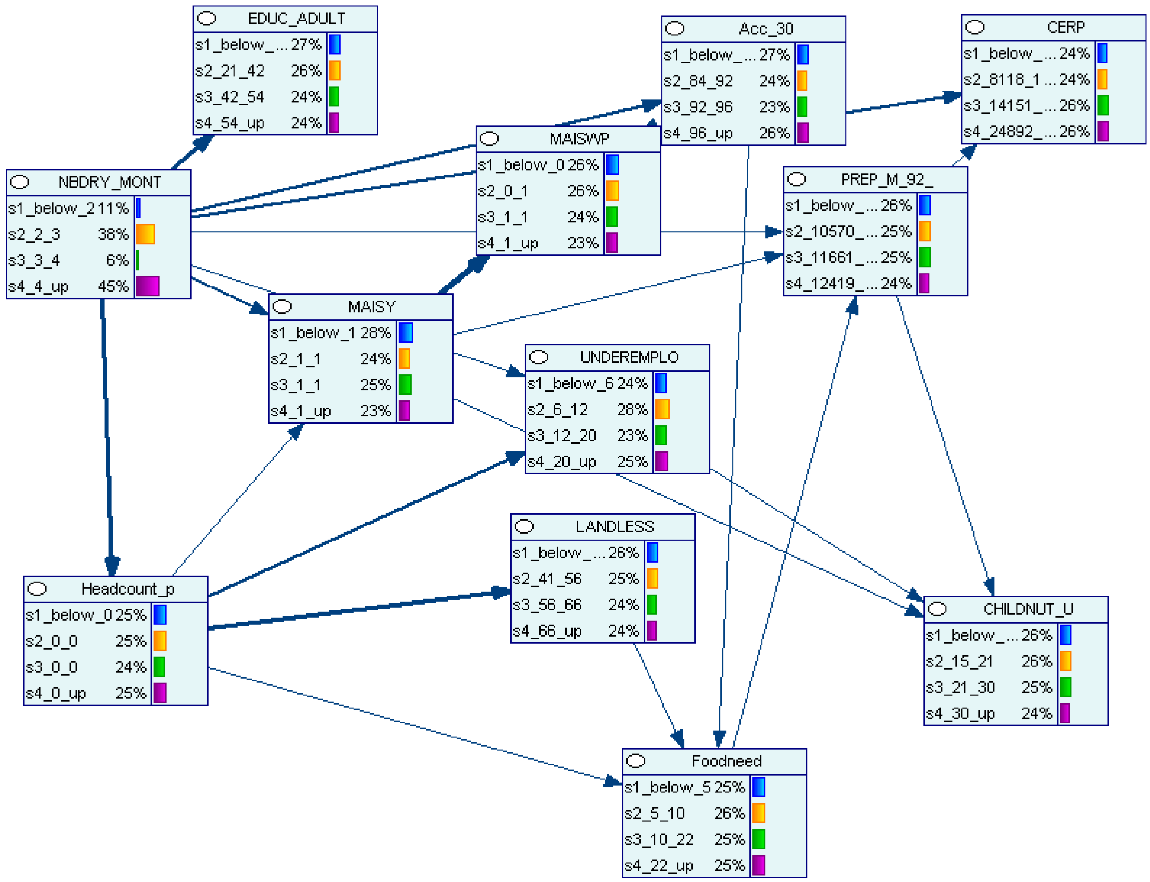


Figure 10. Augmented Bayesian Network to describe the relation of poverty with potential causal variables in Ghana. The thick arrows have stronger relations than the thin ones.

Figure 10 shows an expanded BN for Ghana that includes yield, water productivity and rainfall. Some features:

- Poverty (measured as headcount ratio, Headcount_p) seems related directly to drought (NDRY_Mont), crop yield (MAISY), employment status (UNDEREMPLO) and landlessness (LANDLESS).
- Other variables are associated through intermediates. For example, Water Productivity (MAISWP) is related to poverty headcount, but through crop yield.
- Some associations are explained by geographical co-variation, not logic. For example, the relationship between drought and adult education (EDUC_ADULT)or access (Acc_30) is difficult to explain unless they are linked by an underlying geographical variable.
- Variables could be grouped according to the WPI concepts of resource (rainfall, drought), access (no variable included, but could include m^3/cap/yr), capacity

(employment, nutrition), use (crop productivity), and environment (no variables included).

- Poverty *per se* is not the target variable but water poverty**.** This is not measured directly but may be inferred indirectly from the above net. Work is on-going on this aspect.

BN for poverty in Burkina Faso.

It was not possible to produce a BN for BF that included data from both the survey results and from data such as climate, crop productivity, water productivity, extracted from maps.

The BN for the survey data is shown in Figure 11. Variables are ordered approximately according to the WPI. It should be noted that inference is weaker because of the smaller data-set.

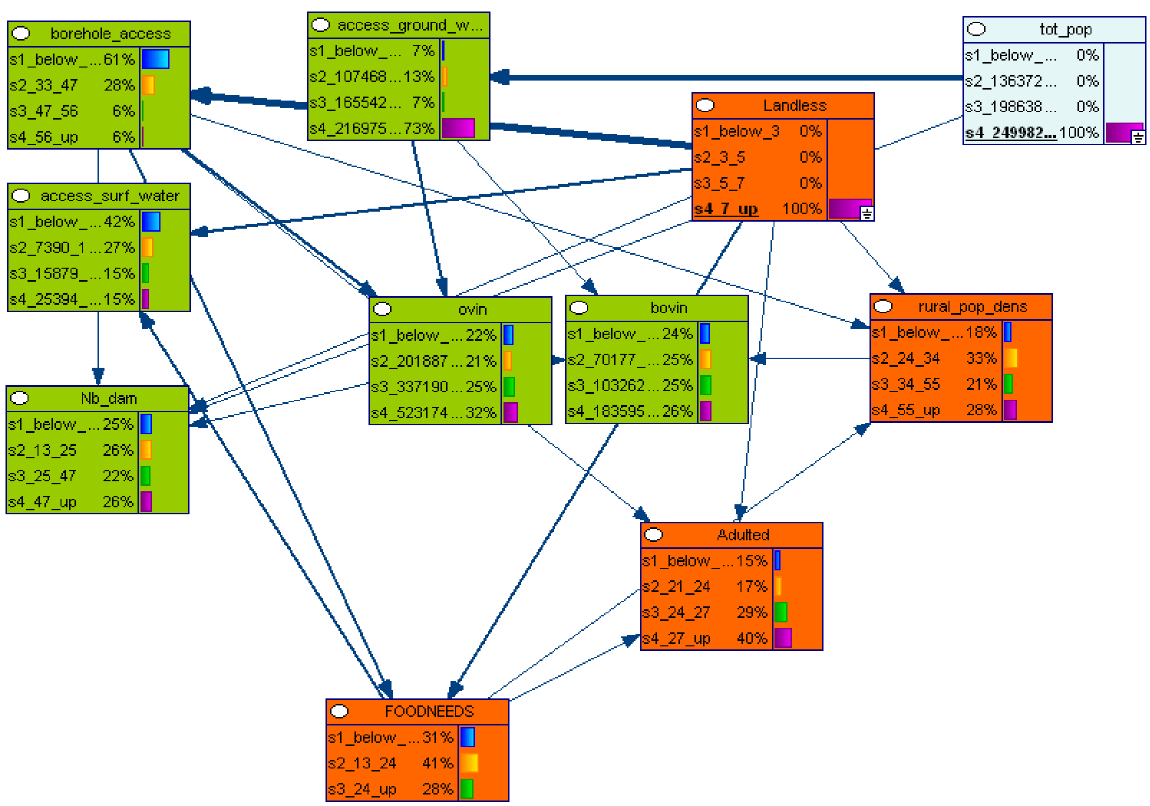


Figure 11. Augmented Bayesian Network to describe the relation of poverty with potential causal variables in Burkina Faso. The thick arrows have stronger relations than the thin ones. Colours are used to identify variables such as drought, that are targetted for improvement with further analysis.

Nevertheless, we were able to draw some conclusions.

- Access to ground and surface water (access_surface, access_groundwater) seems important to livestock numbers, but does not translate (in this dataset at least) through these attributes to measured food needs (FOODNEEDS). But data from Ghana also showed that food security is indirectly related to poverty. In other words, water security does not seem to act *directly* through livestock.
- Landlessness is obviously related to (lack of) access to water, conversely population density is strongly correlated with access to groundwater.

Data of rainfall, drought, cereal and maize productivity were available but were not co-registered for the same locations. A separate report is being prepared.

Step 6. Mapping the BN model.

A separate report is being prepared to produce maps for Ghana and Burkina Faso similar to that produced for Ecuador by Farrow *et al.* (2005).

Discussion

In terms of the application of Bayesian Networks to explain linkages between water and poverty, it was agreed that a standardised framework such as that provided by the Water Poverty Index was useful. Through the application of Bayesian Networks, it was possible to show the strength of the influence between certain variables, and it was clear that a strong relationship existed between cereal yield and rainfall and between the area sown to cereals and population. It was also shown that this reflected the greater degree of drought tolerance of millet and sorghum than maize, in that areas with more dry months per year grow more millet and sorghum and less maize. These very simple results provided evidence that there was merit in the approach, and thus the objectives of this workshop had been achieved. It is reasonable to conclude from this that more time spent on the development of this approach would be likely to not only lead to a valuable analysis of other basins in the Challenge Programme, but also perhaps lead to the development of a methodology with more widespread generic applications.

The workshop also revealed that more care should be taken in data collection activities over the unit of analysis, and ideally, the conceptualisation of the model to be used should be conducted before the data are collected. It was also important to note that there were major differences between the countries in terms of parameter definition, and this was not simply a question of semantics, but also due to cultural differences between the two countries.

With respect to information on water, a number of issues arose. Firstly, monthly average rainfall data is too coarse to provide adequate information of water availability to either crops or people, and the addition of other data would be needed to create a parameter to reflect variability and reliability of water resources. It is also important to note that this data needs to be for specific years to enable us to compare it with poverty measures for the same period. Furthermore, it was very difficult to find useful information on groundwater. There is little hydrogeological information available, and although maps of boreholes were available, there was no information on either their seasonal water levels or their reliability. This was especially important since groundwater is often accessed through hand-dug wells and springs, neither of which are included in the borehole data. With respect to surface water, although information on small dams was available, it was not harmonised between the two countries, and it would therefore need more manipulation before it could be used. This is likely to be similar in other CP basins in which the basin is made up of a number of countries each collecting data with little commonality.

We concluded that would be most useful to have a homogenous data set across the different countries of the basin rather than separate data. Boundaries within the basin should be removed within the basin model. In addition, it would have been valuable to be able to include longer period of precipitation data (only 30 years' data were available) than we were able to include in this analysis.

Conclusions

The question that must be asked is: What would people benefit from further development of these methods? The answers seem to be:

- For *project scientists, theme leaders and basin coordinators*, this analytical process enables them to understand and visualise the water-related causes of poverty within the basin;
- for *NGOs*, the analysis enables them to quantify knowledge about their constituents and describe specific problems that are amenable to solution;
- for *donor* stakeholders, analysis provides objective advice on the dynamics of poverty and its distribution within basins.

When considering links between water and poverty, there are many factors to be considered. To this end it would be useful to include more local representatives in the discussions of what are the important factors.

It was shown that the Bayesian Network approach was found to be very useful tool for the engagement of stakeholders, and much animated discussion was generated through its use and the associations that it generated. It is of course important to realise that the method itself is not the solution, and it is important to include local knowledge and expert opinion, not only as a way of verifying the robustness of the conclusions, but also as an input into the process.

It is also clear that different places need different variables, and an important characteristic of any approach is to maintain flexibility. It is clear that there are many factors that influence the links between water and poverty across the Volta basin. When the methodology is further refined, it should have much relevance as a decision support tool for policy makers. In particular, the newly formed Volta Basin Authority is a potential end user of the outputs from this work, as well as the Ghana and Burkina Faso National Water Research Commissions. This work will also be useful as a contribution to other work attempting to reveal other factors as well as water itself that influence poverty. The kind of data gaps that emerged, such as the lack of good groundwater data, are important, especially as groundwater is an important, yet underused source of domestic water throughout the basin.

Recommendations and next steps

The analytical process should be run within all basins to explore and represent what is known about poverty and agricultural water management.

It was clear from the workshop that the BFPs are generating large volumes of data, and there is a need to manage the large datasets and archive them for future research.

A number of issues arose about data, some of these which need further improvement. These include:

- Food needs (% of population expressing some or frequent food shortages);
- Livestock units/person (cattle/sheep/chickens to be weighted and summed);
- A measure of resource variability or drought risk;
- Volume of small dams per person (with a focus on rural populations);
- Groundwater access (% of population having access) including boreholes, wells and other groundwater sources, and their reliability;
- Grain productivity/rural inhabitant (kg/ha/person);
- Rural population density; and
- Crop production/rural inhabitant (kg/person).

In a country where economic activities are based on agriculture and livestock, improving management of water resources is a high priority. It is thus necessary to identify appropriately-adapted indicators that are likely to provide more accurate information for the development of policy. In addition, it is important to identify the impact of integrated water resource management (IWRM) and good governance needed to promote it. This must involve users in both decision-making and operation and maintenance of infrastructure . Nevertheless, in countries such as those in the Volta basin, it is also very important to recognise the value of traditional governance strategies (for example, relating to small dams), and this remains an important research theme. Another issue that could be fruitful would be to examine how this approach can be applied at different scales, for example to a micro-catchment, to a district, and to the basin as a whole.

In all cases, analysis will support and quantify existing concepts about the role of water in livelihood systems. The WPI is a useful concept that could serve as an underlying framework for the analysis of attributes through a complementary concept of the livelihood system analysis, proposed by the Comprehensive Assessment (Castillo et al, 2007, see figure below).

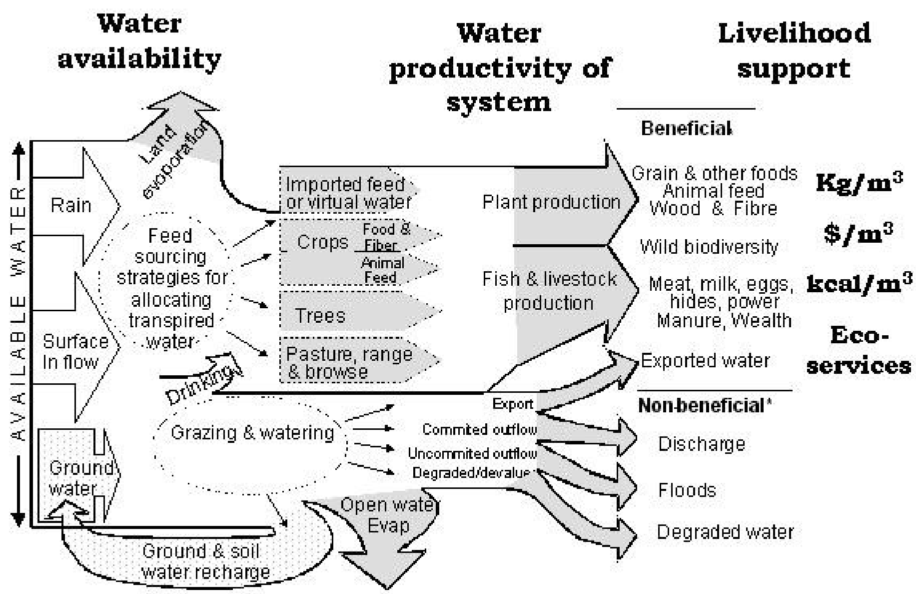


Figure 12. Framework for improving water productivity of livestock. (Modified from Peden *et al.*, 2002.)

Workshop Participants

All participants are authors of this report. Authors' affiliations and email addresses are included in the title page.

Data sources and References

Assante, F.A. (2006) Socio-economics of fisheries dependent communities in the volta basin of Ghana. *International journal of Ecology and Environmental Sciences*, 32(1):127-132 New Delhi

Assante, F.A. (2007) Analysis of Water Related Poverty in the Volta Basin of Ghana ISSER, University of Ghana.

Bonham-Carter, G. F. (1994). "Geographic information systems for geoscientists," Pergamon, New York.

Corner, R. J., Moore, G., Cook, S. E., and Raaff, E. (1998). "Expector - A manual for the expector soil attribute mapping method. First draft," CSIRO Land and Water, Perth.

Coulombe, H. (2005). Ghana Census Based Poverty Map, District and Sub District Levels. GSS/DFID.

Farrow, A., Larrea, C., Hyman, G. G., and Lema, G. (2005). Exploring the spatial variation of food poverty in Ecuador. *Food Policy* **30** 510

Ghana Statistical Service(2005) Core Welfare Indicators Questionnaire (2003) Survey Report. Ghana Statistical Service, Accra, August 2005

GLSS (2000) GLSS 4 Ghana Living Standards Survey, 4th round(1998/99). GSS Housing and population census, 2000 GSS

Heckerman, D., Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). Bayesian networks for knowledge discovery. In "Advances in knowledge discovery and data mining", pp. 273. AAAI Press / The MIT Press, Menlo Park.

INSD, *Enquêtes Burkinabè sur les Conditions de Vie des Ménages* (EBCVM), (2003)

INSD, *La pauvreté au Burkina Faso*, (2003) 70 p

Nikiema, A. and JM. Dipama, (2007) *Pauvreté et pauvreté hydrique au Burkina Faso*

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence* **29**, 241.

Sullivan C.A. and Meigh, J.R. (2007) Integration of the biophysical and social sciences using an indicator approach: Addressing water problems at different scales *Journal of Water Resources Management*

Sullivan C.A., (2001). The potential for calculating a meaningful Water Poverty Index. *Water International*, 26, 471-480.

Varis, O., and Keskinen, M. (2005). A Bayesian Network Model For Intersectorial Policy Analysis In The Tonle Sap Area, Cambodia. In "Proceedings of the 2005 International Conference on Simulation and Modeling", V. Kachitvichyanukul, U. Purintrapiban, P. Utayopas, eds."

Appendix

Workshop Activities

Day 1.

The workshop started with an introduction and summary of expectations, and then some existing insights into the analysis of water, agriculture and poverty were provided by Felix Asante, Aude Nikiema, Eric Kemp-Benedict and Caroline Sullivan.

Following that, an introduction to Bayesian Methods was provided by Jeremy Wallace, followed by a worked example from a case study in Ecuador. A lively open discussion took place on the methods, models and data requirements.

Day 2.

This day was taken up by intensive data organisation and development of various conceptual models. This culminated in a preliminary testing and learning Bayesian techniques. The socio- economic data from Ghana were linjed with those from Burkina Faso, and as far as possible, were then harmonised. Biophysical data including crop data (maize and other cereals) rainfall, water productivity and so on were combined with population and other socio-economic data to build a combined dataset for the Ghana and Burkina Faso parts of the basin. A calculation was made on the number of dry months (based on months with less than 25mm rainfall) as a first attempt to represent length of the growing season and droughtiness.

During the workshop various participants were involved in seeking data from different sources and consulting people outside the workshop group for clarification. Concurrently, other participants reviewed and reduced the data by choosing those variables and units that would be relevant to Bayesian networks to understand the factors involved in water poverty. A first iteration of a network was then constructed and run as a GeNie model, based on Burkina Faso data only. The outcomes of this preliminary exercise were discussed in depth in a session on the interpretation of the results at the end of the day.

Day 3

The objective was to develop an integrated dataset based on data from both Ghana and Burkina Faso, with a view to represent the Volta Basin[16]. Where necessary, the data from the different sources were transformed to make them compatible between countries, and then the variables were divided into quartiles each containing equal numbers of districts. These data were then converted into 4x4 contingency tables for each variable and imported into the GeNIe package to make trial runs. We tested various alternative models, some iteratively to improve the fit.

---

[16] While this excludes data from the four other countries of the basin, Ghana and Burkina Faso between them occupy 335,000 km^2 or 85% of the land area of the Volta basin, see http://www.fao.org/docrep/W4347E/w4347e0u.htm.

Selection of the data appropriate to create a Bayesian Network is a subjective process, with the objective of capturing the knowledge that the existing data represent. The team assembled the combined Ghana and B urkina Faso dataset and spent a lot of time discussing the relevance of individual variables to water poverty and therefore whether to include them in the Bayesian network. This was an iterative process. The first step was to create preliminary models, which showed some promise in demonstrating links between water and poverty across the various districts of the countries. This was followed by reorganization and refinement of the data and consideration of alternative combinations of variables. Some of these combinations are shown in Table A1.

Table A1. Selected examples of water poverty variables modelled in a Bayesian Network

Example 1	Example 2	Example 3	Example 4	Example 5
– Areas of sorghum and millet (av for 1992-01) – Area of maize (av for 1992-01) – Pop total – Maize water productivity – Maize yield – Maize productivity – Precipitation (av 1992-2000) – Number of dry months per year (less than 25mm)	– Total pop – % of children underweight – Number of dry months – Precipitation (av 1992-2000)	– % Landless – Number of dry months – Food shortage gap – % of children underweight – Population – Precipitation (av 1992-2000)	– Population – Food needs – Adult education – Under employment – Unemployment – % of children underweight – % children stunted – % landless	– Poverty headcount – Number of dry months – Adult education (to be rural only) – Maize water productivity – Maize yield – Water access within 30 minutes – Underemployment – Level of millet and sorghum production – % landless – % of children underweight

Some of these models did not show any important linkages between components, but others were more instructive. For example, the provinces with highest populations have lower than expected underweight children, and also places with lowest rainfall had the highest number of dry months and the highest food needs. While the latter seems entirely reasonable, it is encouraging that the Bayesian network captures it and provides numerical evaluation of the probabilities.

Issues identified following from the workshop activities

The workshop identified many positive outcomes and some difficulties.

Positive outcomes

All participants said that the lively and open discussions throughout the workshop were a valuable part of the learning and networking process. Furthermore, participants identified limitations in the data, which would help to strengthen data collection and collation in the future. Participants were very interested in the utility of the Bayesian approaches and felt that the Volta basin model will be a useful prototype, especially with further refinement. Participants noted that linking physical and social data is always a challenge, but that a strong interdisciplinary team was able to make considerable progress. The workshop highlighted the benefits of a WPI-like structure, which is not observable directly but can be inferred from other data. This aspect requires more work for more precise definition.

Problems identified from the workshop activities

Preliminary organization of the data took a lot of time. Clearly, more pre-preparation could have been done, but the scale and resolution of the data needed for the exercise were not defined prior to the workshop.

It was difficult to combine the data from Ghana and Burkina Faso. Not only were there differences between the two countries in the way variables were defined, but in Burkina Faso in 1996 the provinces boundaries changed so that survey results before 1996 are not comparable with more recent data. For agricultural data, however, the old provincial areas were still used until 2001, adding further difficulties for comparisons both within Burkina Faso and between Ghana and Burkina Faso. There were further problems in data integration because of different definitions of the administrative units of district, region etc. between the two countries apart from differences in the metadata available. Some of these problems might have been overcome had more time been available in the workshop, and more time could have been spent by the participants gaining a better understanding of the software and its limitations.

Mapping of Bayesian model outputs

It was not possible to map the output of the Bayesian Network analysis during the workshop because of technical problems in the preparation of input data for the mapping packages. We are preparing a separate report, which will be available on the BFP Wiki site in due course. Although it requires more effort to make a map, the Ecuador example ((Farrow et al., 2005), which shows where different forms of water poverty occur within basins, demonstrate how valuable it is.