

# Hydrologic Models for Emergency Decision Support Using Bayesian Networks

Martin Molina<sup>1</sup>, Raquel Fuentetaja<sup>2</sup>, and Luis Garrote<sup>3</sup>

<sup>1</sup> Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain  
mmolina@fi.upm.es

<sup>2</sup> Departamento de Informática, Universidad de Carlos III, Madrid, Spain  
raquel.fuentetaja@uc3m.es

<sup>3</sup> Departamento de Ingeniería Civil: Hidráulica y Energética,  
Universidad Politécnica de Madrid, Spain  
garrote@caminos.upm.es

**Abstract.** In the presence of a river flood, operators in charge of control must take decisions based on imperfect and incomplete sources of information (e.g., data provided by a limited number sensors) and partial knowledge about the structure and behavior of the river basin. This is a case of reasoning about a complex dynamic system with uncertainty and real-time constraints where bayesian networks can be used to provide an effective support. In this paper we describe a solution with spatio-temporal bayesian networks to be used in a context of emergencies produced by river floods. In the paper we describe first a set of types of causal relations for hydrologic processes with spatial and temporal references to represent the dynamics of the river basin. Then we describe how this was included in a computer system called SAIDA to provide assistance to operators in charge of control in a river basin. Finally the paper shows experimental results about the performance of the model.

## 1 Introduction

The SAIH National Programme (Spanish acronym for Automatic System Information in Hydrology) has been developed in Spain with the goal of installing sensor devices and telecommunications networks in the main river basins to get on real time in a control center the information on rainfall, water levels and flows in river channels. One of the main tasks of this type of control centers is to help to react in the presence emergency situations as a consequence of river floods.

During a river flood, operators in charge of control use knowledge about the physical system and hydrologic processes of the river basin to estimate future states and make decisions about defensive actions. The exact details about the physical system and behavior are normally difficult to know and therefore certain simplifications are made in order to provide quick and efficient decisions in the presence of problems. Operators use their experience trying to identify similar situations, either measured in past events or simulated with models, in order to forecast similar outcomes.

This is a case of reasoning about the behavior of a complex dynamic system (the whole river basin) with uncertainty and real-time constraints using data recorded by a limited number of imperfect sensors. To help operators in this task with automatic tools, a solution based on traditional mathematical models with deterministic simulation (e.g. [1] [2]) cannot be directly applied. The probabilistic nature of the rainfall forecast, the uncertainty on model parameters, the noise of sensor measures and the discrepancy between model results and observations are difficult to incorporate into a decision-support system that uses deterministic simulation models, especially if the problem area is composed of many small basins that need to be monitored simultaneously for flash flood warning. In addition to that, numerical forecasts obtained via deterministic simulation models do not include an assessment of their accuracy, so it is a task for decision makers to assign degrees of credibility to the values based on their experience in the operation with the models.

As an alternative approach, we describe in this paper a solution where hydrologic models are formulated as bayesian networks. Bayesian networks can be appropriate to model the intuitive understanding of physical hydrologic processes with an explicit representation of this uncertainty together with a natural representation corresponding to the causal relations typically present in river basins. Based on this approach, we have developed a computer system called SAIDA to provide assistance in making decisions about hydraulic actions during floods.

In this paper we describe first the type of bayesian networks that we have considered with spatial and temporal references to model different hydrological processes. Then we describe how they are integrated and used in the SAIDA tool to help operators in decision-making during floods. Finally we show experimental results corresponding to the evaluation of the performance of the bayesian model.

## 2 Modeling Hydrologic Processes as Spatio-Temporal Bayesian Networks

In order to provide an acceptable level of decision support in a real-time context we have designed a model considering the different meaningful hydrologic variables associated to the physical processes of a river basin. For each process, one or several types of causal relations have been identified that constitute the basic pieces for the complete bayesian model. Each variable  $X^i_t$  corresponds to a state (rain, flow, volume, potential damage, etc.) at location  $i$  in time  $t$ . In the model, time is divided into intervals of fixed duration  $\Delta t$  (for example,  $\Delta t=1h$  according to the time interval of the data collection network). The current time interval is identified as time  $t$  and past intervals are referred to as  $t-1$ ,  $t-2$ , etc.

As a result of an experimental analysis of physical influences, the general format of the causal relations that we have considered to estimate the value of a physical variable  $X^i_t$  with the upstream variable  $X^j$  is  $P(X^i_t|X^j_{t-1}, X^j_t, X^j_{t-2}, \dots, X^j_{t-k})$  (more than one upstream variable can be considered). This type of relation can be used together with a conditional probability that relates  $X^i_t$  with the observation  $E^i_t$  corresponding to gauge stations in the river basin:  $P(E^i_t|X^i_t)$ . This relation is especially useful when the hydrologic variable cannot directly measured by a gauge station as it happens for example with raingages.

Process	Causal relations	Partial network
<i>Runoff generation</i>	$P(N_t^i   R_t^i, M_t^i, C^i),$ $P(M_t^i   R_{t-1}^i, M_{t-1}^i)$	
<i>Runoff concentration</i>	$P(Q_t^i   Q_{t-1}^i, N_{t-1}^i, \dots, N_{t-k}^i)$	
<i>Discharge propagation</i>	$P(Q_t^j   Q_{t-1}^j, Q_t^i, Q_{t-1}^i)$	
<i>River junction</i>	$P(Q_t^i   Q_t^j, Q_t^k)$	
<i>Reservoir operation</i>	$P(V_t^i   V_{t-1}^i, Q_{t-1}^i, Q_{t-1}^j),$ $P(Q_t^i   Q_{t-1}^i, V_{t-1}^i, T^i, Q_{t-1}^j)$	
<i>Potential damages</i>	$P(D_t^i   Q_t^j)$	

**Fig. 1.** Examples of basic types of causal relations for hydrological processes

Six basic processes have been considered: (1) runoff generation, (2) runoff concentration, (3) discharge propagation, (4) river junction, (5) reservoir operation and (6) potential damages. The first three processes resemble the equations applied in conventional lumped rainfall-runoff modeling (Hortonian infiltration, linear response to rainfall excess and hydrologic flood routing). The *runoff generation* process represents the causal influence between rainfall and net rainfall and includes two basic relations to estimate basin moisture content and infiltration. Three variables are included in the infiltration model: basin average rainfall  $R_t^i$ , cumulative basin moisture content  $M_t^i$ , and average net rainfall  $N_t^i$ , all of them corresponding to the current temporal interval  $t$  and the spatial location  $i$ . Each variable is formulated in a qualitative domain composed of a finite set of discrete values relevant for decision support purposes. For instance, rainfall during a time step may have the following discrete set of significant values  $\{0, 3, 6, 9, 12, 20, 30, 50, >50\}$ , all of them expressed in mm. Two additional variables are required for the basin moisture model: basin moisture content  $M_{t-1}^i$  and basin average rainfall  $R_{t-1}^i$  in the previous time interval  $t-1$ . According to this model, runoff generation is assumed to be Hortonian, and net rainfall  $N_t^i$  is directly explained by rainfall intensity  $R_t^i$  and basin moisture content  $M_t^i$ . In its turn, cumulative moisture content  $M_t^i$  is directly explained by moisture in the previous time interval,  $M_{t-1}^i$ , and rainfall in the previous time interval  $R_{t-1}^i$ . Initially, these causal relations were formulated as  $P(N_t^i | R_t^i, M_t^i)$  and  $P(M_t^i | R_{t-1}^i, M_{t-1}^i)$ .

The results of model calibration of this model showed a lot of variability that was attributed to the basin initial condition. If the bayesian network is built using the full range of inter-storm curve number variability, the dispersion in the result is so large that many forecasts show flat probability distributions. In this case, instead the first relation of the the previous bayesian model, an alternative causal relation was used for  $N_t^i$  in the form of  $P(N_t^i | R_t^i, M_t^i, C^i)$  which includes as additional cause the variable  $C^i$  (SCS curve number). In real time, the bayesian model uses explicitly an estimate of this parameter as input, which can be provided by the operator using knowledge of initial conditions with the help of a simulation model. In this model, conditional independence between  $R_{t-1}^i$  and  $R_t^i$  is assumed considering that the time interval  $\Delta t$  is large enough (e.g., one hour). This assumption is based on empirical studies of the behavior of torrential rain that presents low persistency and consequently low level of correlation between consecutive values of rain.

The *runoff concentration* represents the response to rainfall excess. In this case, the variables are  $N_{t-1}^i, \dots, N_{t-k}^i$ , which correspond to net rainfall for  $k$  previous time intervals and  $Q_t^i$  the average discharge (in  $m^3/s$ ) in the current time interval. The number of temporal intervals of net rainfall ( $k$ ) is chosen balancing the need to represent the length of the unit hydrograph (an hydrological parameter associated to each river basin) and the need to limit the number of explaining variables to a manageable size. In practice, it should be reduced to three or four intervals. The causal relationship is expressed as  $P(Q_t^i | Q_{t-1}^i, N_{t-1}^i, \dots, N_{t-k}^i)$ .

Another type of relation corresponds to the process of *discharge propagation*. This model represents the flow transportation from a certain location (spatial location  $j$ ) to a downstream location (spatial location  $i$ ), assuming hydrologic routing. The variables are  $Q_t^i$ , that corresponds to the flow at location  $i$ , and  $Q_t^j$  that corresponds to the flow at location  $j$ . The causal relationship is expressed as  $P(Q_t^i | Q_{t-1}^j, Q_t^j, Q_{t-1}^j)$ . For a *river*

junction another dependency can be established as  $P(Q_t^i | Q_t^j, Q_t^k)$  where  $Q^j$  and  $Q^k$  are upstream flows of the flow  $Q^i$ .

Reservoir operation is included in the model as an additional set of causal relations. A model of reservoir behavior was formulated with the following variables:  $Q_t^j$  inflow discharge,  $V_t^i$  stored volume,  $T^i$  target volume and  $Q_t^i$  outflow discharge. The bayesian network includes two types of conditional probabilities for causal relations  $P(V_t^i | V_{t-1}^i, Q_{t-1}^j, Q_{t-1}^i)$  and  $P(Q_t^i | Q_{t-1}^i, V_{t-1}^i, T^i, Q_{t-1}^j)$ . Note that this model uses the decision variable  $T^i$  (target volume) that describes the management strategy expressed as the desired volume in the reservoir.

Another application in this type of models is the interpretation of the prediction in terms of potential damages. For this purpose, additional relations were included to interpret the hydrologic values. There are two variables for each location with potential flood problems in the river basin:  $Q_t^j$  flow at location  $j$  and  $D_t^i$  damage level at location  $i$ . The values of  $D_t^i$  represent levels of problems with qualitative values such as normal, material damages, severe material damages, personal damages, severe personal damages, etc. The interpretation of the flow values in terms of problem levels is expressed by the conditional probability  $P(D_t^i | Q_t^j)$ .

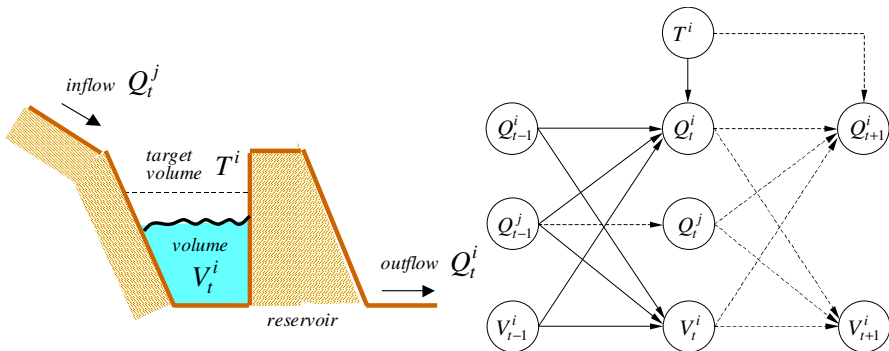


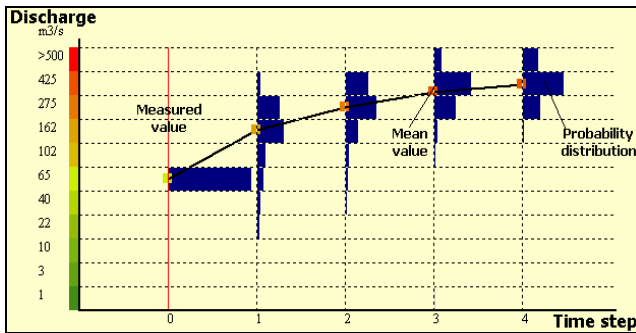
Fig. 2. Example of temporal extension for the bayesian network of the reservoir operation

In the context of prediction for decision support, it is normally required to make a forecast for several consecutive time steps. In order to perform this process, besides the spatial references of nodes corresponding to the specific locations of physical variables, a temporal extension is required as it is done in dynamic bayesian networks [3] [4]. For this purpose, the elementary bayesian network for each physical process is considered with additional nodes and causal relations corresponding to consecutive timeslices. Figure 2 shows this idea for the case of the reservoir operation. In dynamic bayesian networks, the first order Markov property indicates that the parents of a variable in timeslice  $t$  must occur in either slice  $t$  or  $t-1$ . This is a property that is not always satisfied by the hydrologic processes presented here<sup>1</sup>. For example, in the runoff concentration we have identified the causal relation  $P(Q_t^i | Q_{t-1}^i, N_{t-1}^i, \dots, N_{t-k}^i)$  (in the particular model for Guadalhorce river  $k=3$ ).

<sup>1</sup> Nevertheless, variables can be transformed to satisfy this property as it is described by [5].

### 3 Operation with the SAIDA Application

SAIDA is a computer system that was developed in a three-year project during 1998-2000 and promoted by the Spanish Ministry of the Environment with the purpose of operating in connection with the information hydrologic systems in several Spanish basins (details about the operation and the complete software architecture of SAIDA can be found at [6] [7] [8]). SAIDA receives as input the available data provided by sensors about discharge, water level and rainfall at different locations in the river basin. SAIDA provides answers that evaluate the current situation, predict a short term evolution and recommend control actions. The answers are produced with time constraints and the conclusions are justified at a reasonable level of abstraction given that the operator must take the final responsibility of decisions.

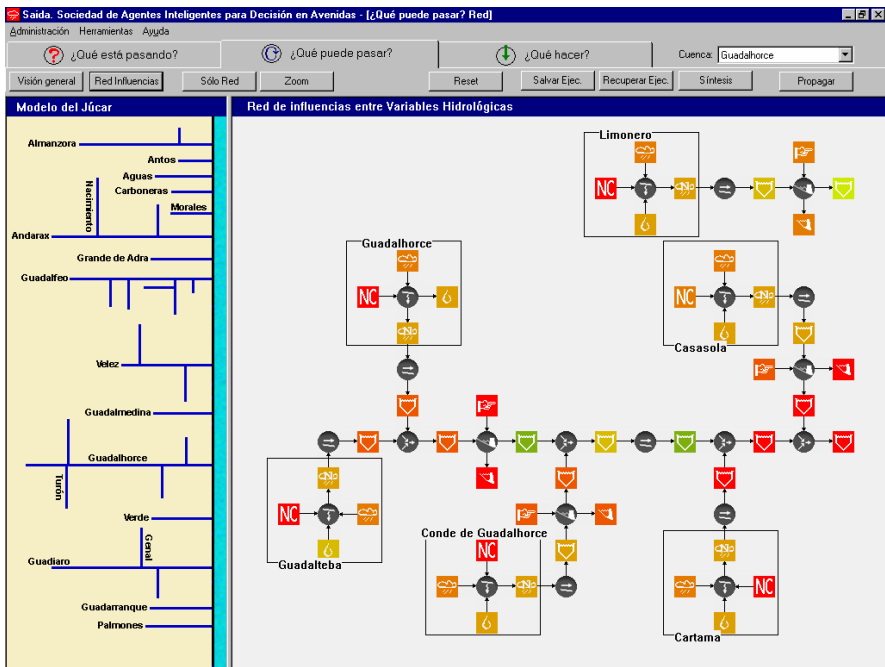


**Fig. 3.** Example of presentation of the predicted values for a variable in consecutive time steps

The bayesian approach was applied for the development of models for prediction as part of the SAIDA system. In this context, SAIDA receives as input: (1) *values recorded by sensors* about past and recent rainfall in different areas, current discharge at significant locations and water level in reservoirs, and (2) *hypotheses of future behavior*, i.e. the operator makes hypotheses of values for significant cause variables like future rain based on global meteorological information, future discharge policy of reservoirs (target volume), basin condition expressed in terms of model parameter values (e.g., curve number), etc.

SAIDA uses the model of bayesian networks to determine values about future volumes stored in reservoirs and flows at certain locations. The model also provides information about potential damages in areas at risk. All these values are expressed as probability distributions showing a range of potential behaviors according to model uncertainty. Figure 3 shows an example of how SAIDA shows the future evolution of a variable at certain location. The graphical representation shows the probability distribution associated to each time step. The mean values are explicitly connected to show the temporal trend of the variable. This graphical representation is a synthetic image that covers a wide range of potential behaviors for a particular variable taking into account the uncertainty of different processes.

In order to perform total predictions for the whole river basin the local bayesian models are connected and linked to the real-time hydrologic information network. Individual bayesian models are combined in a larger network, that connects the set of variables according to river basin topology. Inference is carried out with an adaptation of a general inference algorithm for multiply connected networks [9]. SAIDA shows a complete view about the causal relations in a global image (Figure 4). This global view corresponds to a summarized view of the instantiation of the type of bayesian networks described in the previous section for a particular river (e.g. the Guadalhorce River in Málaga). The model for a particular river basin is built by linking together several instances of the bayesian networks according to the topology of the river basin. Each specific bayesian network for a particular physical process at certain location presents differences (e.g., discrete values and conditional probabilities) compared to another network for the same process at a different location.



**Fig. 4.** Interactive analysis tool for hydrologic prediction provided by the SAIDA user interface with bayesian networks

The window of figure 4 shows a visualization using a color code for each variable that goes from the lowest value (green) to the highest value (red). Each node of the diagram corresponds to a physical process (runoff generation, reservoir operation, etc.). This provides a global image of the causal explanation of flows at different locations. The operator can consult individually the temporal evolution of input, output and intermediate variables displaying additional windows where the probability distributions for different time steps are presented. This user interface is actually an interactive analysis

tool where the user can also change the values of some of these variables to produce a new prediction. This feature is very useful to analyze different hydrologic scenarios in an appropriate level of abstraction in the presence of problematic situations.

## 4 Experimental Results

Following the previous approach several models were developed for the control centers located in Valencia and Málaga (Spain). This section describes details of the case of Málaga to show results about an experimental evaluation. Málaga is located in a flash-flood prone area, at the outlet of two rivers, Guadalhorce and Guadalmedina. Contributing areas to the Guadalhorce and Guadalmedina basins are of 3,158 km<sup>2</sup> and 147 km<sup>2</sup> respectively. The climate is semiarid, with steep slopes covered by brush at the headwaters and irrigated land at the floodplain. Several reservoirs have been built to regulate the Guadalhorce basin and to protect Malaga from flooding. The *Confederación Hidrográfica del Sur* is the management authority responsible for the operation of the reservoirs during floods.

Physical Process	Spatial location	Causal relations	CE	A
Runoff generation	Guadalhorce	$P(N_{t-1}^i   R_t^j, M_t^i, C^i)$	0.16	87
		$P(M_{t-1}^i   R_{t-1}^i, M_{t-1}^i)$	0.04	
	Guadalteba	$P(N_{t-1}^i   R_t^j, M_t^i, C^i)$	0.19	87
		$P(M_{t-1}^i   R_{t-1}^i, M_{t-1}^i)$	0.04	
	Conde de Guadalhorce	$P(N_{t-1}^i   R_t^j, M_t^i, C^i)$	0.15	86
		$P(M_{t-1}^i   R_{t-1}^i, M_{t-1}^i)$	0.05	
	Casasola	$P(N_{t-1}^i   R_t^j, M_t^i, C^i)$	0.19	87
		$P(M_{t-1}^i   R_{t-1}^i, M_{t-1}^i)$	0.04	
	Cártama	$P(N_{t-1}^i   R_t^j, M_t^i, C^i)$	0.18	87
		$P(M_{t-1}^i   R_{t-1}^i, M_{t-1}^i)$	0.04	
	Limonero	$P(N_{t-1}^i   R_t^j, M_t^i, C^i)$	0.17	87
		$P(M_{t-1}^i   R_{t-1}^i, M_{t-1}^i)$	0.04	
Runoff concentration	Guadalhorce	$P(Q_{t-1}^i   Q_{t-1}^i, N_{t-1}^i, N_{t-2}^i, N_{t-3}^i)$	0.2	89
	Guadalteba	$P(Q_{t-1}^i   Q_{t-1}^i, N_{t-1}^i, N_{t-2}^i, N_{t-3}^i)$	0.2	89
	C.Guadalhorce	$P(Q_{t-1}^i   Q_{t-1}^i, N_{t-1}^i, N_{t-2}^i, N_{t-3}^i)$	0.16	90
	Casasola	$P(Q_{t-1}^i   Q_{t-1}^i, N_{t-1}^i, N_{t-2}^i, N_{t-3}^i)$	0.22	89
	Cártama	$P(Q_{t-1}^i   Q_{t-1}^i, N_{t-1}^i, N_{t-2}^i, N_{t-3}^i)$	0.19	90
	Limonero	$P(Q_{t-1}^i   Q_{t-1}^i, N_{t-1}^i, N_{t-2}^i, N_{t-3}^i)$	0.16	90
Reservoir operation	C.Guadalhorce	$P(Q_{t-1}^i   Q_{t-1}^i, V_{t-1}^i, T^i, Q_{t-1}^i)$	0.54	78
		$P(V_{t-1}^i   V_{t-1}^i, Q_{t-1}^i, Q_{t-1}^i)$	0.01	99
	Casasola	$P(Q_{t-1}^i   Q_{t-1}^i, V_{t-1}^i, T^i, Q_{t-1}^i)$	0.05	97
		$P(V_{t-1}^i   V_{t-1}^i, Q_{t-1}^i, Q_{t-1}^i)$	0.04	98
	Limonero	$P(Q_{t-1}^i   Q_{t-1}^i, V_{t-1}^i, T^i, Q_{t-1}^i)$	0.18	91
		$P(V_{t-1}^i   V_{t-1}^i, Q_{t-1}^i, Q_{t-1}^i)$	0.07	96
River junction	Cartama Guadalhorce	$P(Q_{t-1}^i   Q_{t-1}^i, Q_{t-1}^i)$	0.12	93
	Cuadalhorce Campanillas	$P(Q_{t-1}^i   Q_{t-1}^i, Q_{t-1}^i)$	0.12	92
	Guadalhorce G., Conde G.	$P(Q_{t-1}^i   Q_{t-1}^i, Q_{t-1}^i)$	0.13	92
Discharge prop.	Guadalhorce	$P(Q_{t-1}^i   Q_{t-1}^i, Q_{t-1}^i, Q_{t-1}^i)$	0.45	82

**Fig. 5.** Experimental results for the model of the South of Spain (CE: conditional entropy, A: accuracy)



Data gathered from an automatic data collection network are analyzed at a control center to provide assistance to decision makers in selecting the best management strategies for reservoir operation and to issue warnings to Civil Defense authorities and to the population. Hydrologic information is received at one hour time intervals from 29 raingages, 5 reservoirs (Guadalhorce, Guadalteba, Conde de Guadalhorce, Casasola and Limonero) and from a gaging station in the Guadalhorce river located near Málaga, in Cártama.

A deterministic simulation model was taken as the basic framework to build the probabilistic decision model. Hydrological knowledge about a river basin is typically encoded in deterministic simulation models. A great deal of expert knowledge and effort is applied in model formulation, discretization and calibration, using information about the basin, field surveying and data from observed events. After the calibration process, the values of model parameters are only partially known, and they are best described by a confidence interval or a probability distribution.

The deterministic model was run with random parameters and forced with a stochastic rainfall simulator, creating a large database of synthetic storms. During the simulations, parameter values were sampled randomly from their estimated probability distributions to obtain an ensemble of basin behaviors consistent with the results of the calibration process. The database of simulated events contains a variety of basin behaviors expressed in numerical values that were converted to the discrete domains of the bayesian network variables. The qualitative time series generated were processed to collect cases as combinations of values for cause variables and the corresponding value for the effect variable.

The resulting models were validated to determine their ability to produce probability distributions that describe accurately the behavior of the deterministic model and are useful for decision making. Two different types of model evaluation were performed: (1) evaluation of bayesian network structure and (2) evaluation of prediction quality. The first type of evaluation was useful to compare different versions of structures of bayesian networks and discrete domains. This evaluation was applied to different versions of bayesian networks that were accordingly refined until a satisfactory version was obtained. In order to evaluate the structure of each bayesian network, the *conditional entropy*  $H$  is computed with the following equation [10]:

$$H(X/Y_1, \dots, Y_n) = - \sum_{Y_1=y_1, \dots, Y_n=y_n} P(y_1, \dots, y_n) \sum_{X=x} P(x | y_1, \dots, y_n) \ln P(x | y_1, \dots, y_n)$$

where  $X$  represents a node of the network,  $Y_1, \dots, Y_n$  is the set of parent nodes of  $X$  and  $Y_j = y_j$  expresses that the variable  $Y_j$  gets the qualitative value  $y_j$ . This parameter estimates the disorder of information, so lower values are considered better results.

The evaluation of the network prediction quality was computed with the *accuracy* parameter. The *accuracy* parameter  $A$  evaluates the quality of the answers of the bayesian network. This parameter is computed with the formula:

$$A = \frac{\sum P_i}{N} 100$$

where,  $i$  designates a case,  $P_i$  is the probability assigned by the bayesian network to the corresponding value of the effect variable of case  $i$ , and  $N$  is the total number of cases.

Bayesian models were calibrated with a set S1 of about 300,000 cases produced by simulation. Another different set S2 with the same number of cases was generated for evaluation of model performance. The number of cases in these sets was adjusted verifying that all combinations of discrete values for each set of nodes cause in S2 should be also be present in S1. This guarantees that the bayesian network learned from S1 includes all the physically possible situations (this requirement was efficiently verified with the help of a particular data structure for the bayesian network that included the combinations derived from S1).

The evaluation of the bayesian network structure was applied to different versions of structures for bayesian networks with different discrete values that were refined until a satisfactory version was obtained. The resulting final values of evaluation parameters for the case of the Guadalhorce and Guadalmedina basins are shown in figure 5. As shown in the table, all local bayesian networks have a good behavior for the degree of accuracy in accordance to each model level of uncertainty. The resulting values for these parameters after the evaluation process prove that the bayesian networks provide a satisfactory behavior.

## 5 Conclusions

The approach for hydrologic prediction presented in this paper is a practical solution to be used in a context of real-time decision support. The proposed model is a case of spatio-temporal bayesian network, i.e. a bayesian network where nodes have both spatial and temporal references. This is a solution that facilitates rational decisions in probabilistic terms as it required in the field of hydrology about future states of a river basin [11].

A number of solutions have been proposed to generate probabilistic forecasts using deterministic models [12] [13] [14]. However, these solutions show a high degree of artificial mathematical sophistication that makes that, from a practical point of view, a computer system with this approach operates like a black box. In a decision context, on the contrary, it is very important to use a natural representation model closer to the background of decision makers, in order to build confidence in the results produced by the system. Bayesian networks as presented in this paper provide a natural and intuitive description of hydrologic processes based on a symbolic representation with qualitative variables and causal relations. This is very useful to formulate decision models with high levels of abstraction and explicit meaning. The bayesian representation shows explicitly the uncertainties of the information, which is a novelty, compared to classical deterministic models (e.g. [1] [2]). This feature is useful to show explicitly the degree confidence that the system gives to its own answers. This task is normally performed by operators who give partial credibility to the answers of deterministic simulation models according to their experience with those tools.

Bayesian models can be automatically created using information currently available in flood control centers. For example, these types of models can take advantage

of the knowledge about the river basin encoded in a classical deterministic simulation model but also they can easily take advantage of historical information recorded in control centers (e.g., in Valencia the SAIH infrastructure has been recorded hydrological data of near 20 years). This feature favors the transfer of the technology to the operational stage.

The experimental evaluation of the bayesian networks associated to hydrologic processes with data obtained from the river basins in the South of Spain (Guadalhorce and Guadalmedina) showed a satisfactory performance for prediction. This approach was applied to develop part of a software environment called SAIDA which besides the capability of prediction using bayesian networks includes additional features (identification of problem scenarios, recommendation of hydraulic actions, etc.).

Bayesian networks have also been applied in the field of meteorology [15] [16] but, to our knowledge, our approach to model physical processes in the field of hydrology is an original contribution. The success of this development suggests to continue with this work in the following lines: (1) a more extensive use for new river basins in different parts of Spain with additional physical processes (for this purpose the Spanish Ministry of Environment is currently opening a new project), (2) according to the particular type of dynamic bayesian network, alternative inference methods can be applied to gain efficiency, and (3) automatic tools can be designed to facilitate the construction of models (with a suite of software tools for model edition, simulation, and machine learning).

**Acknowledgements.** The development of the SAIDA system was mainly supported by the Ministry of Environment of Spain (*Dirección General de Obras Hidráulicas y Calidad de las Aguas*) and local public organizations from river basins (*Confederación Hidrográfica del Júcar* and *Confederación Hidrográfica del Sur de España.*) with the collaboration of the private companies SYNCONSULT and PAGESEI. It was also partially supported by the Ministry of Science and Technology of Spain within the RIADA Project.

## References

1. Brath, A., Rosso, R.: "Adaptive calibration of a conceptual model for flash flood forecasting". *Water Resources Research*, 29(8) 2561-2572, 1993.
2. Madsen, H.: "Automatic calibration of a conceptual rainfall--runoff model using multiple objectives" *Journal of Hydrology*, (235)3-4, 276-288, 2000.
3. Dean T., Kanazawa K.: "A model for reasoning about persistence and causation". *Computational Intelligence*, 5(3): 142-150. 1989.
4. Z. Ghahramani. "Learning dynamic Bayesian network". In C.L. Giles and M. Gori, editors, *Adaptive Processing of Sequences and Data Structures*, volume 1387 of *Lecture Notes in Computer Science*, pages 168-197. Springer, 1998.
5. Murphy, K. P., "Dynamic Bayesian Networks: Representation". Inference and Learning, Ph.D. thesis, UC Berkeley, Computer Science Division, July, 2002.
6. Cuenca, J. Molina, M.: "A Multi-agent System for Emergency Management in Floods", in "Multiple Approaches to Intelligent Systems", Iman I., Kodratoff Y. (eds.). *Lecture Notes in Artificial Intelligence*, Springer, 1999.

7. Molina M., Blasco G: "A Multi-agent System for Emergency Decision Support". Proceedings of the Fourth International Conference Intelligent Data Engineering and Automated Learning IDEAL 03. LNCS, Springer, 2003.
8. Garrote L., Molina M.: "A Framework for making probabilistic forecast using deterministic rainfall-runoff models". Proceedings of the ESF LESC Exploratory Workshop held at Bologna, Italy, October 24-25, 2003.
9. Lauritzen S. L., Spiegelhalter D. J.: "Local computations with probabilities on graphical structures and their application to expert systems". *Journal of the Royal Statistical Society B* 50(2): 157-224, 1988.
10. Herskovitz E.H., Cooper G.F.: "Kutat6: an entropy-driven system for the construction of probabilistic expert systems from data". Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, pp 54-62, 1990.
11. Krzysztofowicz R.: "The case for probabilistic forecasting in hydrology". *Journal of Hydrology*, 249, 2-9, 2001.
12. Georgakakos K. P., Bras R. L.: "A Hydrologically Useful Station Precipitation Model 1. Formulation". *Water Resources Research*, 20, 1585-1596, 1984.
13. Lardet P., Obled C.: "Real-time flood forecasting using a stochastic rainfall generator". *Journal of Hydrology*, Volume 162, Issues 3-4, Pages 391-408, November, 1994.
14. Krzysztofowicz R.: "Bayesian theory of probabilistic forecasting via deterministic hydrologic model". *Water Resources Research*, 35(9) 2739-2750, 1999.
15. Kennett R., Korb K., Nicholson A.: "Seabreeze Prediction Using Bayesian Networks: A Case Study". Proceedings of the 5<sup>th</sup> Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining PAKDD, Springer Verlag, 2001.
16. Cano R., Sordo C, Gutiérrez J.M.: "Applications of Bayesian Networks in Meteorology" In J.A. Gámez, S. Moral and A. Salmer6n, eds., *Advances in Bayesian Networks*, 309-327, 2004 Springer Verlag, 2004.