Investigating the Role of Statistical Models in Water Distribution Asset Management: A Semi-structured Interview Approach

Vikram M. Rao*, **and Royce A. Francis**
The George Washington University, Washington DC, USA

Abstract: A robust asset management plan needs to be in place for water utilities to effectively manage their distribution systems. Of concern to utilities are broken pipes, which can lead to bacteria entering the water system and causing illness to consumers. Typically, water utilities allocate a portion of funds every year for renewal of pipes and valves. However, pipe renewal is largely based on replacing current broken pipes, and long- term asset management planning to replace pipes is not a priority for water utilities. Water utilities are beginning to use probabilistic break models and other statistical tools to predict pipe failures. These models incorporate variables such as pipe length, diameter, age, and material. These models are emerging in the water industry; however, their direct impact on long term asset planning remains to be seen. In addition, the effectiveness of these models is questionable, as there is currently little research done to evaluate the ability of these models to assist in asset management planning. This paper discusses the role of probabilistic pipe break models in structuring long-term asset management decisions. We determine that there are many factors that are needed to contribute to the feasibility of statistical models in a water asset management program, including data availability, funds, and shared information.

Keywords: Water Distribution, Asset Management, Pipe Renewal and Replacement, Mental Models, Structured Interviews.

1. INTRODUCTION

Water utilities are facing challenges regarding their water distribution systems and maintenance of their assets such as pipes, valves, and other elements of their water network. One of the major issues is repairing and replacing aging assets that have been placed underground for fifty or more years. All utilities are facing the need to replace assets in order to avoid breaks and disruptions in water flow, and provide consistently high quality water to their customers. As pipes age, it becomes more imperative to find effective solutions to plan repair and rehabilitation schedules, and to make effective use of limited funds and resources.

One of the ways in which utilities can approach the issue of managing assets is to develop a comprehensive asset management plan. EPA and AWWA (American Water Works Association) provide guidelines for developing an asset management plan, including providing replacement and rehabilitation guidelines, providing decision making strategies for capital improvements, and outlining risk assessment and characterization methods[1],[2],[3]. However many utilities, constrained by limited funds derived from revenues or allocated from city or local governance, are not able to create effective management plans. The strategies they employ consist of replacing current broken pipes, with little attention to long term asset management planning. To support such planning, a framework which promotes effective decision making is needed.

A practice that has been introduced recently to support asset management decision making are statistical models that serve to predict pipe break rates[4]. These models are designed to take various inputs including pipe age, diameter, material, as well as covariates such as soil conditions, weather, and geographical location. The output of the model provides information on which pipes are most likely to break, through probability distributions or discrete values, which allows the utility to create effective prioritization strategies. Many utilities are already adopting models and are beginning to

* To whom correspondence should be addressed: vrao81@gwu.edu

implement them as part of their asset management programs. There are many statistical models available, and a review of recent literature shows many models that have been developed and compared with historical break data obtained from a local utility[5],[6],[7],[8],[9]. The accuracy of the models is good, and continuously improving. Table 1 below shows the accuracy of several models developed in the literature.

Table 1: Accuracy of Current Statistical Models in Literature

Type of Model	Accuracy Measure	Result	Conclusion
Time Linear Model[8]	Akaike Information Criterion (AIC)	14,267	Time Exponential Model is the best predictor of non-zero pipe breaks [8]
Time Exponential Model[8]	AIC	1066	
Poisson GLM[8]	AIC	14932	
I-WARP (Individual Water mAin Renewal Planner)[10]	Pipe-dimension coefficient of determination (pR^2), and time-dimension coefficient of determination (tR^2)	$pR^2 = 0.43$ $tR^2 = 0.61$	Rather successful at estimating the total number of breaks/year, not as successful at estimating breaks/pipe[10]
Weibull/Exponential Model[5]	R^2	R^2=0.39	Explained by simplicity of the model and random processes. Future research will consider additional risk factors[5]

However, there is a need to understand the role of models and the value they bring to an organization. This is an area of interest because utilities are constrained by lack of funds and limited capital improvement budgets. Considering that implementing models can take a lot of time and energy, it is worthwhile to explore the role of models and their value towards long term asset management planning. There is also a need to understand the perceived accuracy of the models and whether they are worth the time and monetary cost of implementing them. This information can prove valuable to utilities and decision makers who seek to adapt models in the hopes of mitigating risks associated with pipe breaks such as adverse health effects and diminished water quality. The tradeoffs associated with adopting models and using them need to be understood in order to provide decision makers with the best information on how to direct their asset management plans.

The objective of this article is to investigate the role of statistical methods in drinking water distribution asset management. We use semi-structured interviews as part of a mental models approach to understand the issues facing water asset managers. These include the role of statistical models in asset management, their value to an organization, and their perceived accuracy. We highlight the issues that are obtained from our interviews to show what can be improved with models, and how utilities can best use them as part of their asset management plan. The value of our research lies in the need to evaluate pipe break models as to their effectiveness and usefulness in long term asset management planning. We believe our results will help models become more effective tools for utilities to employ, leading to more prevalent use among water utility firms.

2. BACKGROUND

Pipe break models have been developed and commented on in recent literature. Clark[6] utilized a condition assessment model to analyze risk of pipe breaks. Their model was a Cox Proportional

Hazards model which was incorporated into a survival model. The authors used a frailty term, which refers to random, unidentified factors which may cause failure (such as soil) as part of the model. Data came from a utility in Laramie, WY.

Debon[7] used several models including the Cox proportional hazard model, the accelerated lifetime model, and generalized linear models to compare and contrast risk factors such as age, size, and pipe diameter. The goal was to examine which variables contributed most to the incidence of pipe breakage. The authors used receiver operating characteristics curves to study the accuracy of each of the models. Their findings showed that the characteristics that contribute to pipe breaks are short pipe length, large pipe diameters, and low pressure.

Yamijala[8] carried out an assessment of several statistical models of pipe break failure. The goal was to compare how accurate various models were at predicting pipe breaks. The models included time linear ordinary least squares regression, time exponential ordinary least squares regression, generalized linear models, and logistic GLM. Data came from a water utility in Texas. The authors fit the various models and compared actual versus predicted breaks, discovering that Poisson GLM is a better predictor of breaks (zero breaks and non-zero breaks) than the time linear model. But the Logistic GLM is a better fit for the data, since it can handle a large percentage of zeroes in the data set (referring to pipes that do not break).

Kleiner[9] developed a cost effective distribution network renewal plan using a dynamic programming approach, which sought to minimize the total discounted costs associated with rehabilitation of pipes. A procedure known as Multistage Procedure for Rehabilitation Analysis of Water Distribution Systems was used to solve. Kleiner[10] also developed a new study using a computer model called Individual Water mAin Renewal Planner, which is used to examine historical pipe breakage patterns using a non-homogenous Poisson process to model individual breaks. The model uses dynamic factors such as climate and pressure change as part of its analysis.

These models cited in literature are innovative models that can provide a means for predicting breaks and other events based on historical information. They can be of benefit to water distribution asset managers if they are implemented as part of a formal asset management plan. So far, many of the models used by water utilities are off-the-shelf software packages or basic statistical models such as Weibull analysis.

3. METHODS

Our methodology in this study is to perform semi-structured interviews with water utility firms, specifically including asset managers, to understand the role of statistical models in long term asset planning. The goal is to develop understanding of the role of models and identify the strengths and weaknesses of models, the perceived benefits of using models, and ways to improve models to make them a more effective tool for utilities.

Semi-structured interviews as part of a mental models framework in the context of risk communication are described in Johnson-Laird and Morgan[11],[12]. The goal is to improve risk communication through a five step process, including the creation of an expert model, conducting mental models interviews to elicit beliefs about the hazard, followed by structured initial interviews, and drafting and evaluating risk communication measures.

In mental models interviews, the goal is to examine a person's mental models, or psychological representations of situations. Researchers have used a mental models approach to determine a person's system of beliefs regarding a topic at hand. The mental models framework consists of a semi-structured interview process designed to give the respondent the most flexibility in providing a response, while limiting bias that the interviewer may inadvertently provide. Mental models interviews have been used to gather information on subjects' beliefs about a number of topics. One such example, done by Magzan[13], is to use mental models to explore leadership effectiveness in business environments.

The first step in developing a mental models framework is to create an expert model, or influence diagram, that summarizes the current expert knowledge. Based on our literature review, we propose an influence diagram, shown in Figure 1 below, which captures the current state of expert knowledge regarding models in water distribution asset management. This influence diagram serves as a basis for developing our interview questions.

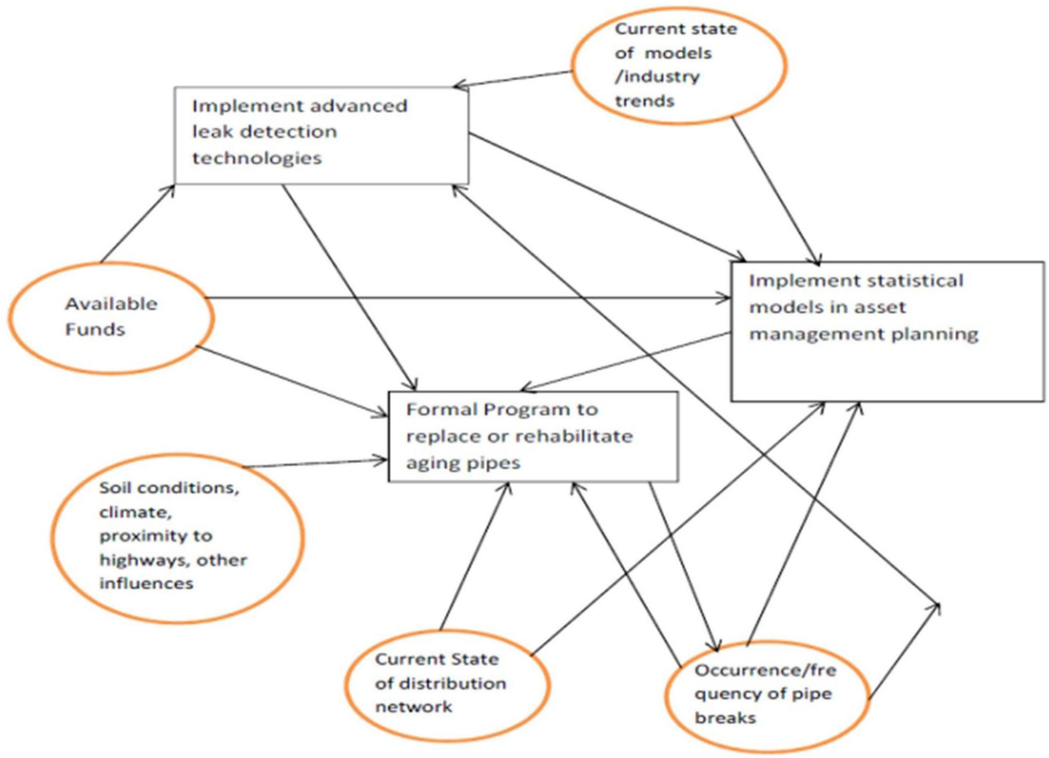

Figure 1: Influence Diagram for Water Asset Management

Our approach was to contact various water utility firms located across the United States, each serving a medium to large sized city. We began by holding informal discussions with asset program managers or other individuals deemed knowledgeable with asset management programs and the use of models. We chose six participants based on their degree of agreeability to the study and knowledge and experience about the subject matter. We provide a list of the utilities and qualifications of the various individuals in Table 2 below.

Table 2: Description of Participants in our Study

	Title of Individual	Years of Experience at Utility
Utility A	Asset Management Coordinator	5
Utility B	Senior Engineer	9
Utility C	Asset Management Program Manager	4
Utility D	Manager, Water Infrastructure Planning	10
Utility E	Asset Strategy Manager	15
Utility F	Managing Engineer, Infrastructure	4.5

We found that the utilities were agreeable to participate and were receptive to our inquiries. We provided information about our study including our data collection, which consisted of audio recordings of our conversation. We ensured that the confidentiality of the data would not be compromised, and that the utilities would not be mentioned by name in our report.

Our mental models interviews were developed upon completion of the informal interviews, which provided background information to help develop our formal interview questions. Following the guidelines provided by Johnson-Laird and Morgan, we kept the questions as open-ended as possible, to exclude biases or intervention on the part of the interviewer. The questions were worded to allow the respondents to answer as freely as they wished. Our interview checklist consisted of several question topics and subsequent follow up questions. We began with an initial, open-ended prompt, to begin the discussion, and as the respondents brought up various topics, we followed up with the topic questions. In this way the respondents dictated the flow of the conversation. We have included a copy of the interview checklist in Appendix A.

We archived the data recordings and transcribed the conversations into documents. These transcriptions were as verbatim as possible, though we removed words such as 'uh' and 'um' which we decided were irrelevant and inapplicable to our study. The rest of the transcription captured the recording as closely as possible, and we verified by comparing the audio recording with our written transcription. We did not conduct member checks, however we felt our transcriptions were as accurate as possible and did not compromise the integrity of the data. We ensured rigor by verifying that our questions captured the spectrum of issues pertaining to statistical models with our participants. All participants either admitted they had no additional information, or added information that was already mentioned and deemed a relevant theme, so we considered this as no new information.

Our goal was to understand themes that pertain to facilitating or inhibiting the use of models in water asset management. These themes were derived from coding the interviews using Atlas software which allows for analysis of qualitative information. We found codes that represented the salient points of the respondents' answers, including areas for concern or areas in need of improvement. We validated these codes by counting the frequency of occurrence across all interviews, providing support that these were indeed themes that were regarded as important by the majority of respondents.

4. RESULTS

4.1 Codes

Our analysis began by compiling a list of codes and their relative frequency. A Pareto chart showing the codes, their relative frequency, and the cumulative distribution is shown in Figure 2.

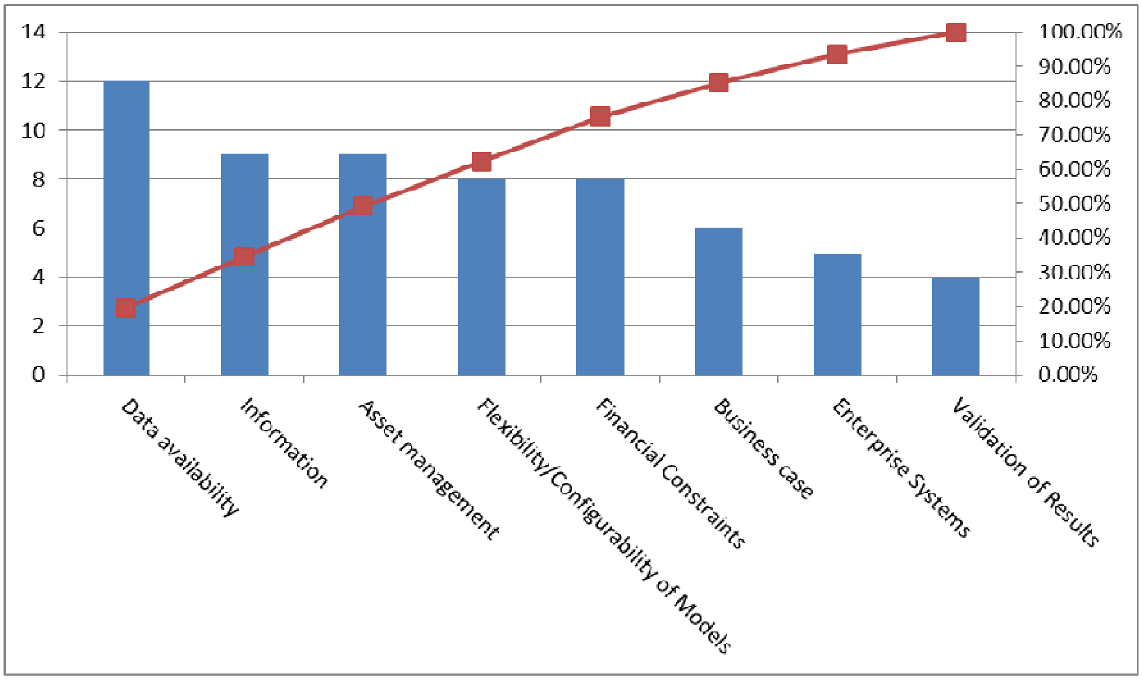

Figure 2: Pareto Chart for Descriptive Codes

We describe the codes in detail. First, data availability refers to the idea that the effectiveness and accuracy of the models is dependent on the availability of the data. The model can only work if the right data is available and readily accessible. Numerous respondents claim that since this is the input to the model, a lack of proper data can hinder the accuracy of the models, or make models unsuitable for use by a utility.

Next is information, which refers to information about the various models, how they work, what can they inform utility managers, and what data is required. Asset managers need to know this information before they can plan to use models in their asset programs. A common theme among the respondents is that there is not enough information about models and how to use them. One participant recommended a project be underway to review models to see what they require and what they produce. This would serve to inform asset managers how to use models and what it takes to use them.

Asset management refers to the understanding that an asset management program framed using statistical models as a basis for determining replacement prioritization is recognized as necessary. Most participants we surveyed recognized that models are becoming more useful to asset management. Predicting a schedule of prioritization is especially beneficial since utilities are dealing with limited capital improvement budgets, and planning using models can help determine the most effective way to use available funds.

Flexibility and configurability of models refers to the idea that utilities want to see models that are user friendly and can be customized to meet the specific needs of the utility in question. Water utilities want to be able to configure models to handle various inputs, and be able to adapt to changing geographical or spatial conditions Too many packages are canned and do not allow for user configurability. Since each utility has different sets of pipes, different data collection methods, and different needs for models, each model needs to be tailored for a specific utility, or have the ability to undergo customization to meet specific needs.

Financial constraints refers to working under limited financial resources. Utilities are working with restricted capital expenditure budgets which limit how much they can spend on models and training. Some utilities understand that empirical testing of pipe, using methods such as acoustic leak detection, can be very expensive, and are seeking other means of assisting asset planning. Models can be a cheaper alternative to physical testing, and all the utilities surveyed agreed that a balanced approach to asset management, using both modelling and physical testing, is ideal.

Similar to financial constraints, there is an idea that a business case needs to be in place to justify the expense of models. Part of this involves communicating the need for models to executives or a management team who may not be able to understand the technical details of models and may not understand their value.

Another theme is enterprise systems. This refers to a trend towards developing enterprise systems-based asset management, which encompasses business processes and information flow across multiple systems, including engineering, accounting, and operations and maintenance. An enterprise system based asset program would allow all departments to have information readily available and shared within the system's network, allowing for more collaboration and data flow. Modelling would be easy to share with other departments, resulting in prompt response from financial departments towards predicted capital allocation budget requests and operations and maintenance response towards predicted pipe failures. Models need to be able to fit in the context of enterprise systems, in order to play a role in future asset management.

Lastly is validation of results. Most of the participants cited want to see validation of the statistical results by comparing them with historical data or pipe condition assessment (physical pipe inspection). The utilities surveyed had varying responses to this theme. One stated satisfaction with the results, another saw a need to improve the accuracy of the models, but was generally satisfied, and another claimed there was no satisfaction with models and accuracy needed to improve greatly.

4.2 Questions on Current State of Models and Future

Many of our interview questions were specific towards the use of models, how accurate they were, and what benefit was derived from them. However, we did have a cohort of questions that were directed towards the current feelings towards models as well as their future. We feel these questions are among the most important in our interview list, and we have selected three questions and presented them here. The responses serve to highlight the issues that need to be addressed before models become more prevalent among water utility firms.

- **Question 1:** Are water utilities quick to embrace models?

  Answers: "*I think yes and no. It depends on where you are in the utility. From a planning context, people have a very high regard, an early adopter phase to modeling. From an operational context, the value of modeling is not seen as much. However that is changing, as technology is becoming more commonplace.*"

  "No, generally utilities are comfortable doing what they have been doing. Moving beyond that is hard. Especially with a model that takes understanding that people might not have. It's less likely that it will be embraced. There's recognition that there is value there but it's hard to know what to do."

  "I want to say no. In our utility, we do use them. We are a very large utility, and have therefore the need and resources, based on the number of assets, we can apply statistical models, rather than rely on observation. With smaller utilities, with a smaller number of assets or a smaller geographical region, there is a thought that people know their assets well, better than a computer model. This is an old school mentality."

Based on these responses, it is evident that there is still some hesitation on the part of utilities to embrace models. Many users recognize the value of models, but there are some who are used to practicing old ways of managing assets, and are not comfortable embracing new technologies or new methods.

- **Question 2:** How do you feel about the future of models in asset management planning?

  Answers: "*Well I think there's a place for them. But they can't live as standalone entities. The biggest challenge is that the public utilities lag behind the private utilities by a decade or more, in terms of their IT systems or internal systems and the enterprise connectivity of them. None of our data is accessible with other departments. We are working to improve that with the master data management system, so everyone is working with the same source data. Everyone's data goes into it, so there is a system of record, so you can see what is happening."*

  "The models are going to be critical for developing asset management plans for the future. The systems are getting more and more complex, the operations and maintenance of the systems and rehabilitation are getting more and more expensive, the funds are going to be limited, even in the future so the models are going to play a critical role in the prioritization of your asset management program."

  "I think there's a great future for it. It is the way of the future. Modeling can give you answers to questions that there is no substitute for."

These responses are more optimistic about the future of models than the first question above. There is recognition that models will play a role in future asset management. Since water systems are complex, and utilities have limited funds to address capital improvements, facilitating prioritization of pipe replacement with models will become more important. Universally, the future for models is favorable, and there are beliefs that more utilities will embrace them to assist in managing complex systems.

- **Question 3:** How easy or difficult is it to incorporate models into your current asset management approach?

  Answers: *"It's very difficult, or easy and difficult. The difficult part is that you have to have data, and you may not have the data, or the data is somewhere and you don't know where it is. And it costs money, and these days for us you have to plan ahead to get a budget allocation to do a project of that type. That's the difficult part, access to data and money."*

  "It's a challenge for us because we have so many different assets, it's not a challenge where we use statistical analysis. It is a challenge where we bring data together, our challenge is we are drowning in data. We need to bring it together, in one area, so we can apply analysis. Not a challenge to do modeling, it's more of management of information."

  "It's a lot of work, because the more thorough and sophisticated the model is, the complexity is exponential. An additional factor requires connections to other systems, you just can't do it out of the blue. So it can get complicated quickly."

We see from these responses that there are issues with data collection and storage, and that the model is only as useful as the data that is provided. Models need the right data available as well as systems in place to allow data sharing to ensure they can provide meaningful results. Management of information is just as important as the model itself. This is a reason why enterprise data systems are valuable – they allow sharing of information with teams and departments. Shared information allows departments to collaborate and meet needs in a more effective manner.

5. DISCUSSION AND CONCLUSIONS

Based on the information we have collected, it is clear that utilities encounter difficulties introducing models into their asset management plans. The reasons for this include lack of available data, lack of user-friendliness and configurability of models, difficulties tying models into current asset management plans, lack of information about models and their benefits, and lack of funds to pursue modeling.

There are several ways to ameliorate these issues. A first step would be to provide guidelines to water utilities as to how models are used and what benefits they can provide. Utilities can also be informed about what data they need to have collected before they can effectively use models. Introducing such information can help clear up misconceptions and assuage fears that utilities may have, encouraging more adoption of models. A second step would be to encourage the use of enterprise systems to facilitate data sharing and information flow. Models would be more valuable in such a context because their predictions and outputs would be seen by many departments and can be addressed in a timelier manner. The benefit would be swifter action to address issues such as pipe breaks and loss of water flow.

However, despite the issues stated above, all the utilities surveyed remain optimistic about the future of models, citing the need for developing more effective asset management plans. There is recognition that models can provide a useful tool in conjunction with condition assessment to allow utilities to better manage their assets.

We can take the information that has been synthesized from this survey and create a network diagram, as shown in Figure 3 below.

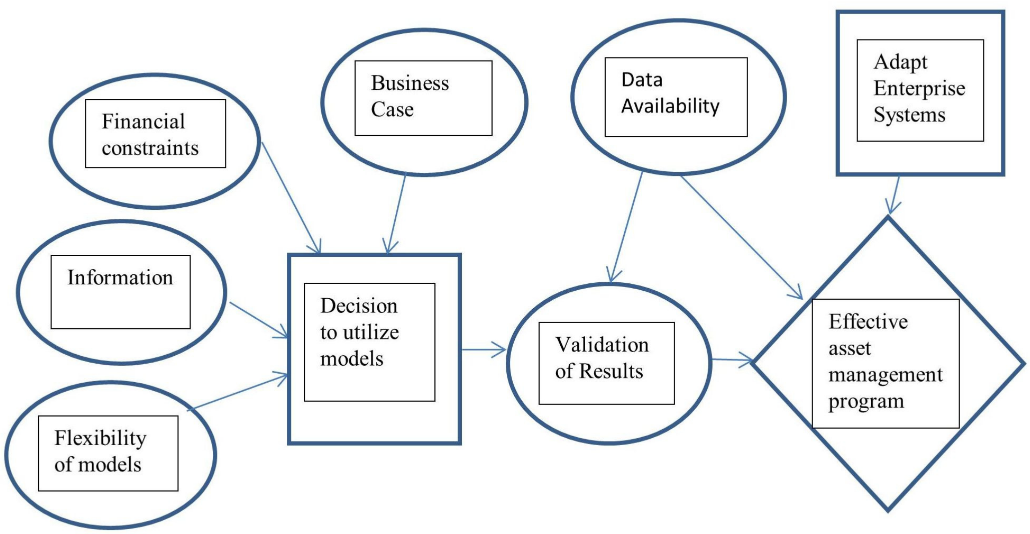

Figure 3: Network Diagram for Models in Water Distribution Asset Management: Influences That Contribute to an Effective Asset Program .

This diagram shows the various influences that play a role in the decision to adopt models and their impact on developing a more comprehensive and effective asset management program. We derive this diagram based on our knowledge of the major issues obtained from our interviews. We have labelled several of the nodes such as financial constraints, information, and data availability as unknowns since in many cases they are not readily quantifiable and information is not readily available to the decision maker. This diagram shows the major factors that contribute to the success of statistical models in asset management planning, and what information decision makers need to be aware of before they choose to implement models. When comparing Figure 3 to our influence diagram (Figure 1), we see that we have gained more information about the types of influences that are directly affecting models and their use in water asset management, including data, enterprise systems adaptation, and flexibility of models. These influences were not apparent from our literature review, which explains the value and contribution of our semi-structured interviews.

In conclusion, we have shown that our mental models interview approach can help illuminate answers to questions regarding the role and outlook of statistical models in asset management planning. We have seen a number of requirements that need to be satisfied for models to be viable. These include robust and available data, more information about the model requirements and needed inputs, and user-configurability and ease of use of the model. We also note that utilities are constrained by limited financial resources and many times a business case must be made to justify the use of models. Utilities also want to see increased accuracy of models when compared with physical testing results. This serves as validation that the model is working properly. Utilities also want to see models become part of advanced enterprise systems that connect multiple parts of the utility firm to allow sharing of data and information. Despite these issues, we see that all of our study participants agree that the future of models is optimistic. Models will allow utility firms to make more informed decisions regarding rehabilitation and repair of pipes, and will be a valuable part of future long term asset planning.

Our next stage of the study would be to use a Q-sort methodology to elicit opinions and determine shared ways of thinking among water utility firms regarding the use of models. Current existing knowledge of the Q sort methodology is widespread, and the methodology is widely used in areas relating to humanities, social sciences, and psychology. However, to the authors' knowledge, no current Q sort study has been performed on water industry experts. We believe this will shed more information about the role and future of models, and how they can better assist utility firms.

Acknowledgments

The authors are supported for work on this project by a grant from the National Science Foundation Project # 1031046.

References

[1] Stubbart, John M. *AWWA water operator field guide.* AWWA, Denver, CO. 2004.

[2] *Rehabilitation of Water Mains: AWWA Manual M28.* AWWA, Denver, CO. 2001.

[3] *Water Infrastructure at a turning point: the road to sustainable asset management.* AWWA, Denver, CO. 2006.

[4] Alegre, Helena. *Strategic Asset Management of Water Supply and Wastewater Infrastructures*. IWA Publishing, London, 2009.

[5] Pelletier, Genevieve, A. Mailhot, and J.P. Villeneuve. *Modeling Water Pipe Breaks – Three Case Studies,* Journal of Water Resources Planning and Management. Vol. 129, No. 2, pp. 115-123, (March 2003).

[6] Clark, Robert et al. *Condition assessment modeling for distribution systems using shared frailty analysis,* Journal AWWA Vol. 102, No. 7, pp. 81-91, (2010).

[7] Debón, A. Et al. *Comparing risk of failure models in water supply networks using ROC curves*, Reliability Engineering and System Safety, vol. 95, no. 1, pp. 43–48, (Sep. 2009).

[8] Yamijala S., S. D. Guikema, and K. Brumbelow, *Statistical models for the analysis of water distribution system pipe break data*, Reliability Engineering and System Safety, vol. 94, no. 2, pp. 282–293, (Feb. 2009).

[9] Kleiner Y., B. Adams, and J. Rogers, *Water distribution network renewal planning*, Journal of Computing in Civil Engineering, vol. 15, no. 1, pp. 15–26, (2001).

[10] Kleiner, Y and B. Rajani, *I-WARP: Individual Water Main Renewal Planner*, Drink. Water Eng. Sci., vol. 3, no. 1, pp. 71–77, (2010).

[11] Johnson Laird, P.N. *Mental Models: towards a cognitive science of language, inference, and consciousness.* Harvard University Press, 1983, Cambridge MA.

[12] Morgan, M. Granger, et al. *Risk Communication: A mental models approach.* Cambridge University Press, 2002, Cambridge, UK.

[13] Magzan, Masa. *Mental Models for Leadership Effectiveness: Building Future Different Than the Past.* Journal of Engineering Management and Competiveness. Vol.2, No. 2, pp. 57-63, (2012).

[14] Saldana, Johnny. *The Coding Manual for Qualitative Researchers.* 2nd Ed. Sage Publications, 2013, Los Angeles.

[15] Seale, Clive and D. Silverman. *Ensuring Rigor in Qualitative Research.* European Journal of Public Health. Vol.7, pp. 379-384, (1997).

Appendix A: Our Interview Checklist

The following is our interview checklist for our semi-structured interview approach. We check off each topic as it is mentioned by the participant, and proceed with the set of sub-questions below each section, one by one.

The interview begins with an initial prompt: "What I'd like to ask you is just to talk about statistical models and their role in water asset management. That is, just tell me what you know about models."

__I__ Benefits/Costs

- __ Can you tell me what are the benefits of implementing statistical models?
- __ Do you feel that models are worth the cost of implementing them?
- __ Do you feel you are able to make better decisions regarding your asset management plan when using models?

__I__ Information

- __ How did you find out about statistical models and their use in asset management?
- __ Did you hear about models from industry trade journals or conferences?
- __ Do you feel there is enough information about models that utility firms are well-informed about them?

__I__ Asset Management

__ What role do statistical models play in asset management planning?

- __ Do you feel a balanced approach between models and empirical testing of pipes is ideal?
- __ Are water utilities quick to embrace models?
- __ How do you feel about the future of models in asset management planning?

__I__ Implementation/Choosing Models

- __ How easy or difficult is it to incorporate models into your current asset management approach?
- __ How do you choose a model and evaluate it?
- __ What do you feel about the accuracy of current models?
- __ Are you satisfied with the performance of models that you have used?
- __ What can be done to improve models so that they can better assist utility firms?