BAYESIAN MODEL OF URBAN WATER SAFETY MANAGEMENT

B. TCHORZEWSKA-CIEŚLAK

1Department of Water Supply and Sewage Systems, Rzeszow University of Technology, 35-950 Rzeszow, Al.Powstancow Warszawy 6, Poland
e-mail: cbarbara@prz.edu.pl

EXTENDED ABSTRACT

Water supply system is a critical infrastructure. Main task of urban water system is to provide consumers with drinking water in adequate quantity, at the required quality and pressure corresponding to current standards.

For the purposes of this paper, operational reliability of the water supply system is defined as the ability to supply a constant flow of water for various groups of consumers, with a specific quality and specific pressure, according to consumers demands, in specific operational conditions, at any or at a specific time.

The main aim of this paper is to present a method for risk analysis using Bayesian process. The proposed method made it possible to estimate risk associated with the possibility of partial or total loss of the ability of water supply system operation. The paper proposes to consider two types of risk: the first type, associated with the possibility of interruptions in water supply and the second type, associated with the possibility of tap water contamination.

Urban water safety management system is introduced on the level of the local water companies. Risk management is part of a modern and well-developed system of safety management of water supply systems. It is a multi-step procedure aimed at improving the system safety, including quantitative and qualitative aspects of drinking water. This process is based primarily on the risk analysis, risk assessment or risk estimation, making decision on its acceptability, periodic control or reduction. Risk identification is based on a selection of representative emergency events that may occur during continuous operation of WDS, including initiating events that could cause the so-called domino effect [9, 10].

Keywords: water supply systems, risk analysis, Bayes model

1. INTRODUCTION

The Bayesian networks - BRA (Bayes Risk Analysis) are used in risk analysis due to the ability to model the dependent events. The Bayesian network is upgraded by means of experience and acquired knowledge. The network is modelled by a directed acyclic graph in which vertices represent events and edges represent causal connections between these events. The occurrence of the event $X_j$ (cause) has some impact on the occurrence of the event $X_i$ (effect). If this impact is not "certain" and can only be determined by the probability, then such an arrangement of events and the relation between them can be modelled by a directed graph $D$ [3,11,12]. Each event is represented as a graph vertex. Relations between events are represented by edges. If the occurrence of the event $X_j$ has some impact on the occurrence of the event $X_i$ ($X_i$ depends on $X_j$), then there is an edge $(X_j, X_i)$ in the graph model, exiting the $X_j$ and entering the $X_i$ (direction is indicated by the arrow). The vertex $X_j$ is called 'parent' of the vertex $X_i$. The set of all 'parents' of the vertex $X$ is marked as $\pi(X)$. Most often every event is identified with the corresponding random variable having the same name, on the assumption that all the random variables corresponding to the events are divalent (1 - an event that occurs, 0 – an event opposed to the event that occurs). The relations between the vertices (events) are expressed by means of the conditional probability. For the vertex $X$, whose parents are in the set $\pi(X)$,
these relations are represented by the conditional probability tables (CPT). In CPT, for
the variable $X$, all the probabilities $P(X|\pi(X))$ (for all the possible combinations
of variables from the set $\pi(X)$) must be specified. The table for the vertex that does not have
parents includes the probabilities that the random variable $X$ will take its particular values.
If the network has $n$ vertices, $X_1, ..., X_n$, the total probability distribution of all the random
variables is shown as the relation $[2]$: 

$$
P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | X_{\pi(i)})$$

To determine the total probability distribution without using the Bayesian network it is
necessary to know all the values of $P(X_1, ..., X_n)$ for all the possible combinations
of variables $X_1, ..., X_n$, which gives $2^n$ values of the probabilities. Using the Bayesian
network it is sufficient to know the conditional probabilities for each vertex.

At the macro scale, safety concerning water supply is defined as a state of water
management that allows to cover current and future customers demands for water, in a
technically and economically justified way, and by the requirements for the protection of
the aquatic environment $[8,13]$. The primary and basic subject to which the notion of
water safety is concerned is a consumer. The secondary subject is a supplier – a
manufacturer of water. In this respect, one can consider the risk of the consumer and the
risk of the producer $[8,13]$. The important elements in this regard are also the
environmental aspects and the principles of sustainable development in widely
understood water management. The definition of water supply safety is the following:
"safe operation of water supply systems means ensuring continuity of water supply to the
consumer while the following criteria are met: system reliability (for quantity and quality),
socially acceptable level of prices per m$^3$ of delivered water, taking into account the
aspects arising from the requirements for public safety, natural aquatic environment
protection and the standard of quality of life $[7,8,13]$.

2. ANALYSIS OF WATER CONSUMERS RISK

Consumer's risk (individual) $r_K$ is the sum of the first kind risk associated with the
possibility of interruptions in water supply, and the second kind risk associated with the
consumption of poor quality water $[13]$. Consumer's risk is a function of the following
parameters: probability $P$ or frequency $f$ of undesirable events in water distribution
subsystem which are directly felt by water consumers, related losses $C$ (e.g. purchase of
bottled water, possible medical expenses after consuming unfit for drinking water or
immeasurable losses, such as living and economic difficulties and loss of life or health),
the degree of vulnerability $V$ to undesirable events $[1,4,5,8,13]$. Consumer's risk $R_C$ is
given by the formula:

$$
R_C = R_{CI} + R_{CII}
$$

where:

$R_{CI}$ – the risk of the first type,
$R_{CII}$ – the risk of the second type

For the risk of the first type, associated with quantity of supplied water, and for the risk of
the second type, associated with quality of supplied water , the three parametric definition
was assumed:

$$
R_{CI,II} = P_{I,II} \cdot C_{I,II} \cdot V_{I,II}
$$

where:
$P_{I,II}$ - likelihood of event occurrence that may cause the risk of the first type or the risk of the second type,

$C_{I,II}$ - losses caused by the undesirable event that may cause the risk of the first type or the risk of the second type,

$V_{I,II}$ - vulnerability associated with the occurrence of the undesirable event that may cause the risk of the first type or the risk of the second type.

3. RISK ANALYSIS OF WATER DISTRIBUTION SUBSYSTEM FAILURE USING THE BAYESIAN NETWORK

The Bayesian network can be used in the decision-making model analysing the risk of failure in urban water system (UWS). Risk analysis model for the risk of the first type and the risk of the second type that can be used in making decisions by water supply companies (for the modernization or repairs), was developed. This model was inserted into the calculation program JavaBayes [13]. In Figures 1 and 2 the developed Bayesian network schemes, used for failure risk analysis of water distribution subsystem, from the water consumer point of view, are presented.

![Figure 1](image1.png)

**Figure 1.** Bayesian network for the risk of the first type (JavaBayes Widow).

![Figure 2](image2.png)

**Figure 2.** Bayesian network for the risk of the second type (JavaBayes Widow).
Symbols used in Figure 1 and 2 mean:
- $R_{CI,II}$ – consumer's risk (the first or second type) in a five-point scale: $R_{CI,II} = \{r_{C1,II}, r_{C2,II}, r_{C3,II}, r_{C4,II}, r_{C5,II} \}$
  - neglected risk $r_{C1,II}$
  - tolerable risk $r_{C2,II}$
  - controlled risk $r_{C3,II}$
  - intolerable risk $r_{C4,II}$
  - unacceptable risk $r_{C5,II}$
- $X_1$ – interruption in water supply:
  - $X_{11}$ – failure of the water supply network,
  - $X_{12}$ – lack of water supply from the water treatment plant,
  - $X_{13}$ – failure of zone pumping stations.
- $X_2$ – consumers protection from the existing threat:
  - very little – $x_{21}$,
  - little – $x_{22}$,
  - medium – $x_{23}$,
  - large – $x_{24}$,
  - very large – $x_{25}$,
- $X_3$ – water quality parameters specified in the relevant Regulations of the Minister of Health are exceeded:
  - $X_{31}$ – physico-chemical parameters are exceeded,
  - $X_{32}$ – microbiological parameters are exceeded.

The following assumptions were made:
- the event in the given node takes exactly one of the possible values,
- 1 means that the event occurs, 0 - that the event does not occur

For each vertex the CPT should be defined:
For the risk of the first kind (fig. 1):
- $P(R_{CI} \mid X_1, X_2)$,
- $P(X_1 \mid X_{11}, X_{12}, X_{13})$,
- $P(X_2)$,
- $P(X_{11})$,
- $P(X_{12})$,
- $P(X_{13})$,
For the risk of the second kind (fig. 2)
- $P(R_{CII} \mid X_2, X_3)$,
- $P(X_3 \mid X_{31}, X_{32})$,
- $P(X_{31})$,
- $P(X_{32})$.

Using the formulas (1) the values of probability for each risk category (the first or second kind), i.e. $P(R_{CI,II} = r_{C1,II}, r_{C2,II}, r_{C3,II}, r_{C4,II}, r_{C5,II})$, were calculated. An example of reasoning is as follows: the probability that the consumer's risk of the first type is neglected is the sum of the product of the conditional probabilities $P(R_{CI} = r_{C1} \mid (X_1 \land X_2))$ (the probability that $r_{C1}$ is $r_{C1}$, if event $X_1$ ($X_1 = 1,0$) and events $X_2$ ($X_2 = x_{21}, x_{22}, x_{23}, x_{24}, x_{25}$)) and the probabilities for events $X_1$ and $X_2$: $P(X_1) \cdot P(X_2)$. The summation is performed for all the possible values of $X_1$ and $X_2$, according to the general formula [13]:

$$P(R_{CI} = r_{C1,II}, r_{C2,II}, r_{C3,II}, r_{C4,II}, r_{C5,II}) = \sum_{X_1=1}^{X_1} P(R_{CI} = r_{C1} \mid X_1 = 1 \land X_2 = j) \cdot P(X_1 = 1) \cdot P(X_2 = j)$$

(5)

where:
- $i$ – event $X_1$ occurs or does not occur ; $i = 1,0$, 

- $j$ – event $X_2$ occurs or does not occur ; $j = 1,0$, 

- $r_{C1,II}, r_{C2,II}, r_{C3,II}, r_{C4,II}, r_{C5,II}$ – risk categories.
j – given value of event \( X_j \);

For example, that the probability that the consumer's risk of the first kind is neglected, according equations (5) we obtain:

\[
P(r_{CI} = r_{CI1}) = P(r_{CI} = r_{CI1} \mid X_1 = 1 \land X_2 = x_{21}) \cdot P(X_1 = 1) \cdot P(X_2 = x_{21}) + P(r_{CI} = r_{CI1} \mid X_1 = 1 \land X_2 = x_{22}) \cdot P(X_1 = 1) \cdot P(X_2 = x_{22}) + \ldots
\]

The rest of the values are calculated in the same way. The developed model allows to determine the probability of the particular risk level. The result of modelling are the probability values for each risk level. The risk assessment is based on the interpretation of the result (risk with the highest and the lowest probability of occurrence is given). For example, the result shows that for the first type of risk the highest probability is for a tolerable level and the lowest for the unacceptable level. The developed model enables also determining the partial probabilities for events included in the defined Bayesian network.

### 4. The application example

The analysed water supply system is located in the south-eastern Poland. Surface water intake from the river. The system is currently used by about 200 thousand residents of the city and nearby towns. The average daily treated water production is about 34 600 m³/d and it fully meets the customers demand for water. The water supply system with a total length of about 840 km. 80% of the network operates in a closed system. Two teams of water tanks located in the eastern and western part of the city work with the water pipe network. To perform risk analysis in terms of water consumers safety, the developed model using the Bayesian networks was applied. Calculations were performed by means of JavaBayes, using the model for the risk of the first and second kind. For each vertex of the Bayesian network (Figs. 1 and 2) the CPT were defined (table 1-7) [13].

<table>
<thead>
<tr>
<th>Nr</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( P(r_{CI} \mid X_1, X_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.05 0.05 0.2 0.3 0.4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1 0.05 1 0.3 0.2 0.1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0 0.2 0.3 0.2 0.1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0.3 1 0.4 0.2 0.1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.4 0.3 0.3 0 0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0.6 0.4 0 0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.65 0.35 0 0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.7 0.3 0 0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0.8 0.2 0 0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.9 0.1 0 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( P(X_1 \mid X_{11}, X_{12}, X_{13}) )</th>
<th>( X_{11} )</th>
<th>( X_{12} )</th>
<th>( X_{13} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.05</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.005</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.245</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3. The CPT for $X_{11}$, $X_{12}$, $X_{13}$

<table>
<thead>
<tr>
<th>$X_2$</th>
<th>$P(X_2 = 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{21}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$X_{22}$</td>
<td>0.5</td>
</tr>
<tr>
<td>$X_{23}$</td>
<td>0.14</td>
</tr>
<tr>
<td>$X_{24}$</td>
<td>0.05</td>
</tr>
<tr>
<td>$X_{25}$</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 4. The CPT for $X_2$

<table>
<thead>
<tr>
<th>$X_3$</th>
<th>$X_31$</th>
<th>$X_32$</th>
<th>$P(X_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = 1)$</td>
<td>0.97</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>$P(X = 0)$</td>
<td>0.02</td>
<td>0.98</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. The CPT for $r_{CII}$

<table>
<thead>
<tr>
<th>Nr</th>
<th>$X_31$</th>
<th>$X_32$</th>
<th>$X_33$</th>
<th>$X_34$</th>
<th>$X_35$</th>
<th>$r_{CII}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>7</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>8</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>9</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 6. The CPT for $X_3$

<table>
<thead>
<tr>
<th>$X_31$</th>
<th>$X_32$</th>
<th>$P(X_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = 1)$</td>
<td>0.09</td>
<td>0.90</td>
</tr>
<tr>
<td>$P(X = 0)$</td>
<td>0.91</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 7. The CPT for $X_{31}$ and $X_{32}$

5. RESULTS
The calculation results are presented in Table 8.

Table 8 The results of the first kind risk analysis and the second kind risk analysis performed using the Bayesian network

<table>
<thead>
<tr>
<th>Risk scale</th>
<th>$r_{KI}$</th>
<th>$r_{KII}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>neglected</td>
<td>0.630</td>
<td>0.583</td>
</tr>
<tr>
<td>tolerable</td>
<td>0.340</td>
<td>0.329</td>
</tr>
<tr>
<td>controlled</td>
<td>0.011</td>
<td>0.022</td>
</tr>
<tr>
<td>intolerable</td>
<td>0.010</td>
<td>0.034</td>
</tr>
<tr>
<td>unacceptable</td>
<td>0.0095</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

Risk analysis, made with the adopted assumptions, for the analysed water supply system, both in terms of the lack of water supply and the possibility of consuming water with parameters inconsistent with the existing regulation, shows that risk is at the neglected level.
5. CONCLUSIONS
The proposed method of risk analysis using the Bayesian network is mainly used in decision-making processes. The Bayesian network shows the cause-effect relation of events. Using the developed method, one obtains information on the level of risk (in the adopted scale) and its probability. In this way, the proposed model can be an important element in the decision making by the subsystem operator. Models can be modified for all the elements of the water supply system. Two models have been developed, for the first kind risk analysis and for the second kind risk analysis. The program also allows to determine the likelihood of the intermediate events included in the Bayesian network.

ACKNOWLEDGEMENTS
Scientific work was financed from the measures of National Center of Research and Development as a development research project No N R14 0006 10: “Development of comprehensive methodology for the assessment of the reliability and safety of water supply to consumers” in the years 2010-2013.

REFERENCES