International Congress on Environmental Modelling and Software

6th International Congress on Environmental Modelling and Software - Leipzig, Germany - July 2012

Jul 1st, 12:00 AM

Assessing variation in biofilms development in a drinking water distribution system by an object oriented Bayesian network approach

E. Ramos-Martínez

M. Herrera

Joaquín Izquierdo

Rafael Pérez-García

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Assessing variation in biofilms development in a drinking water distribution system by an object oriented Bayesian network approach

E. Ramos-Martínez[a], M. Herrera[b], J. Izquierdo[c], R. Pérez-García[d]
Fluing–IMM,Universitat Politècnica de València. C. de Vera s/n. Edificio 5C, bajo. Valencia. Spain[abcd].
evarama@upv.es[a], mahefe@upv.es[b], jizquier@upv.es[c], rperez@upv.es[d]

Abstract: Biofilms develop in drinking water distribution systems (DWDSs) as layers of microorganisms bound by a matrix of organic polymers and attached to pipe walls. Biofilms are ubiquitous in DWDSs, regardless of the type of treatment or disinfection employed. Many problems in these systems are microbial in nature. Biofilm growth within DWDSs could lead to operational problems such as deterioration of bacterial water quality, generation of bad tastes and odors, and proliferation of macroinvertebrates and other undesirable impacts in DWDSs. The presence of substantial and active attached biomass can protect pathogenic microorganism, create anaerobic zones, lead to the formation of high biocorrosion zones, and consume residual disinfectant. Numerous studies have been carried out on the reasons and effects of the biofilms in DWDSs, both in the microbiological and in the engineering fields. Several factors have been found related to biofilm development in DWDSs, but the complexity of the disinfectant microenvironment under study, and the use of different methodologies and biofilm growth systems lead to ambiguous or not easily comparable results. Our aim is to compile the information available nowadays about biofilm ecology since it has been simplified in practical approaches, softening the biofilms' role. To this purpose, an object oriented Bayesian network (OOBN) is proposed. This framework for knowledge representation uses a Bayesian network to describe the probabilistic relations between the attributes of an object. These attributes can themselves be objects, providing a natural way for managing a pipe, a district area, and a whole DWDS in the same encoding features. Thus, we propose a tool that allows us to identify network areas whose characteristics tend to develop more biofilm.

Keywords*: biofilm; water distribution systems; object oriented Bayesian networks*

1. INTRODUCTION

Biofilms develop in drinking water distribution systems (DWDSs) as complex communities of microorganisms bound by an extracellular polysaccharide polymer, the glycocalyx, which provides them with structure, protection and helps retain food. These communities of organisms form spontaneously in DWDSs due to the presence of moisture, bind strongly against the initial repulsion at the inner pipe wall, and modify the pipe as they capture more nutrients and new bacteria.

A developed biofilm is very resistant and may pose a significant problem when a clean and disinfected environment is needed. In addition to the health risk that biofilms create due to their role as microbial pathogens reservoir [Momba et al., 2000], biofilms are also responsible for many other DWDSs’ problems. For example: aesthetic deterioration of the water, proliferation of higher organisms [Gelves, 2005], operational problems, biocorrosion [Videla, 2005], consumption of disinfectant [de Beer, 1994], among others. Currently, biofilms represent a challenging paradigm in the management of water quality in all DWDSs.

Different systems have been devised to study biofilms and the influence of the large spectrum of conditions that are known to alter their development potential. However, most studies assess only one variable at a time [Simoes et al., 2007]. The aim of this research is to study the effect of the interaction of the physical and hydraulic conditions of DWDSs on biofilm development, and, thus, to identify the pipes that are prone to biofilm development according to the characteristics considered above. To achieve this goal the currently available information of the DWDSs' physical and hydraulic conditions that affect biofilm development has been compiled. The results obtained in different studies are abundant. Nevertheless, in many cases they are difficult to compare [Simoes et al., 2007] due to the complexity of the environment studied and the use of different methodologies and biofilm growth reactors. In this work we propose Bayesian networks (BNs) as a statistical tool to help achieve our purpose. Then, object oriented Bayesian networks (OOBNs) are used to study the whole supply network and its classification in areas of various risk levels of biofilm development.

The present study follows the structure described below. In Section 2, materials and methods, we consider Bayesian networks and object oriented Bayesian networks, and their involvement in the study of biofilm development in DWDSs. In Section 3, results and discussions, we focus on the probabilities obtained by the network implementation and analysis, both, at pipe and network levels. Finally, in Section 4, conclusions, we review the implications of BNs and OOBNs as new tools in the study of biofilms development in DWDSs and possible future applications in this area.

2. MATERIALS AND METHODS

2.1 Brief introduction to Bayesian networks

Bayesian networks [Pearl, 1988, Heckerman, 1995] are a type of probabilistic graphical models characterized by modeling causal relationships. These models arise as a result of the blending between Graph Theory and the Theory of Probability. When a probabilistic mathematical model is built it is essential to take into account two important components on the information provided: qualitative and quantitative information of the problem [Heckerman, 1995]. Qualitative information introduces information associated with the dependency relationships between the variables of the model. Based on Graph Theory, this information can be summarized by a graph in which nodes represent variables of the problem and the arcs represent the dependency and causality between them. Accordingly, the lack of arcs means independence relationships. To build a mathematical model, probabilistic information is available concerning the probability distribution of the variables of the problem, also called quantitative information of the problem. When modeling a problem using a Bayesian network (BN), experts in the field of the network application should determine the dependency relationships between variables, which will be later reflected in a directed acyclic graph (DAG) representing the qualitative part of the network [Jensen, 2001].

2.2 Bayesian networks; application to our experimental study

By using BNs, we aim at making efficient use of the information and at reducing the uncertainty associated with biofilm development processes. To configure the Bayesian network structure (Figure 1) information has been gathered from experts. The result of the aggregation of this information, despite being rich in knowledge, presents certain ambiguities due to the fact that they come from different sources, further justifying the use of the BNs as a suitable method for analysis.

The Bayesian network proposed contains the following nodes (Figure 1):

- Water age (WAGE).This node represents the residence time of the water in the DWDS. The older the water, the greater decay of disinfectant residual, sediment deposition and temperature increase [EPA, 2002]. All of them are aspects that favor the biofilm development. In our case, to increase the number of data available for the analysis, we decided to create a synthetic index, which we call water age. The information used to this purpose was the Hydraulic Retention Time (HRT) and the distance to the point of disinfection (Km). In each case, the minimum value was subtracted to the current value and divided by the difference between the maximum and the minimum. HRT values thus obtained were weighted by a factor of 0.7 and the distance to the disinfection point by 0.3, before aggregating them. Values closer to one will be those who have older water. Less weight has been given to the distance from the point of disinfection, multiplying it by a lower factor, because the water may recirculate. This index was discretized in three categories by dividing evenly the range (Table 1).

- Flow velocity (VELOCITY). Working on a pilot distribution system, Lehtola found [Lehtola, 2006] that the formation of biofilms increased with the flow velocity of water. It may be because there is an increase of mass transfer nutrients. Although specific velocities between 3-4 m/s may favor its release [Cloete et al., 2003]. The flow velocity discretization was based on experts' criteria. From 0 to 0.7 m/s was considered low, medium from 0.8 m/s to 1.7 m/s and high above 1.7 m/s (Table 1).

- Hydraulic regime (FLOW). It can be turbulent or laminar (Table 1). It has been demonstrated that some biofilms in turbulent flow tend to be more active, having more mass per cm^2, increased cell density and distinct morphology than biofilms in laminar flow [Simoes et al., 2007].

- Roughness / Deposits (roughness).The DWDS pipes with a rough surface have greater potential for biofilms growth [Chowdhury, 2011]. The deposits can increase the roughness by providing more surface to biofilm growth and reducing water drag force on the biofilm. Due to the limited quantitative information available regarding these factors and the biofilm development in DWDSs, this node will be treated as a hidden or latent node. A node is hidden if no case in the dataset contains any observation on the node. This node will depend on pipe material and age. In Figure 1 it is represented as a circle, instead of a rectangle like the other nodes.

- Pipe material (MATERIAL). The pipe materials can be classified into metallic, plastic, or cement (Table 1). In general, we can say that metallic pipes produce more biofilm than cement pipes and these more than plastic pipes [Niquette et al., 2000].

- Pipe age (PAGE).The accumulation of corrosion products and dissolved substances in the older pipes can increase the roughness of the pipe [Christensen, 2009], thus favoring the development of biofilm. In addition, older deposits may have greater biomass and contain more bacteria [Chowdhury, 2011]. The pipes age was discretized in young, medium and old based on experts criteria as shown in Table 1.

- Biofilm (BIOFILM). Heterotrophic plate count (HPC) has been chosen as a method of biofilm quantification. It was observed that biofilm data obtained followed a normal distribution, so it was decided to take the central values as average values and the tails as extremes. Thus, from 0 to 10^3 biofilm development was considered low, from 10^4 to 10^6 medium, and from 10^7 onwards high.

The resulting Bayesian network structure is shown in Figure 1.

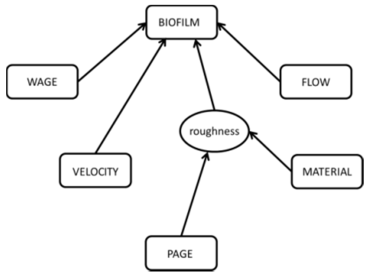

Figure 1.- Bayesian network structure

The study has been performed by with Hugin Lite 7.5 program [Andersen et al., 1989] using discrete data with no missing values. The different categories of each variable are shown in Table 1.

Table 1.- Variables and categories

P. MATERIAL	P. AGE (years)	FLOW VELOCITY (m/s)	BIOFILM (HPC/cm^2)
metalic	high [≥ 31)	high [1.8-3.5]	high [≥ 10^7]
cement	medium [11-30]	medium [0.8-1.7]	medium [10^4-10^6]
plastic	low [0-10]	low [0-0.7]	low [0-10^3]
HYDRAULIC REGIME		WATER AGE	
laminar		medium [0.4-0.6]	
turbulent		low [0-0.3]	

2.3 Object oriented Bayesian networks application.

Applying BNs to real-world domains often requires modeling large and complex systems. The object oriented (OO) software engineering paradigm provides a framework for the large-scale construction of robust, flexible, and efficient software [Korb & Nicholson, 2011]. From this point of view, object oriented Bayesian networks (OOBNs) are presented as a powerful and general approach for large-scale knowledge representation using BNs [Koller & Pfeffer, 1997].

While an ordinary BN is made up of ordinary nodes, representing random variables; an OOBN class is made up of both ordinary nodes (simple objects: random variables) and objects (complex objects, instances of other classes or nodes belonging to other level of abstraction, whose attributes can enclose another objects – a class can be thought of as a self-contained instance or template for an OOBN object). Thus, an object can encapsulate multiple sub-networks, and OOBNs can model domains with hierarchical structure and redundancies. In our application case, we understand a pipe as a simple object. Thus, a structured set of pipes would be defined as the class of hydraulic sectors, where properties of pipes are inherited. At this abstraction level we can say that hydraulic sectors can be defined by the pipes they are made of. Similarly, the whole DWDS is composed by a structured set of hydraulic sectors.

OOBNs can be very close representations of a real-world problem, achieving interesting approaches of a huge variety of problems. Nevertheless, OOBNs have drawbacks such as: work with large and messy networks, obtain outputs composed of complex modeling processes, all evidence is treated at the same level, and their implementation and interpretation may be hard.

Our proposal to implement the OOBN is based on a junction tree of Bayesian networks [Flores et al., 2003] (Figure 2). As a result, we will have instances of classes inside a unique encapsulating class. This structure can be nested once again, finding an easy way to propagate changes in one class to all objects belonging to it. This leads to more efficient inference and a tree-modular structure that increases the robustness and the flexibility of the algorithms. The interface structure of junction trees of Bayesian networks is also well-suited to implement dynamic domains [Bangso et al., 2004].

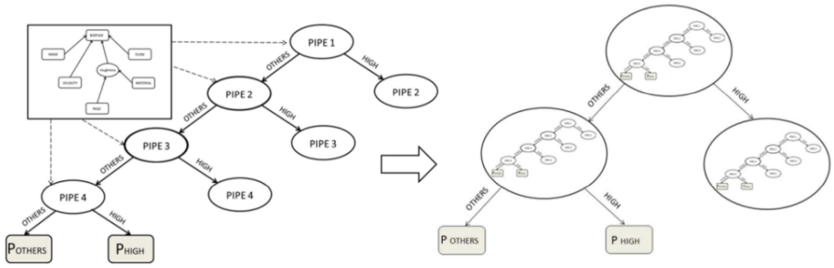

Figure 2.- Junction tree

This analysis has been performed in a small section of the Celaya center DWDS, Mexico (Figure 3).

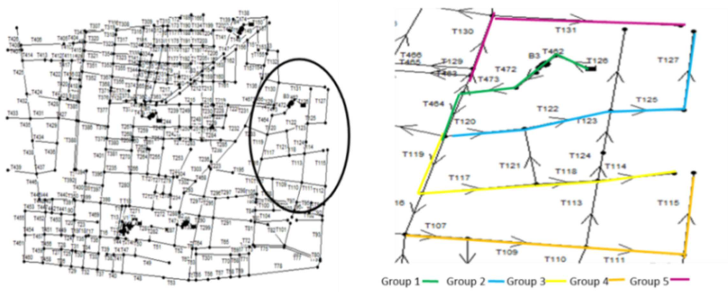

Figure 3.- Celaya center network and pipes used for OOBN analysis

The number of studied pipes was 20 and all of them were located in the same area. The pipes were divided in groups of four. The pipes belonging to the same group were consecutive pipes, joined downstream (Figure 3).

First, the previously trained BN was run in each pipe of the chosen section to obtain the expected biofilm development probabilities. After that, the groups of four pipes were studied. The condition for a group to have high biofilm development was that the probability of having at least one pipe with high biofilm development must be higher than 0.6. Next, we studied the biofilm development in the whole sector using the results obtained in the analysis of the pipe groups. In this case, a sector was considered to have a high biofilm development if the probability of having at least three groups of pipes with at least one pipe with high biofilm development was higher than 0.6. It is worth noting that the criteria for high biofilm development used in these OOBN analyses were based on the opinion of the experts of our group. Thus, they can change, depending on the interests and water quality objectives of the water utilities' stakeholders.

3. RESULTS AND DISCUSSION

3.1 Bayesian network results

The subsequent distribution of our Bayesian network that represents the biofilm development in a pipe differs significantly from the null hypothesis of independent nodes (if we reject H_o the probability to be wrong is just around 0.07, Table 2).

Table 2.- p-value of the test for marginal independence in relation to biofilm

Node	roughness	WAGE	VELOCITY	PAGE	MATERIAL	FLOW
p-Value	-	0.06	1.26E-12	3.53E-05	1.12E-07	0.07

It is worth noting that according to the theory the highest probability of high biofilm development would be expected in the case of turbulent flow, high velocity and medium water age. Instead, that highest probability is found when the velocity is high, the water age medium, and the flow laminar (Table 3). This may be explained by the fact that when the flow is turbulent the shear force is higher than when the flow is laminar. As a result, the biofilm detachment is favored in the case of turbulent flow (but not when the flow is laminar). This result enhances the idea that studying the combined effect of the different variables provides better understanding of the processes that really happen in the pipes of a DWDS.

Table 3.- Remarkable results of the Bayesian network

FLOW	VELOCITY	WAGE	roughness	H	M	L
T	L	L	STATE 1	**0.0889**	0.8005	0.1106
T	L	M	STATE 0	**0.0383**	0.7360	0.2257
T	H	L	STATE 1	**0.0714**	0.1780	0.7506
L	L	L	STATE 1	**0.0989**	0.7557	0.1454
L	M	M	STATE 1	**0.5672**	0.3842	0.0486
L	H	M	STATE 0	**0.7917**	0.1232	0.0852

Similarly, the lowest probability of high biofilm development was expected to be in the case of laminar flow, low velocity, and low water age. In this case, as expected, very low probability of high biofilm development is found, particularly in the case of roughness state 1 (Table 3). However, the case which has fewer probability of high biofilm development corresponds to turbulent flow, low velocity, medium water age and state 1 for roughness (Table 3). Finally, it is worth pointing out that except for this case, for all the others cases with very low probability of having high biofilm development (Table 3), the roughness state is 1 and the water age is low. Moreover, in the cases where the probability of high biofilm development is large (Table 3), the water age is medium. This fact coincides with theoretical results that state that if the water age is low less biofilm develops. This may mean that this variable is especially influential in biofilm development.

In relation to the role of the pipe material and pipe age it is worth noting that it is hidden by the latent node, roughness, and it is not easy to see a clear relationship between the states of roughness and the probability of biofilm development.

The appearance of groups of pipe characteristics with a tendency to have high or low biofilm development reinforces the idea that studies of biofilms in a DWDS must be considered as composed of several heterogeneous subsystems that influence biofilm development in different ways. This is the reason why an OOBN is a very useful tool to study biofilm development in the whole DWDS because it takes into account this fact and the interactions between the different subsystems.

3.2 Object oriented Bayesian network results

From the obtained results it can be concluded that, according with the criteria explained above, 4 out of the 5 pipe groups studied have high biofilm development and the whole sector also can be classified as having high biofilm development ($p=0.791$). The OOBN analysis besides allowing us to study the DWDS as a whole, gives us the opportunity to see the pipe material and age effect that we could not appreciate well in the BN because the latent variable. In relation to the obtained pipe groups probabilities, it is worth enhancing the fact that the group that, according with our criteria, does not have high biofilm development ($p=0.51$) differs from the others mainly in having one pipe with medium water age, while the others, which have similar characteristics, do not. It is remarkable that, once again, water age seems to have special effect in biofilm development.

Moreover, the group with the highest probability of having high biofilm development (p=0.669) is the only group that consists mostly of metal old pipes. Both are characteristics that the experts stand out as especially driving forces of the biofilm development. The probabilities obtained are quite similar (p=0.603; p=0.613; p=0.644), among the other groups. It is probably explained by the similar characteristics of the pipe groups studied, which are mainly formed by cement and medium age pipes. This type of pipe is expected to have quite high biofilm development. It may explain the relative high probabilities found. These results suggest that pipes, areas, or/and networks with high biofilm development appear more frequently than expected.

4. CONCLUSIONS

The purpose of this paper is to attain a deeper understanding of the real consequences of a number of factors interacting in biofilm development. So far, most of the approaches in this area were focused in the effect of just one or two factors on biofilm development. To our knowledge, ours is the first one which tries to study the combined effect of a number of physical and hydraulic characteristics of the DWDSs on biofilm development.

In addition, this study provides an overview of an innovative work with the introduction of BNs and OOBNs as new tools that allow using the level of knowledge gained on the development of biofilms in DWDS in a practical and efficient way. It is worth pointing out that the implementation of OOBNs, always difficult, through the Trees Junction offers the possibility to make an approximation of the available information at various levels of aggregation. This fact allows a better implementation and interpretation, making easier its application by non-highly qualified people in this methodology, and allowing better understanding of the processes and interactions that occur in the DWDSs. Through this knowledge the bases for future improvement of drinking water quality are developed. Acquiring a tool which is able to identify and to predict the DWDSs' conditions which favor high biofilm development, and thus, the areas of DWDSs which are prone to it, could increase the effectiveness of the drinking water utilities management. There are a lot of DWDSs' management issues that could be improved thanks to this. It may be hygienically relevant since the hydraulic flushing of pipes could be directed to those areas that are known to develop high biofilm, saving money and time, and increasing its efficiency. In the same way, since biofilms are known to consume disinfectant residual, knowing which areas of the DWDSs develop more biofilm may be useful to decide with more accuracy the location of the chlorination points. The same may happen with the maintenance operation since biofilms are related to the increase of corrosion rates by biocorrosion processes. A good deal of improvements could be carried out through such a tool in drinking water management. Although more work is needed, the results obtained are promising and it seems that we are a little bit nearer to the final aim.

The OOBNs can extend the study of a pipeline to the study of a sector (understood as a set of pipes distributed in space, which inherits the properties of the pipe as a unit). Similarly, from the study of sectors the whole network (the entire network will be an ordered set of sectors in space) may be considered. We claim that in the short term, detection and location of areas where the mitigation of the effects associated with high biofilm development will be intensified. In following stages of this work, it is also planned to include the physico-chemical characteristics of water in order to have a better quantification of biofilm in different areas of DWDSs.

ACKNOWLEDGEMENTS

This work has been performed under the support of the project IDAWAS, DPI2009-11591 of the Dirección General de Investigación del Ministerio de Ciencia e Innovación (Spain) and ACOMP/2011/188 of the Conselleria de Educació of the Generalitat Valenciana. We want to express our gratitude to the research grant (FPI), Ministerio de Ciencia e Innovación (ref.: BES-2010-039145).

REFERENCES

Andersen, S. K., Olesen, K. G., Jensen, F. V., & Jensen, F. "HUGIN—a shell for building Bayesian belief universes for expert systems. *In Proceedings of the eleventh international joint conference on artificial intelligence*, Detroit, Michigan, August 20–25, 1080–1085, 1989.

Bangso, O., J. Flores, and F.V. Jensen. "Plug & Play object oriented Bayesian networks", *L.N. in Artificial Intelligence*, 3040, 457-467, Springer-Verlag, 2004.

de Beer D., Srinivansa R. and Stewart P. S., "Direct Measurement of chlorine penetration into biofilms during disinfection". *Appl. Environ. Microbiol*. 60(3), 4339, 1994.

Chowdhury, S. "Heterotrophic bacteria in drinking water distribution system: a review", *Environ Monit. Assess*, DOI 10.1007/s10661-011-2407-x, 2011.

Christensen Ryan T.: "Age Effects on Iron-Based Pipes in Water Distribution Systems". Paper 505. Utah State University, 2009.

Cloete, T.E, Westard D. and van Vuuren S.J.: "Dynamic response of biofilm to pipe surface and fluid velocity". *Water Science and Tech.*, 47(5), 57-59, 2003.

Flores, M., J. Gamez, and K. Olesen. Incremental compilation of a Bayesian Network. In Proc. of the 19th Conference on Uncertainty in Artificial Intelligence, 233-240. Morgan Kaufmann Publishers, S. Francisco,2003.

Gelves, M.F. "Deterioro de la calidad del agua por el posible desprendimiento de las biopelículas en las redes de distribución de agua potable". *Universidad de los Andes*, Bogota, Colombia, 2005.

Heckerman, D. "A Tutorial on Learning With Bayesian Networks". Technical Report, Msr TR-95-06, Microsoft Research, Redmond, WA, 1995

Jensen, F.V. "Bayesian Networks and Decision Graphs". *Springer*, 2001.

Koller, D. and Pfeffer, A. "Object-oriented Bayesian networks", in Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence, UAI-97, pp. 302-313, Providence - Rhode Island, USA. 1997.

Korb, K. and Nicholson, A. Bayesian artificial intelligence, Chapman & Hall/CRC Press, London, UK. 2011.

Lehtola M. J, Laxandera M., Miettinena I. T., Hirvonenc A., Vartiainenb T., Martikainenc P. J.: "The effects of changing water flow velocity on the formation of biofilms and water quality in pilot distribution system consisting of copper or polyethylene pipes". *Water Research*, 40, 2151-2160, 2006.

Momba M.N.B., Kfir R.,Venter S.N., Cloete T.E., "An overview of biofilm formation in distribution systems and its impact on the deterioration of water quality" *Water Research Commission*, 2000.

Niquette P. M., Servais P., Savoir R.: "Impacts of pipe materials on densities of fixed bacterial biomass in a drinking water distribution system". *Water Resources*, 34(6), 1952-1956, 2000.

Pearl, J. "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Representation and Reasoning Series" (2nd printing ed.). San Francisco, California: Morgan Kaufmann, 1988.

Simoes M, Pereira M.O, Vieira M.J.: "The role of hydrodynamic stress on the phenotypic characteristics of single and binary biofilms of Pseudomonas fluorescens". *Water Science & Technology*, Vol. 55(8-9), 437-445, 2007.

United States Environmental Protection Agency. Effects of water age on distribution system water quality. Paper Issue, 2002

Videla Héctor A, Herrera Liz K., "Microbiologically influenced corrosion: looking to the future" *International Microbiology*, 8, 169-180, 2005.