VAS

Donkers, H.H.L.M.; Ferreira, R.; Uiterwijk, J.W.H.M.; van den Herik, H.J.

Published in:
BNAIC'99

Publication date:
1999

Citation for published version (APA):
Donkers, H. H. L. M., Ferreira, R., Uiterwijk, J. W. H. M., & van den Herik, H. J. (1999). VAS: Quantifying a qualitative network. In E. O. Postma, & M. Gyssens (Eds.), *BNAIC'99: Proceedings of the eleventh Belgium/Netherlands artificial intelligence conference, Maastricht, November 3-4, 1999* (Vol. 11, pp. 139-146). IKAT, Universiteit Maastricht.

VAS: Quantifying a Qualitative Network

Jeroen Donkers Rui Ferreira Jos Uiterwijk
Jaap van den Herik

Institute for Knowledge and Agent Technology (IKAT),
Computer Science, Universiteit Maastricht, P.O. Box 616, 6200 MD Maastricht

Abstract

VAS (*Verkennend Analyse Systeem*) is an existing qualitative tool in a decision support system. It is used for rapid problem assessment by policy developers. VAS is built on the knowledge of experts, but the decisions so taken lack a theoretical foundation. In this paper we focus on a scientific reliable foundation. To model the domain, we relate potential consequences with probabilistic means to their causes, and attempt to turn qualitative statements into quantitative ones, especially where end users of the system indicated that quantification of VAS is needed. We developed a method to translate a VAS model into a quantitative Bayesian probability network. This paper deals with the first three steps of gradual quantification. We applied the method to the reformulated VAS model on the problem of the fresh-water inlet to the East-Scheldt estuary.

1 Background

Policy developers in the Netherlands are nowadays facing an increasing number of ambitious large-scale projects, such as the high-speed railway, the Betuwelijn, the second Meuse plane in Rotterdam, and the reconstruction of the West-Scheldt estuary. They have to take into account the viewpoints of different stakeholders and pressure groups, all having their own objectives and criteria. Information technology provides many tools for these policy developers. In order to investigate how new developments in information technology can assist policy developers, the LWI project (*Land Water and environmental Information technology*) was founded by the Dutch ministry of economic affairs.

Our group takes part in the LWI project by investigating how techniques from Artificial Intelligence, and especially from the field of probabilistic AI can provide a formal basis for decision support systems that are useful for policy developers. The branch of LWI in which our group participates concentrates on policy development for estuaries such as that of the West and East Scheldt. For this purpose, our LWI-partners previously produced the EDSS (Estuary Decision Support System), which is a generic decision support system for long-term governmental policy development [9]. The program aims at civil servants who develop or prepare policies at the regional government.

EDSS consists of several analysis tools which are connected by a generic shell. This shell is set up according to the Framework for Analysis, a stepwise approach to policy analysis used by LWI-partner Resource Analysis[1] [9]. This stepwise approach helps policy developers to analyse a problem in a structured way, improving the communication among policy makers and stakeholders.

Our research focusses on a part of EDSS called the VAS (Verkennend Analyse Systeem, *eng: Preliminary Analysis System*). This tool is used for rapid problem assessment (e.g., on the inlet of fresh water in the East Scheldt). It provides the user with a quick overview by qualitatively calculating effects of measurements.

Our research goal is to provide an alternative for VAS, for which we focus on probabilistic methods. Unfortunately, the current VAS module lacks a formal background. Moreover, end users of EDSS and VAS have indicated that a purely qualitative approach does not provide them with sufficient information. It would be desirable that the output of VAS is precise and accurate, especially when more information has become available. We therefore aim at applying a technique that allows a gradual and easy transition from a problem specified purely qualitatively to a completely quantitative specification. The technique that we choose to apply is a Bayesian probability network (or belief network) [6, 7].

In section 2 we explain the VAS module as it is used now. Section 3 shows the method, i.e., how we translate a VAS network into a comparable Bayesian probability network and how we want to implement the gradual quantification of VAS. In section 4 we present a practical application of our method to the problem of fresh-water inlet in the East Scheldt [3]. Finally, section 5 states our conclusions.

2 Verkennend Analyse Systeem

In the existing VAS system, the user starts appointing the elements of interest (*variables*) in the system. First, the variables are grouped together in functional units. For instance, in Figure 1 the variable KREEFTACHTIGEN (crustaceae) belongs to the functional group BRAK SOORTEN (brackish species). Second, the user defines qualitative relations between the variables like "if A increases then B decreases and C does not change." Qualitative relations are indicated by one

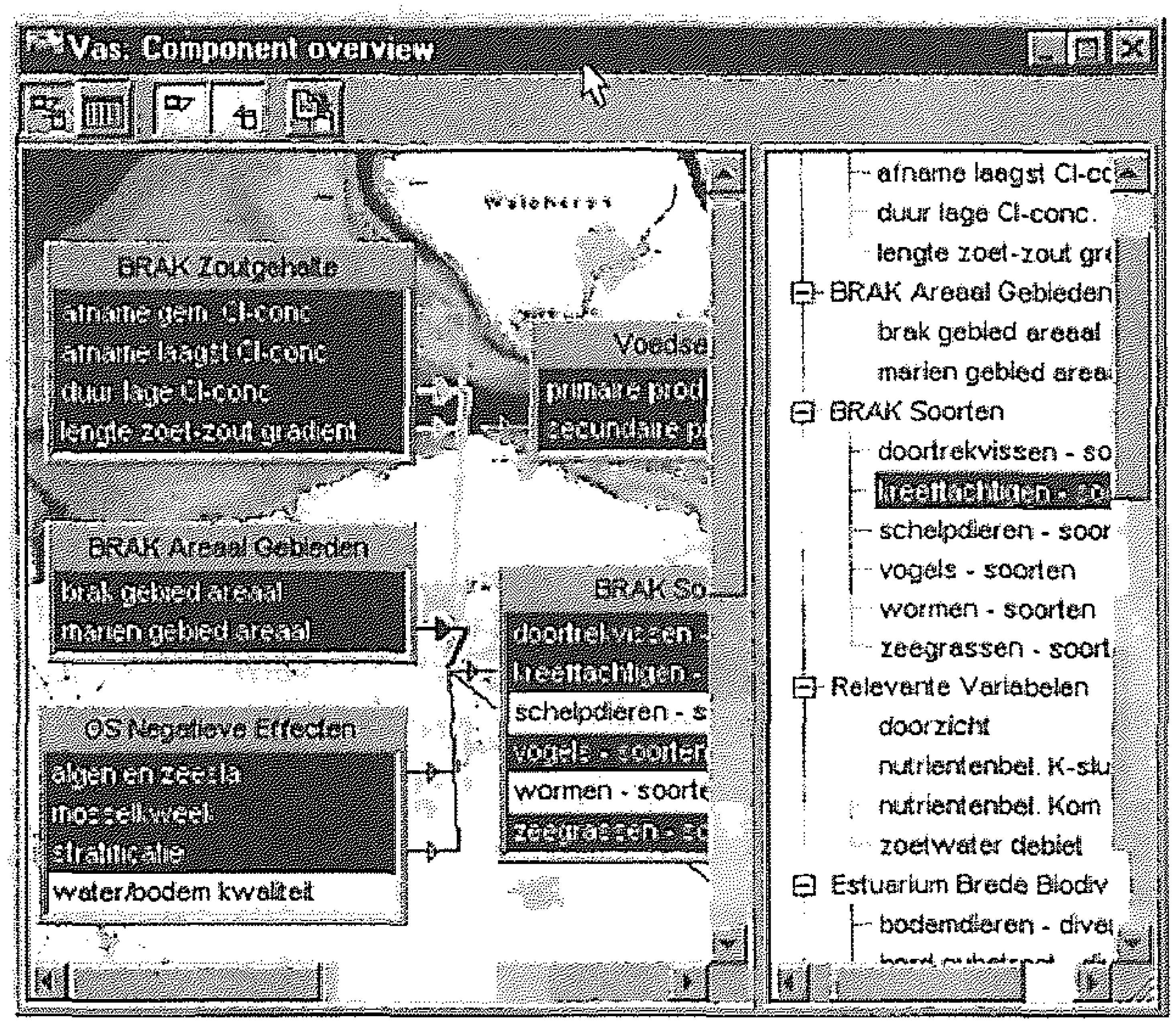

Figure 1: A screendump of the VAS module.

[1]Resource Analysis is a consultancy firm in Delft, the Netherlands, specialized in policy development. (url: `www.resource.nl`)

of seven strength signs ("− − −", "− −", "−", "0", "+", "++", "+++"). In the user interface, these relations are indicated by arrows and their strength is indicated by the colour and the thickness of the arrow. For instance, in Figure 1 an arrow runs from STRATIFICATIE to KREEFTACHTIGEN indicating a "+" relation between these variables.

Effects of measurements are expressed as qualitative changes in the values of the variables. The change-in-value is called the *state* of the variable; we use the same seven strength signs for the indication of a state. Table 1 shows the combination rules for a state variable, given the strength of the relation and the state of the influencing variable.

The relations between variables in the VAS are directional. The relation $A \longrightarrow B$ is not the same as $B \longrightarrow A$. The strength of the first relation could be "+" while the latter relation could have strength "− −". Such a configuration would indicate some feedback relation between variables A and B.

infl. state / strength	− − −	− −	−	0	+	++	+++
− − −	+++	++	+	0	−	− −	− − −
− −	++	++	0	0	0	− −	− −
−	+	0	0	0	0	0	−
0	0	0	0	0	0	0	0
+	−	0	0	0	0	0	+
++	− −	− −	0	0	0	++	++
+++	− − −	− −	−	0	+	++	+++

Table 1: Combination rules as used in VAS.

When more than two variables have a relation with the same variable, the state of that variable depends on the strength of all relations and the states of the related variables. In VAS, the output will be a *state range* that indicates the lowest and highest amount of change that may be caused by applying the rules of Table 1 to each of the relations separately. For instance, if B has a relation to A with strength "+++" and C has a relation to A with strength "− − −" , and given the values of B and C are both "+", then the state range of A is "[− .. +]". (In the case of only a single influencing variable, VAS will in fact produce a range too. This range has zero width such as "[+ .. +]".)

The computation of the effects in VAS is performed by first initializing all variables with state "0", meaning no change. Then some variables (according to project criteria or policy criteria) are set to a fixed state, indicating some external event or measurement. Subsequently the rules of Table 1 are applied to compute the *first-order effects*. The computation results in a range of states for every variable. In the following steps, the rules are re-applied to compute second, third and higher-order effects as long as the user wishes to continue.

There is no exact description available of the semantics of the VAS tool. It is up to the user to define what a *variable* is, what the meaning of a *relation* is, and how the *strength* of the relations should be interpreted. Moreover, the computation rules of Table 1 appear to be given by practical motivations (i.e., expert knowledge). In our view, the VAS-approach is a mixture of two concepts. At the one hand, the main question that VAS tries to solve is: "how will the variables change in the future, when a certain measurement is taken?" This is in principle a *static* problem. At the other hand, the rules are applied in such a way that effects are propagated step by step as in a *dynamic* way. The dynamic character also follows from the fact that cycles of relations can exist; even relations between two variables in both directions can be present at the same moment. In fact, the relation $A \longrightarrow B$ seems to be interpreted as "if A changes, then B will change *on a later moment*."

3 Translating VAS into a BPN

Our practical goal is to provide a smooth transition from the qualitative VAS network into a quantitative Bayesian probability network. Our contribution to science aims at finding relevant answers on the following two questions: (1) What is the best methodological approach to transfer a rather arbitrarily-built knowledge-based expert system to a transparent system which can reason quantitatively? and (2) What are the pitfalls that may occur during such a transfer?

A Bayesian probability network (BPN) is a compact representation of a joined probability distribution [6, 7]. It contains a set of (discrete-valued) variables and a set of links between them. The links in a BPN indicate the conditional dependencies that exist in the probability distribution. These links can be interpreted (roughly) as direct causal influences. Every variable in a BPN possesses a table that contains the conditional probability on the values of the variable, given the values of the influencing variables. All probability tables together contain the original joined probability. Although reasoning with BPNs is NP-complete in general, in practice BPNs appear to be effective and many applications of BPNs are in use today [4].

In this paper we present our method consisting of six steps. The first three steps of the transition are emphasized, since they specifically deal with the transition of the internal structure. In the first step we translate the VAS network into a qualitative probability network (QPN) [8, 1]. In the second step we translate the QPN into a comparable BPN. These two steps are in fact a preparation for the process of quantification. In the third step, we perform the first level of quantification by allowing the user to specify the strength of a relation by a real number.

Step 1: From VAS to a QPN. In the first step VAS is translated into a QPN. This is performed by hand. The process is guided by the strict criterion whether a relation is causal or not. All relations in VAS are scrutinized carefully, because the meaning of a relation in a (qualitative) probability network differs from the meaning of a relation in the VAS. In a probability network, the relations between variables represent conditional dependencies between variables. These relations can, with the required care, be interpreted as causal links [2]. In VAS relations are sometimes interpreted as causal links, but more often they are not.

Our goal is to build a QPN in which all links can be interpreted causally. Sufficient conditions are the *Markov* condition and the *Faithfulness* condition (see [5]). In short, the direct children of a node in the QPN should shield off all dependencies to other descendants, when the states of the direct parents are given. For us, these conditions imply that only direct causal relationships between variables can take part in the QPN. All relations in the VAS network that represent indirect causal relationships must be removed. Furthermore, in a probability network, cycles of relations are not allowed; in VAS this restriction does not apply. A cycle of relations exists if there is a relation from A to B and at the same time a direct or indirect relation from B to A. To eliminate cycles, those relations have to be removed from the VAS network that do not indicate a direct causal link.

Step 2: From a QPN to a BPN. In the second step, the qualitative probability network is translated into a BPN. The major task is the transformation of qualitatively specified probabilities into real-valued probability values. Since we wish to preserve the modelling power of VAS as much as possible, we put four requirements on the probability tables.

(1) All variables obtain the same number of states. The number will depend on the size of the network and the available amount of memory. A typical choice is 2 ("−","+"), 3 ("−", "0", "+"), 5 ("−−", ..,"0", ..,"++") or 7 ("− − −", ..,"0", ..,"+ + +").

(2) The states of all variables represent *equidistant* levels of change.

(3) As in VAS, the effect of a change will not be larger than the cause (Table 1).

(4) When two or more variables influence one variable, we wish to discriminate as much as possible between the separate influences.

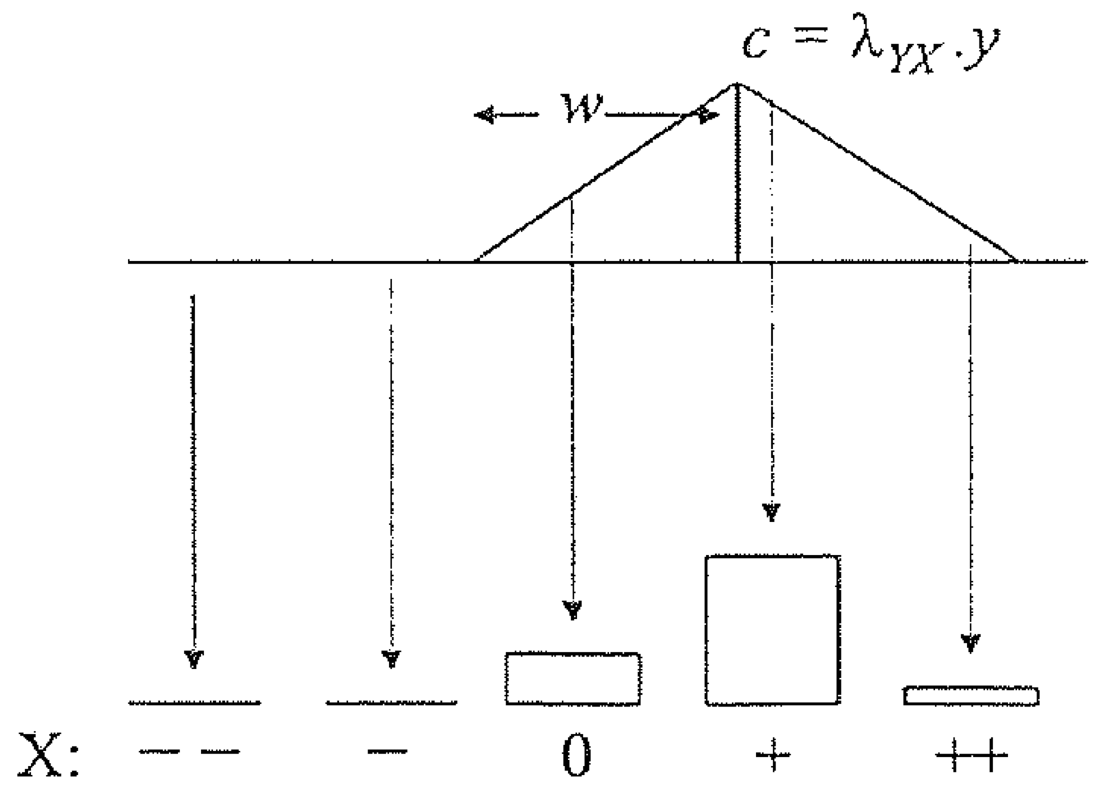

Figure 2: Probability table function for $Y \longrightarrow X$.

Moreover, we decided to develop a single function that produces all probability tables needed.

The construction of this function starts with the case of only one influencing variable ($Y \longrightarrow X$ with strength λ_{YX}). The resulting probability table should contain the conditional probabilities $p[X = x|Y = y, \lambda_{YX}]$. The function that we propose has a fixed triangle shape around a moving centre. The place of the triangle's centre depends on the state of the influencing variable and on the strength of the relation. When we represent the states of variable Y as equidistant points on the interval $[-1, 1]$, and the relation's strength λ as a predetermined number within the same interval, then the centre c of the triangle is given by: $c = \lambda.y$. The width w of the triangle represents the *vagueness* in the system. The wider the triangle, the more the probability will be spread out over states. This width is an external parameter that must be provided by the user.

In Figure 2, the function is drawn for the situation where the state of Y is $y =$"+", width $w = 0.75$, and strength $\lambda = 0.8$. The probabilities of the states of x are then taken as the value of the triangle function at these points (indicated in the lower part of Figure 2). To produce valid probabilities, these values have to be normalized. The total probability table is computed by applying this function to all states of Y.

In case of more than one influencing variable, we compute the function from Figure 2 for every variable separately, add the functions together and normalize the results (see Figure 3).

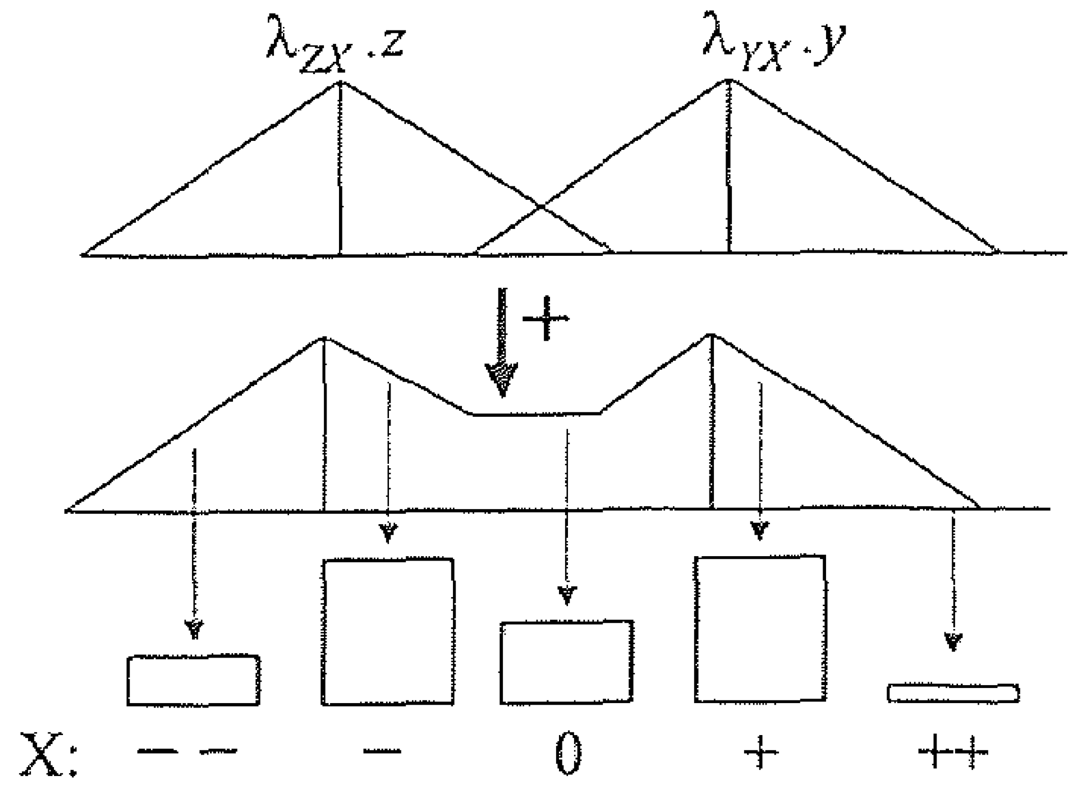

Figure 3: Probability table function for $Y, Z \longrightarrow X$.

The probability tables for almost all variables in the BPN are computed by applying the above functions. The only variables that are not treated this way are the variables that have no variables influencing them. These variables are called *prior* variables. We propose that they receive a table that has equal probabilities for all states.

Having computed the probability tables, the translated network is ready for use. It is given to an existing computer program for BPNs after which the user can consult the BPN by entering evidence to the network and reading out the probabilities of the desired variables.

The output of this translated network will usually differ from the output of VAS because the BPN produces a probability distribution over the states. The exact probabilities are not very important for the user, but the shape of the distribution can give some important hints. Figure 4 shows three example distribution shapes. The first one shows a unimodal distribution in which one state is clearly favourite. The second one shows a bimodal distribution in which two disjunct states are equally favourite. This indicates counteracting forces in the network. The third distribution is almost uniform, which indicates maximal uncertainty for that variable.

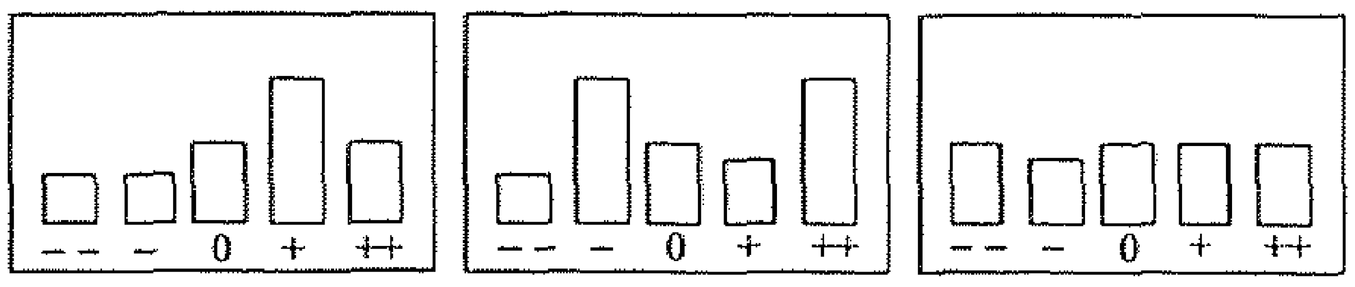

Figure 4: Example output distributions for a variable in a BPN.

Step 3: Quantification of the strength. This step is the first genuine "quantification" step that is performed by the user. It is straightforward and easy to apply: instead of using strength signs ("$- - -$", .. , "0", .. , "$+ + +$") for the relations and translate them to fixed λ-values, we let the user specify a free numeric value for the strength. Although this step does not change anything to the way in which the BPN is produced, for the user this step adds a degree of freedom.

Further possible quantification steps are as follows:

Step 4: Specification of type of synergy. If more than one variable influence the same variable, then the user can add to the network information on whether the influences amplify or neutralize each other.

Step 5: Specification of asymmetry. The user can specify whether the influence on a variable is stronger for increase than for decrease.

Step 6: Specification of variable states. The user can specify exact numerical values for different states, possibly selecting non-equal distances between states.

4 An example: The East Scheldt

We tested our approach on a case for which the VAS already has been applied: the fresh-water inlet to the East Scheldt. The East Scheldt is one of the sea arms of the Rhine / Meuse / Scheldt delta in the province of Zeeland. As part of the Delta-works the East Scheldt has been closed at the east-side by dams and dikes. The closing prevents fresh water to enter the East Scheldt, causing an increase of the salinity of the water. This results in a decrease of the typical and unique brackish ecosystem [3]. The goal of the local government is to restore the brackish ecosystem as much as possible, without increasing the danger of floods and without effecting the recreational and economical functions of the East Scheldt too much. For this problem, a VAS model was developed by Resource Analysis, together with a team of experts. This model counts about 45 variables in 12 groups. Figure 1 shows a part of this model.

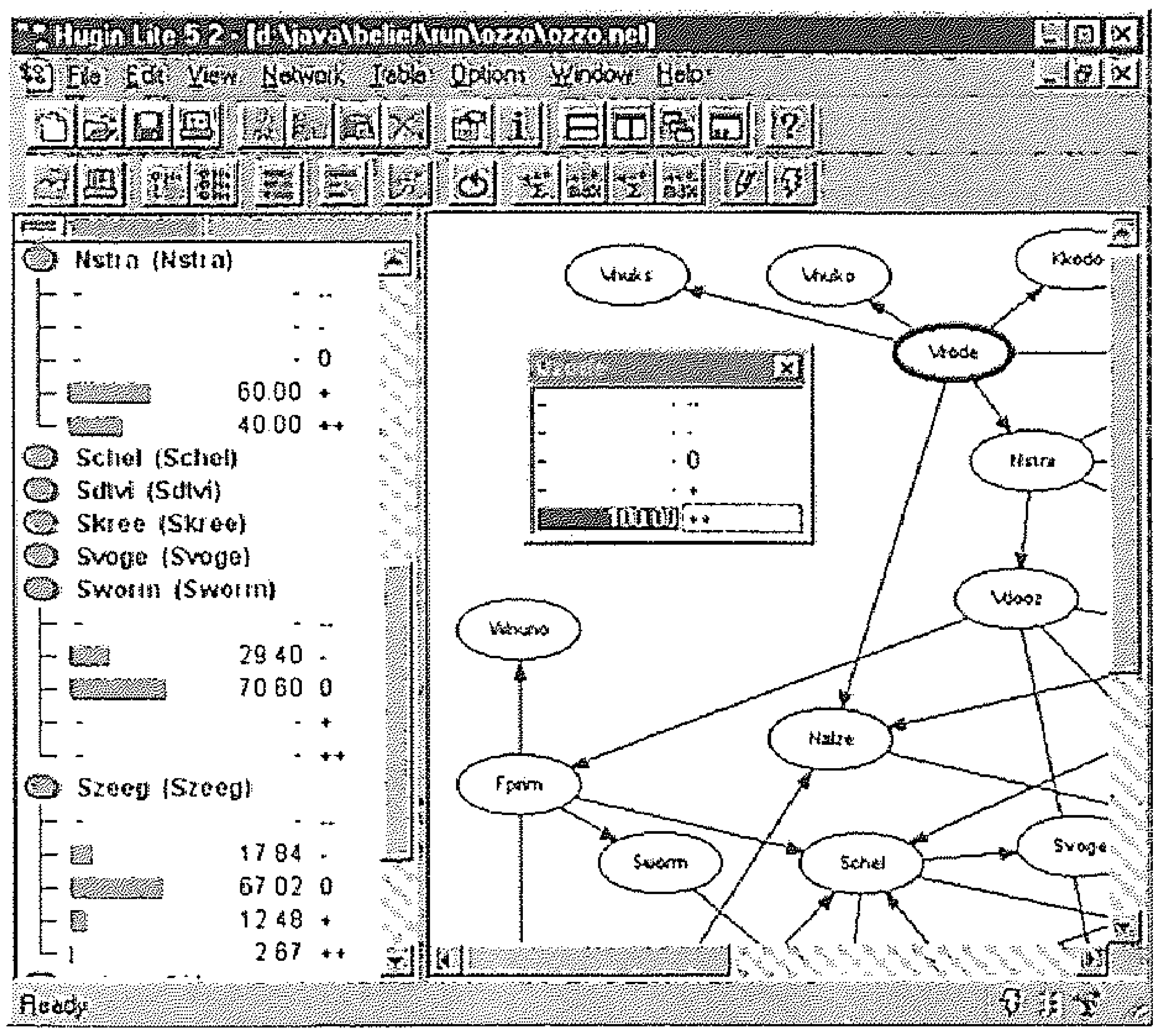

Figure 5: A screendump of the BPN in Hugin.

Applying our method, we first translated the VAS-information into a comparable qualitative network. This requested the removal of about one third of the existing links in VAS and the addition of one extra variable (not discussed in section 3). Then we translated the qualitative network into a BPN. We used the following numerical values for the λs: "+" = 0.3, "++" = 0.7, "+++" = 1.0, and the corresponding negative values for the negative strengths. Finally, We selected 5 states per variable, due to the size of the network.

The resulting BPN was saved into an input file for the program HUGIN, a commercial package for BPNs. Figure 5 shows the results for some variables in case the amount of fresh water (VZODE) is increased. The figure shows that the possibility of stratification (NSTRA) increases, that the amount of sea grass (SZEEG) will remain unchanged or will only slightly increase or decrease, and that the amount of worms (SWORM) will decrease moderately. The result for the same variables in VAS is for the stratification: [+ .. +], a moderate increase. The sea grass and worms both have the range: [0 .. +], which differs from our findings.

5 Conclusions

In the example application, we see that our BPN provides more information to the user than the original VAS, although not much quantification has been added yet to the network.

Our methodology is applied to a specific situation, i.e., the translation of VAS into a BPN. Although this seems a once-only task, the methodology is useful in other situations where qualitatively-specified knowledge is used and where there is need for some level of quantification. The essence is to remodel the system in such a way that variables are easily quantifiable. Of course, there is a pitfall in transferring a rather arbitrarily-built knowledge system into a transparent system which reasons qualitatively. The original semantics are unclear and therefore, the knowledge entered in the system is of poor use for the target system. This means that the careful redesign should involve domain experts.

The application of our method on the East-Scheldt VAS model has not been evaluated by the end users (at the moment of publication). This will be done in the near future. Upcoming research will also include the automation of the first step in the translation: transforming VAS into a qualitative probability network by an intelligent computer program.

Acknowledgements

The research has been supported by the Dutch organization for Land Water Milieu Informatietechnologie LWI (task number: LWI141). Thanks are due to our partners in this organization, especially to Peter Kouwenhoven and Marcel Taal from Resource Analysis. Hugin is a product of Hugin company in Aalborg, Denmark and is available at `www.hugin.dk`.

References

[1] Marek J. Druzdzel and Max Henrion. Belief propagation in qualitative probabilistic networks. *Qualitative Reasoning and Decision Technologies*, pages 451–460, 1993.

[2] Marek J. Druzdzel and Linda C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In *UAI-95*, pages 141–148, 1995.

[3] H.A. Haas. Zoet water naar de oosterschelde: een verkenning naar de effecten op natuur en visserij. Technical Report RIKZ 98.03, Rijksinstituut voor Kust en Zee, Middelburg, The Netherlands, 1998.

[4] David Heckerman, Abe Mamdani, and Michael P. Wellman. Real-world applications of bayesian networks - special theme. *Communications of the ACM*, 38-3, 1995.

[5] Scott B. Morris, Doug Cork, and Richard E. Neapolitan. The cognitive processing of causal knowledge. In *UAI-97*, pages 384–391, 1997.

[6] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publ., San Mateo, CA, 1988.

[7] Linda C. van der Gaag. Bayesian belief networks: Odds and ends. *The Computer Journal*, 39-2:97–113, 1996.

[8] Michael P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257–303, 1990.

[9] Harm Albert Zanting. *Estuarium DSS Westerschelde, eindrapport fase 3*. LWI, Gouda, The Netherlands, 1998.