

# 10

---

## Optimization of Enhanced Coagulation in Water Treatment using Bayesian Networks

Zoe J. Y. Zhu, Edward A. McBean, and Hongde Zhou

As a result of the potential health effects of disinfection byproducts, there is extensive interest in alternative treatment technologies which lessen their formation. The potential for enhanced coagulation to improve the removal efficiency of organic matter and thus decrease the formation of disinfection byproducts is identified. A Bayesian Network model is used to simulate different coagulation conditions for purposes of determining optimal conditions for enhanced coagulation. The causal dependence relationships amongst the variables (e.g. coagulant, pH, temperature, etc.) are encoded in a Bayesian network structure to identify the most advantageous configuration of treatment options.

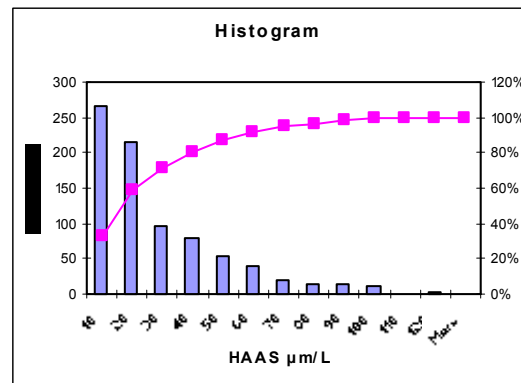
### 10.1 Introduction

Concerns regarding the potential health effects of DBPs have resulted in a number of regulations promulgated by the United States Environmental Protection Agency (USEPA, 1998). For example, the Disinfection/Disinfection By-Product (D/DBP) Rule (Rule) has set maximum contaminant levels for total trihalomethanes (TTHMs) and total

Zhu, Z.J., E. McBean and H. Zhou. 2006. "Optimization of Enhanced Coagulation in Water Treatment using Bayesian Networks." *Journal of Water Management Modeling* R225-10. doi: 10.14796/JWMM.R225-10.  
© CHI 2006 www.chijournal.org ISSN: 2292-6062 (Formerly in Intelligent Modeling of Urban Water Systems. ISBN: 0-9736716-2-9)

haloacetic acids (HAAs) of 0.080 and 0.060 mg/L, respectively with the potential for lowering to more stringent limits in the future. The Rule mandates utilities using disinfectants, to reduce NOMs (as measured by TOC) as a means of reducing DBP precursors. As depicted in Figure 10.1 as an example, exceedance of the 0.060 mg/L guideline occurs more than 10% of the time in Ontario water treatment systems (Zhu and McBean, 2004). In response to the potential health effects of the DBPs, there is extensive interest in alternatives which lessen their formation. Such interest includes “Enhanced Coagulation”, which involves the formation of precipitates which may adsorb the organic matter, and hence decrease the formation mechanisms of the DBP precursors.

The four conventional coagulation mechanisms for water treatment are (i) double layer compression, (ii) charge neutralization, (iii) interparticle bridging, and (iv) precipitate enmeshment. Historically, the conventional coagulation processes were designed primarily for particulate/turbidity removal. However, it is now recognized that organic matter may impact the quality of drinking water, especially via the formation of disinfection byproducts (DBPs). As well, when ozone reacts with natural organic matter (NOM), components such as humic and fulvic acids, and biodegradable organic matter (BOM) may be produced, promoting bacterial regrowth and coliform occurrences within water distribution systems. The mechanisms implicit in enhanced coagulation would improve standard coagulation, and be followed by conventional precipitation and adsorption.



**Figure 10.1** Frequency of occurrence of total HAAs DBPs in Ontario water systems.

However, while the USEPA has recognized the merit of enhanced coagulation/softening for controlling precursors of DBPs (USEPA, 1998a), determining the optimal conditions for enhanced coagulation requires evaluation of a number of factors which influence the effectiveness including, but not limited to, the pH, the coagulant dose, and the temperature, for each given Total Organic Carbon (TOC) concentration. Tests to identify the optimum coagulation conditions involve jar tests.

As an alternative, a model that incorporates the causal relationships amongst these factors, and simulates different coagulation conditions can be used to decrease the costs associated with jar tests. Bayesian Networks (BN) provide one such computational model as described herein, by assisting in identifying optimal conditions for enhanced coagulation by encoding the dependencies among key variables in the coagulation process.

## 10.2 Overview of Enhanced Coagulation

To build a computational model of coagulation, the interrelationships among the relevant factors must be characterized. The removal of NOM in conventional water treatment processes by the addition of coagulants has been demonstrated by laboratory research, by pilot-demonstration, and by full-scale studies. Researchers have shown that TOC in water, used as an indicator of NOM, exhibits a wide range of responses to treatment with aluminum and iron salts (as per Chowdhury et al., 1997; Edwards, 1997; Owen et al., 1996; Krasner and Amy, 1995; Owen et al., 1993; James M. Montgomery, 1992; Hubel and Edzwald, 1987; Duan and Gregory, 1996; Knocke et al., 1986; Chadik and Amy, 1983; Semmens and Field, 1980; Young and Singer, 1979; and Kavanaugh, 1978). The majority of these studies have been conducted, using commercial or reagent-grade alum ( $\text{Al}_2[\text{SO}_4]_3 \cdot 14\text{H}_2\text{O}$  and  $\text{Al}_2[\text{SO}_4]_3 \cdot 18\text{H}_2\text{O}$ , respectively) as the coagulant, but iron salts have also been shown to be effective in removing TOC from water. Polyaluminum chloride (PACl) and cationic polymers also can be effective in removing TOC (Edzward, and Tobiason, 1999). Cationic polymers (as well as anionic and non-ionic polymers) have proven to be valuable in creating settleable floc when high dosages of aluminum or iron salts are used.

Vickers et al. (1995) used membrane filtration (MF) in conjunction with coagulation processes for improved NOM removal. They indicate that NOM removal, and subsequent control of the processes to minimize DBP

formation can optimize enhanced coagulation practices and result in a finished water of higher quality. Ma and Wei (2001) demonstrated enhanced removal of algae using potassium ferrate (VI) (ferrate) pre-oxidation followed by coagulation with alum. The pretreatment with ferrate increased remarkably the algae removal even for very short times of pre-oxidation. The removal effectiveness was further increased using a prolonged pretreatment time. Pretreatment with ferrate resulted in a reduction of alum dosage required to cause an efficient coagulation for algae removal.

Efficient removal of algae by potassium ferrate pre-oxidation is believed to be a consequence of several process mechanisms. Ferrate pre-oxidation inactivates algae and induces the formation of a coagulant aid, cellular components secreted by algal cells. The coagulation was also improved by increasing particle concentration in water because of the formation of the intermediate forms of precipitant iron species during pre-oxidation. In addition, it was also observed that ferrate pre-oxidation caused algae agglomerate formation before the addition of coagulant and the subsequent application of alum resulted in further coagulation.

Lee et al. (2000) addressed the issue of the relationship between membrane permeability and coagulation under various coagulation conditions such as alum dosage and pH, using both dead-end and cross-flow membrane filtration (MF) systems. The membrane permeability depended upon the coagulation condition because it gave rise to different specific cake resistance. The lower specific cake resistance was attributed to the formation of a less compressible, but more porous cake, with the charge neutralization condition. Under the dead-end MF, the effect of coagulation was demonstrated by the rising rate of transmembrane pressure at constant flux. According to Lee et al. (2000), membrane permeability with charge neutralization is much better than the sweep-floc mechanism. The permeate water quality was examined in terms of removal of natural organic matter ( $UV_{254}$ , TOC) and residual aluminum concentration. Lee et al. (2000) also addressed the issue of optimal particle size. In cross flow MF, particle transport depends on two main actions: the moving of the particles toward the membrane surface (negative direction, the permeation drag) and then shift away from the membrane surface (positive direction, the Brownian diffusion, shear-induced diffusion, and lateral inertial lift, etc).

Particle movement depends on the particle size. Particles in the range of 0.1-1 $\mu$ m have a strong tendency to move toward the membrane; thus they are the major foulants through internal fouling ( $R_f$ ) and cake deposit ( $R_c$ ) in the MF raw water. These particles were also reported to be the major flocculant for flux decline. Particles larger than 45 $\mu$ m could not approach

the membrane surface and thus do not cause the flux decline. Small particles (0.1 - 1 $\mu$ m) are responsible for the membrane flux decline and hence enhanced coagulation with alum might be the solution to remove the small particles. Lee et al. (2000) observed the membrane permeability largely depends on the coagulation condition as it influences specific cake resistance. Table 10.1 shows, under several coagulation conditions, the percent removal of turbidity and natural organic matter (UV<sub>254</sub>, TOC) by MF. Edzwald and Tobiason (1999) provided a broader view of enhanced coagulation, indicating that enhanced coagulation has multiple objective responses, namely: (i) maximization of particle and turbidity removals by downstream solid-liquid separation, (ii) maximization of removal of TOC and DBP precursors, (iii) minimization of residual coagulant, (iv) minimization of sludge production, and (v) minimization of operating costs.

James and John (1999) proposed a two-step process to determine enhanced coagulation requirements for waterworks as detailed below.

**Table 10.1** Water quality of feed and permeate in alum coagulation-MF system.

Coagulation condition	MF without coagulation	Coagulation-MF (10mg/L alum pH 5 charge neutralization)	Coagulation-MF(30mg/L alum, pH 7.5, sweep-floc)
Turbidity (NTU)/feed	3.5	2.0	1.6
Turbidity (NTU)/permeate	0.13	0.13	0.11
Turbidity (NTU)/removal (%)	96.3	93.5	93.1
UV <sub>254</sub> , (cm-1)/feed	0.058	0.044	0.04
UV <sub>254</sub> (cm-1)/permeate	0.036	0.02	0.019
UV <sub>254</sub> (cm-1)/removal (%)	37.9	54.5	52.5
TOC, (mg/L)/ feed	2.10	2.01	2.04
TOC, (mg/L) /permeate	1.90	1.30	1.56
TOC, (mg/L) /removal (%)	9.52	35.3	23.5
Aluminum (mg/L)/permeate	0.042	0.045	0.145

Ref. Lee, et al. (2000)

*Step 1. TOC removal requirements are summarized in Table 10.2.*

The TOC percent removals necessary are dependent upon raw water TOC concentration and alkalinity. The alkalinity is an economic consideration based on the use of alum or ferric salts. Water utilities monitor the TOC of their source and treated waters to determine TOC removals. Compliance is based on a running average, computed quarterly from monthly sampling.

Water systems that meet or exceed the Step1 requirements listed in Table 10.2 are demonstrating enhanced coagulation.

**Table 10.2** Enhanced coagulation: required TOC percent removals.

Raw water TOC (mg/L)	Raw water Alkalinity (mg/L CaCO <sub>3</sub> )		
	<60	60-120	>120
>2-4	35	25	10
>4-8	45	35	25
>8	50	40	30

Ref. Edzward and Tobiason et al. (1999)

*Step 2: Alternative TOC performance criteria using jar tests or pilot studies.*

If alum is the coagulant, alum is dosed in 10mg/L increments, and TOC measurements made on raw and settled waters. For ferric coagulants, the dosage increments are 9.1 mg/L for FeCl<sub>3</sub>·6H<sub>2</sub>O and 9.5mg/L for Fe<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub>·6H<sub>2</sub>O. The pH conditions depending on raw water alkalinity are as indicated per Table 10.3. A second step, Step 2, is needed when it is technically infeasible to meet the Step 1 TOC removal requirements. The Step 2 process does not apply for coagulants such as polyaluminum chlorides (PACls) or dual coagulation with organic cationic polymers, a shortcoming of Step 2.

The interdependencies indicate that Step 2 TOC removal criteria are established by determining the alum dosage. When additional 10mg/L increments (equivalent dosages for ferric coagulants) do not decrease the TOC by at least 0.3mg/L or a coagulant dosage is used where the settled water specific UV absorbance (SUVA) value is less than 2 L/mg.m, the result is a “point of diminishing return”. The point of diminishing return defines an alternate enhanced coagulation level and establishes alternative minimum TOC removal criteria for demonstrating compliance.

**Table 10.3** pH conditions for jar tests.

Alkalinity Range (mg/L as CaCO <sub>3</sub> )	pH
<60	5.5
60-120	6.3
120-240	7.0
>240	7.5

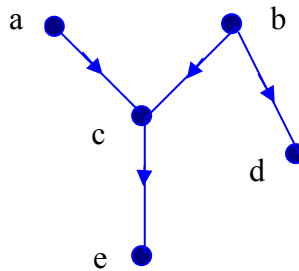
In summary, enhanced coagulation has the primary objective of removing NOM to control DBP formation. In order to achieve the primary objective, the need exists to maximize particle and turbidity removals by downstream solid-liquid separation production; to minimize residual coagulant; to minimize sludge production; and to minimize operating costs. Optimum coagulation is defined as coagulation conditions of dosage, pH that optimize the sub-objectives and the primary objective. Therefore, optimization of the coagulation conditions plays an important role for NOM removal and control of DBP formation but the extensive interaction in system design must be reflected, an ideal application for network theory.

### 10.3 Background to Probabilistic Network Theory

Although probability provides a rigorous formalism to encode uncertain knowledge in a problem domain, the traditional approach to probabilistic inference based on joint probability distributions is intractable in domains with more than a few variables. Probabilistic networks offer a mathematical framework and a feasible computational formalism for reasoning in uncertain problem domains with a large number of variables. A probabilistic network encodes concisely, a probability distribution over a set of domain variables, by using a graphical representation. The causal or dependent relations in a probabilistic network between variables are encoded in the graphical structure and the strength of the uncertain relations is encoded by probability distributions associated with the graph, such as in a Bayesian network (BN). In particular, the network structure of a BN is a directed graph (where a link between two nodes is directed from one to the other). Each directed link represents the causal or asymmetric dependence between the two variables. The directed graphs may include undirected cycles (a sequence of nodes where the last is identical to the first, each is connected to the next in the sequence by a link, and not all links are in the same direction relative to the sequence), but not self-loops (a self-loop is a link leaving and coming into the same node). A graph that contains no directed cycles is called acyclic. A graph that is both directed and acyclic is called a directed acyclic graph (DAG). The network structure of a BN is a DAG.

In a BN, the dependence among the variables is encoded with a DAG as indicated in Figure 10.2. Each node in the graph corresponds uniquely to a variable in the problem domain, and hence a node and a variable are terms which are used interchangeably. Loosely speaking, the parent nodes of a node 'x' are those variables that represent the direct causes of x. The

strength of the causal dependence is encoded by a probability distribution  $P(x|\text{par}(x))$ , where  $\text{par}(x)$  denotes the set of parent variables of 'x'. The DAG embeds the semantics that every variable is independent of any non-descendent variables, given the values of its parent variables. Once the probabilistic knowledge (namely, the probabilistic distribution  $P(x|\text{par}(x))$  for each node x) about the domain is specified as a Bayesian network, the joint probability distribution over all variables of the domain can be obtained as a product of probability distributions associated with nodes in the graph. This joint distribution can then be used to reason about possible values of any variable when values of some variables have been observed.



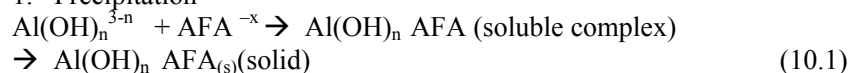
**Figure 10.2.** Schematic of a Bayesian network.

Causal models can serve to predict the effect of interventions. A causal structure of a set of variables  $V$  is a DAG in which each node corresponds to a distinct element of  $V$ , and each arrow represents a direct functional relationship among the corresponding variables. The relationships are characterized by conditional probability distributions that reflect the aggregate response of each variable to changes in its “up-arrow” predecessor, together with the uncertainty in that response. In the domain of enhanced coagulation, the BN method allows an efficient prediction of the consequence of a given coagulation setup as will be shown below.

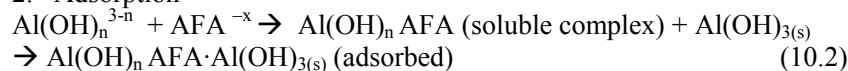
## 10.4 Bayes Network for Enhanced Coagulation

First step: The BN encodes the dependence among key variables as a set of nodes with arrows representing direct functional relationships in the coagulation process. There are two parallel mechanisms of coagulation:

## 1. Precipitation



## 2. Adsorption



Both aquatic humic and fulvic acids (AFA) are the major form of aquatic humic matter. Because of their lower molecular weight and hydrophilic nature, fulvic acids are more difficult to remove by coagulation. AFA has a high total negative charge of about  $15 \mu\text{eq/mg C}$  (Edzwald and Van Benschoten, 1990). Its charge is pH-dependent from carboxyl and phenolic groups with approximate values of  $7.5 \mu\text{eq/mg C}$  at pH 5.5 and  $10 \mu\text{eq/mg C}$  or greater, at pH 7.

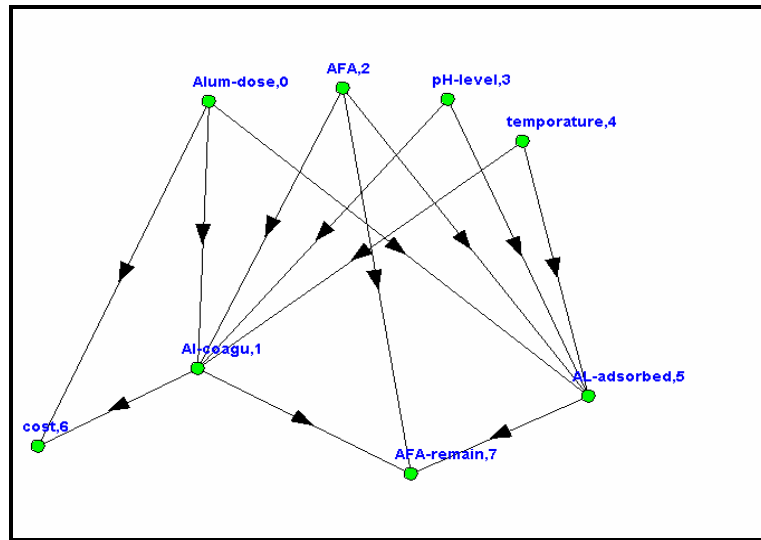
The first mechanism involves reactions between alum hydrolyzed product  $\text{Al(OH)}_n$  and AFA, and hence they are the two causative variables. The reaction produces Al-coagulant,  $\text{Al(OH)}_n\text{AFA}_{(s)}(\text{solid})$ , as Equation 10.1 shows. Al-coagulant (Al-coagu) is causally dependent on  $\text{Al(OH)}_n$  and AFA. Therefore,  $\text{Al(OH)}_n$  and AFA are the parent nodes of Al-coagu in the BN.

The amount of Al-coagu produced through reaction between  $\text{Al(OH)}_n$  and AFA is influenced by the pH level. Hence, the pH level becomes another parent variable of Al-coagu. Temperature also plays a role in Al-coagu produced and hence the fourth parent node of Al-coagu.

The second mechanism involves reactions between  $\text{Al(OH)}_n$  and AFA in which the charge of AFA is neutralized, followed by adsorption of the complex on aluminum hydroxide solids. To represent this output mechanism, the node AL-adsorbed is created in the BN with the same set of parent nodes as Al-coagu.

The overall effect of the coagulation process can be described by the remaining AFA, or AFA-remain, in the water after processing. AFA-remain depends on the original amount of AFA, the amounts of Al-coagu and AL-adsorbed due to reactions. Hence, there are three variables as the parent nodes of AFA-remain. Finally, the cost-effectiveness of the coagulation is directly dependent on the amount of  $\text{Al(OH)}_n$  added. It is also dependent on the Al-coagu produced, as their removal requires filtration. Therefore, both  $\text{Al(OH)}_n$  and Al-coagu are the parent variables of cost. Figure 10.3 shows

the resulting Bayesian network of nodes and interactions for enhanced coagulation.



**Figure 10.3.** A Bayesian network of enhanced coagulation.

Second step: The conditional probability table for each variable in the BN can be acquired from experimental data but expensive experimental resources are needed to allow this. Alternatively, the probability tables can be based on available knowledge from the technical literature and thereby provide more focused experimental resources as an efficiency step.

The key knowledge applied in the estimation is the nature of the causal influence between a child variable and its parent variable(s). From the literature review in Section 10.2, the following causal influences are identified between the parent variables of *Al-coagu*. (i) If the amount of AFA increases, then the amount of *Al-coagu* produced will increase. (ii) If the amount of  $\text{Al}(\text{OH})_n$  increases, then the amount of *Al-coagu* produced will increase. (iii) If the pH level increases, then the amount of *Al-coagu* produced will decrease. Since the relationship between *Al-coagu* and pH is not monotonic, modeling herein is restricted to this response. (iv) If the temperature increases, then the amount of *Al-coagu* produced will decrease. From the knowledge (i) through (iv), the condition that is the most advantageous in producing *Al-coagu* is identified; this occurs when AFA

and  $\text{Al}(\text{OH})_n$  are at the highest levels and the pH and temperature are at their lowest levels. The condition that is the least advantageous in producing *Al-coagu* occurs when AFA and  $\text{Al}(\text{OH})_n$  are at the lowest levels, and when pH level and temperature are at the highest levels. Once these two extreme conditions are identified (the best and the worst), the probabilities of costs given these two conditions can be estimated, as they provide bounding conditions for estimating the remaining probabilities. Subsequently, other conditions that are between the best and worst are identified and rank-ordered from the best to the worst.

The result is a partial order, rather than a total order of the causal conditions of cost. Assume each of the variables is categorized into a sequence of steps or intervals. For example, if the amount of  $\text{Al}(\text{OH})_n$  increases by one interval from a given condition (which is more advantageous in producing *Al-coagu*), but the pH level also increases by an interval (which is less advantageous in producing *Al-coagu*), then the new condition may be comparable with the given condition, when there is no clear reason to believe one is better than the other. The ranks of both conditions are therefore assigned as identical. However, if  $\text{Al}(\text{OH})_n$  increases by two intervals, then the new condition is more advantageous than the given condition and ranked higher than the given condition. Once the number of different ranks has been determined, the conditional probability distribution for each rank is estimated, while ensuring that the probability values are consistent with the relative ranks.

To illustrate this procedure the assessment of conditional probability distribution of cost is used as an example. We assume that variable cost has a binary domain of

$$\{\text{less than } \$6/10^6\text{L, greater than or equal to } \$6/10^6\text{L}\},$$

which will be denoted as

$$\{< \$6/10^6 \text{ L and } \geq \$6/10^6 \text{ L}\}.$$

In the actual model (below) we use later, cost has a domain of

$$\{< \$6/10^6 \text{ L, } \$6\text{-}8/10^6 \text{ L, and } > \$8/10^6 \text{ L}\}.$$

We use the binary domain above to simplify the explanation.

The cost variable has two parents. One of them is  $\text{Al}(\text{OH})_n$ , whose domain is

$$\{<20, 20\text{-}40, \text{ and } >40\}.$$

Note that under this representation,  $\text{Al}(\text{OH})_n$  is a discrete variable of three possible values and should not be confused with the corresponding continuous variable which can take any real value in the range, say, from 10 to 50. Therefore, the following statements are all equivalent and are used interchangeably:

“Al(OH)<sub>n</sub> takes the value ‘<20’.”

“Al(OH)<sub>n</sub> = ‘<20’.”

“Al(OH)<sub>n</sub> <20.”

Note that in the third statement, Al(OH)<sub>n</sub> appears as a continuous variable but the statement express the same meaning as the first two statements.

The other parent of cost is Al-coagu, whose domain is

$$\{<4, 4-8, >8\}.$$

Therefore, the conditional probability distribution contains a total of  $2*3*3 = 18$  parameters. We demonstrate how to assess

$$P(\text{cost} < 6 \mid \text{Al(OH)}_n, \text{Al-coagu}),$$

which amounts to evaluation of 9 parameters. The other 9 parameters are determined from:

$$P(\text{cost} \geq 6 \mid \text{Al(OH)}_n, \text{Al-coagu}) = 1 - P(\text{cost} < 6 \mid \text{Al(OH)}_n, \text{Al-coagu}). \quad [10.3]$$

The causal impacts of both Al(OH)<sub>n</sub> and Al-coagu on cost are positive. That is, as Al(OH)<sub>n</sub> or Al-coagu increases in value, the cost increases as well. When Al(OH)<sub>n</sub> takes the value <20 and Al-coagu takes the value <4, the cost is likely to be at its lowest level. We assess

$$P(\text{cost} \geq 6 \mid \text{Al(OH)}_n < 20, \text{Al-coagu} < 4) = 0.7.$$

When Al(OH)<sub>n</sub> >40 and Al-coagu >8, the cost is likely to be at its highest level. We assess

$$P(\text{cost} \geq 6 \mid \text{Al(OH)}_n > 40 \text{ and } \text{Al-coagu} > 8) = 0.1.$$

Between the two extreme cases, three intermediate levels of causal influence exist and, hence, there are a total of five levels. To facilitate description, the five levels of causal influence, from the lowest to highest, are referred to as levels 1 through 5.

Level 2 influence consists of the following two cases:

$$\text{Al(OH)}_n < 20 \text{ and } \text{Al-coagu} = 4-8$$

and

$$\text{Al(OH)}_n = 20-40 \text{ and } \text{Al-coagu} < 4.$$

Comparing with level 1, namely,

$$\text{Al(OH)}_n < 20 \text{ and } \text{Al-coagu} < 4,$$

It can be seen that in the first case, Al(OH)<sub>n</sub> has the same value as level 1, but Al-coagu has the value 4-8, which has larger cost than the value <4 in level 1. In the second case, Al-coagu has the same value <4 as level 1, but Al(OH)<sub>n</sub> has the value 20-40, which is more costly than the value <20 in level 1.

Level 3 consists of the following three cases:

$$\text{Al(OH)}_n < 20 \text{ and } \text{Al-coagu} > 8,$$

$$\text{Al(OH)}_n = 20-40 \text{ and } \text{Al-coagu} = 4-8,$$

and

$$\text{Al(OH)}_n > 40 \text{ and Al-coagu} < 4.$$

Note that in the first case, Al-coagu is more costly than the lowest level by two levels and Al(OH)<sub>n</sub> remains the same. In the third case, Al(OH)<sub>n</sub> is more costly than the lowest level by two levels and Al-coagu remains the same. In the second case, Al-coagu is more costly than the lowest level by one level and so is Al(OH)<sub>n</sub>.

Level 4 consists of two cases again: Al(OH)<sub>n</sub> = 20-40 and Al-coagu ≥ 8, and Al(OH)<sub>n</sub> ≥ 40 and Al-coagu = 4-8. In the first case, Al(OH)<sub>n</sub> is more costly than the lowest level by one level and Al-coagu by two levels. We then assess the conditional probability parameters for levels 2 through 4 as 0.55, 0.4, and 0.25. The resulting 9 probabilities for cost being less than '6' are as listed in Table 10.4.

**Table 10.4** Probability of costs conditioned on Al(OH)<sub>n</sub> and Al-coagu.

P(cost = <6   Al(OH) <sub>n</sub> , Al-coagu)	Al(OH) <sub>n</sub> (mg/L)	
	Al(OH) <sub>n</sub> (mg/L)	Al-coagu(mg/L)
0.7	<20	<4
0.55	<20	4-8
0.4	<20	>8
0.55	20-40	<4
0.4	20-40	4-8
0.25	20-40	>8
0.4	>40	<4
0.25	>40	4-8
0.1	>40	>8

Using the above technique, we estimate the probability distributions for Al-coagu, AL-adsorbed, AFA-remain, and cost. In this fashion, the probabilistic tables are established.

The computation of unknown probability distributions over some variables from other known distributions is called *probabilistic inference* or *inference*. The probability tables established above represent the general knowledge on coagulation conditions. Through inference computation, such knowledge can be utilized to estimate the outcomes of coagulation under any given conditions. In particular, we can specify the values of some variables and let the inference computations tell us what are the likely values of target variables. For example, we can increase pH or decrease pH to observe how

other variables change their values. By exploring different coagulation conditions, we can derive the optimal values for the pH, Alum dosage, etc.

Each bar in the histograms in Figure 10.4 represents a probability value in the range [0,1]. Each bar is labeled by the possible value of the corresponding variable. Only the first character of the label is shown. For example, possible values of variable cost are <6, 6-8, and >8. Only <, 6, and > are shown for purposes of explanation. The histogram in Figure 10.4 is produced only as a visual aid; the actual numerical probability distribution is stored as a log file by the inference program.

After we specify the pH value, for example, at  $\leq 7$ , the temperature at  $\leq 5$ , AFA value at  $\leq 4$  and  $\text{Al}(\text{OH})_n$  value at  $\leq 20$ , inference computation provides the probability distributions of the dependent variables, Al-coagu, Al-adsorbed or cost, conditioned on these factors as shown in Figure 10.4. We can observe the following: the probability of Al-coagu > 8 decreases from about 0.58 in Figure 10.4 to about 0.4 in Figure 10.5, the probability of Al-adsorbed > 8 is also decreasing from about 0.57 in Figure 10.4 to about 0.4 in Figure 10.5; AFA-remain > 8 decreases from about 0.32 in Figure 10.4 to about 0.26 in Figure 10.5; and cost > 8 decreases from about 0.35 in Figure 10.4 to about 0.21 in Figure 10.5.

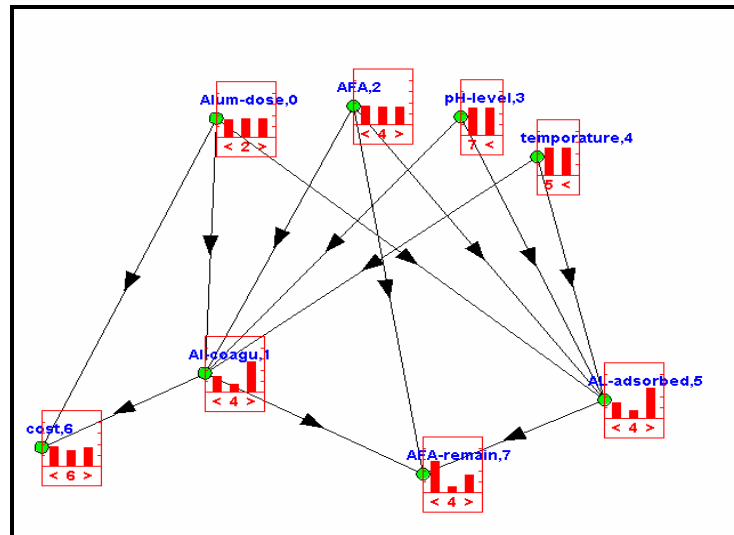
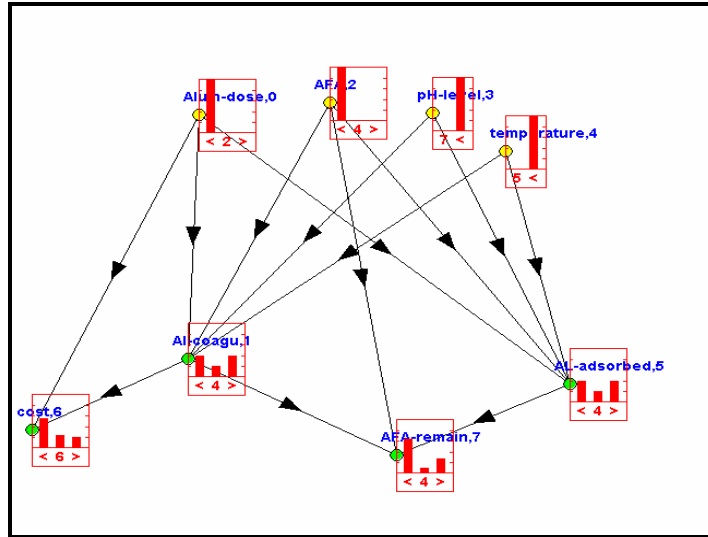


Figure 10.4 The prior probability distribution.



**Figure 10.5** Evidence for  $pH \leq 7$ ,  $temp \leq 5$ ,  $AFA \leq 4$  and  $Al(OH)_n \leq 20$ .

**Table 10.5** Posterior probability computed by BN with the evidence of  $pH \leq 7$ ,  $temp \leq 5$ ,  $AFA \leq 4$  and  $Al(OH)_n \leq 20$

Al-coagu		Al- adsorbed		AFA-remain		Cost	
Interval,	Prob.	Interval,	Prob.	Interval,	Prob.	Interval,	Prob.
<4:	0.4	<4:	0.4	<4:	0.63	<6:	0.55
4-8:	0.2	4-8:	0.2	4-8:	0.099	6-8:	0.24
>8:	0.4	>8:	0.4	>8:	0.26	>8:	0.21
ExpV:	3.6	ExpV:	3.6	ExpV:	2.89	ExpV:	4.17

The expected value (ExpV) in Table 10.5 was derived according to Equation 10.4:

$$Expvalue = \sum_{i=1}^n P(x_i) * x'_i \tag{10.4}$$

where:

- $x'_i$  = the middle of the interval  $x_i$ , and
- $i$  = the index of the individual intervals.

To analyze another condition, we specify pH value at  $>7$ , temp value at  $>5$ , AFA value at 4-8 and  $\text{Al}(\text{OH})_n$  value at 20-40. The inference computation provides different probability distributions on *Al-coagu*, *Al-adsorbed* and cost.

**Table 10.6.** Posterior probability computed by BN with the evidence of  $\text{pH}>7$ , temp  $>5$ , AFA= 4-8 and  $\text{Al}(\text{OH})_n=20-40$ .

Al-coagu		Al- adsorbed		AFA-remain		Cost	
Interval,	Prob.	Interval,	Prob.	Interval,	Prob.	Interval,	Prob.
<4:	0.4	<4:	0.4	<4:	0.5	<6:	0.4
4-8:	0.2	4-8:	0.2	4-8:	0.12	6-8:	0.28
>8:	0.4	>8:	0.4	>8:	0.38	>8:	0.32
ExpV:	3.6	ExpV:	3.6	ExpV:	3.24	ExpV:	4.44

When we compare the two policies (Tables 10.5 and 10.6), the first one is optimal. There are lower AFAs remaining and lower costs. In reality, we can have different combinations according to different coagulation conditions and the optimal policy can be derived using Bayesian Networks.

## 10.5 Conclusions

To predict optimal enhanced coagulation conditions, a computational model for enhanced coagulation as a Bayesian Network was developed. The causal dependence relations among the variables were encoded by the BN network structure. To specify the probability parameters for the distribution of a consequence variable  $x$  given its cause variables  $\pi(x)$ , the configurations of  $\pi(x)$  that are the most advantageous in producing the effect represented by  $x$  were identified. We also identified the configurations of  $\pi(x)$  that are the least advantageous in producing the effect and each group of intermediate configurations of  $\pi(x)$  that has approximately the same strength in producing the effect. Using this partition, the probability distribution  $P(x|\pi(x))$  can be specified using a linear number of parameters in the number of groups instead of requiring an exponential number of parameters. Using this model, the consequence of alternative coagulation conditions can be evaluated very efficiently. It allows the optimization of the coagulation conditions and improves the effectiveness on the enhanced coagulation processes.

## References

- Chadik, P.A., and G.L. Amy, "Removing Trihalomethane Precursors from Various Natural Waters by Metal Coagulants", *J. AWWA*, 75:10:532, 1983.
- Chowdhury, Z.K., Roberson, J.A. and Owen, D.M., "A National Evaluation of Enhanced Coagulation and Enhanced Softening", *Proc. AWWA Annual Conference*, Atlanta, GA, June 1997.
- Duan, Jinming, Gregory, John, "Influence of soluble silica on coagulation by aluminium sulphate, *Colloids and Surfaces*" Elsevier A: Physicochemical and Engineering Aspects 107, 309-319, 1996.
- Edwards, M., "Predicting DOC Removal during Enhanced Coagulation", *J. AWWA*, 89:5:78, 1997
- Edward, J. K. and Van Benschoten, J. B., "Aluminum Coagulation of Natural Organic Matter", In: *Chemical Water and Wastewater Treatment*, H. Hahn and R. Klute (eds), Springer Verlag, NY, pp. 341-359, 1990.
- Edward, J. K. and Tobiason, J. E., "Enhanced coagulation: US Requirements and a Broader View", *Wat. Sci. Tech.* Vol. 40, No. 9, pp 63-70, 1999.
- Hubel, R.E., and Edzwald, J.K., "Removing Trihalomethane Precursors by Coagulation", *J. AWWA*, 79:7:98, 1987
- James Montgomery Consulting Engineers, Inc., "Effect of Coagulation and Ozonation on the Formation of Disinfection By-products", Prepared for the AWWA, 1992
- James K. Edward and John E. Tobiason, *Enhanced coagulation: US Requirements and a Broader View*, *Wat. Sci. Tech.* Vol. 40, No. 9, pp 63-70, 1999.
- Kavanaugh, M.C., "Modified Coagulation for Improved Removal of Trihalomethane Precursors", *J. AWWA*, 70:11:613, 1978.
- Knocke, W.R., S. West, and R.C. Hoehn, "Effects of Low Temperature on the Removal of Trihalomethane Precursors by Coagulation", *J. AWWA*, 78:4:189, 1986.
- Krasner S W, Amy G L (1995). Jar-test evaluations of enhanced coagulation. *JAWWA*, 87(10) 93-107.
- Lee, Jeong-Dae, Lee, Sang-Ho, Jo, Min-Ho Park, Pyung-kyu, Lee, Khung-Hak and Kwak, Jong-Wood, "Effect of Coagulation Conditions on Membrane Filtration Characteristics in Coagulation-Microfiltration Process for Water Treatment", *Environ. Sci. Technol.* 34, 3780-3788, 2000.
- Ma, Jun, Liu, Wei, "Effectiveness and mechanism of potassium ferrate(VI) preoxidation for algae removal by coagulation", *Water Research* 36, 871-878, 2002.
- Owen, D.M., R.S. Summers, Z.K. Chowdhury, "Removal of DBP Precursors by Granular Activated Carbon Adsorption", *AWWARF Report*, 1996.
- Owen, D.M., G.L. Amy, and Z.K. Chowdhury, "Characterization of Natural Organic Matter and its Relationship to Treatability", *AWWARF Report*, 1993.
- Semmens, M.J., and T.K. Field, 1980. "Coagulation: Experiences in Organics Removal." *J. AWWA*, 72:8:476.
- Singer, Philip C. Bilyk, Katya, "Enhanced coagulation using a magnetic ion exchange resin", *Water Research* 36, 4009-4022, 2002.
- USEPA, (1998a). *National Primary Drinking Regulations: Disinfectants and Disinfection Byproducts*; Final Rule. 40 CFR parts 9, 141, and 142, Federal Register, 63(241), 69390-69476.

US EPA, Enhanced Coagulation and Enhanced Precipitative Softening Guidance Manual, Office of Water, EPA 815-R-99-012, 1999.

Vickers, J. C. Thompson, M. A. and Kelkar, U. G., "The use of membrane filtration in conjunction with coagulation processes for improved NOM removal", *Desalination* 102, 57-61, 1995.

Young, J.S., and P.C. Singer, "Chloroform Formation in Public Water Supplies: A Case Study", *J. AWWA*, 71:12:87, 1979.

Zhu. Z. J. Y. , and McBean, E., "Estimation of Censored Data Water Quality Values Using Decomposable Markov Networks", *International Conference on International Society for Environmental Information Sciences*, 2004.