Infrastructure Systems Renewal and Replacement Model Using Probabilistic Forecasting

J.D. Solomon, PE, CRE, CMRP

Adam Sharpe

Key Words: Probabilistic Forecasting; Monte Carlo Simulation; Reliability; Infrastructure; Renewal and Replacement; Planning

SUMMARY & CONCLUSIONS

Common approaches to addressing renewal and replacement rely on limited analysis of maintenance strategies and on deterministic approaches with finite, best approximation inputs. These over-simplified approaches yield results that greatly overestimate, or in many cases greatly underestimate, uncertainty and risk. Decision makers are often placed in a poor position to defend the annual budget as well as to properly program the needed capital investment to maintain and re-establish system reliability. Probabilistic analysis provides an alternative approach which clarifies uncertainty and risk in decision making.

1 INTRODUCTION

All infrastructure assets deteriorate with time and use. To maintain the effectiveness and value of an asset, renewal work should be performed periodically, and when the asset has reached the end of its functional life, it should be replaced. At the heart of any asset management program is the effort to preserve the existing system's performance and reliability by anticipating future renewal and replacement (R&R) needs and to ensure that adequate and timely funding is planned into the capital improvement program.

Organizations continue to be under economic pressures related to the uncertainty and risk associated with the planning for R&R of their infrastructure. Most of the common approaches to addressing this issue rely on deterministic approaches with finite, best approximation inputs. Unfortunately these deterministic approaches yield results that greatly overestimate, or in many cases greatly underestimate, uncertainty and risk. Such results fall short of providing sufficient clarity for decision makers.

In contrast, probabilistic approaches offer a more robust method to address uncertainty and risk. This approach is particularly relevant at the current time, while decision makers continue to face aging infrastructure and widening funding gaps.

2 FORECAST METHODOLOGY

Deterministic models use point estimates for each input parameter and from these estimates a single, specific outcome is predicted. Such models are rooted in the theory of determinism, which holds that every outcome event is the result of preceding, or antecedent, events. Some of the many problems associated with forecasts based on this methodology include: the belief that all preceding events can be precisely determined; the failure to account for interdependence among the input events; input events have been accurately measured, interpreted, and reported; and the misconception that precise models can be developed to capture the relationship of the inputs and output events.

More simply said, the primary problem associated with this approach is its failure to account for uncertainty. It is an oversimplified approach that reduces the applicability and accuracy of the associated forecasts. An equally important ramification is that it normally drives the forecaster to collect as much data as possible, and with a high level of measurement, in order to avoid being considered wrong with the associated prediction of the future. Many organizations waste tremendous amounts of resources collecting data to potentially improve a deterministic forecast while at the same time not gaining any advantage in the decision making process.

Probabilistic models, on the other hand, use statistical distributions associated with input parameters rather than single point estimates. According to ISO 31000 [1], the international risk standard, this approach provides a method of taking into account uncertainty in systems for a wide range of situations. Probabilistic models evaluate the range of possible outcomes and the relative frequency of the values in that range for quantitative measures of a system such as cost, duration, throughput, demand and similar measures. According to ISO 31000, it is primarily used either as an uncertainty propagation on conventional analytical model or for probabilistic calculations when analytical techniques do not work. The primary use of a probabilistic approach in the application for forecasting R&R needs described in this paper is primarily related to the former.

The ability to quantitatively address risk and uncertainty is the most obvious benefit of R&R forecasts. However, one advantage that is often not fully appreciated is the ability to use the model in cases where input data are not as perfect as desired by a purely deterministic approach. The outcomes of the probabilistic analysis can be used to direct data collection activities to the areas of greatest value, thereby saving cost and time by directing limited resources to improve data quality where such data is most critical to the accuracy of the forecast. Such data is classified as highly valued data.

2.1 *Forecast Model Objective*

A best practice for any type of problem solving and decision making is to first establish the problem statement. The model objective is specifically to help decision makers consider meaningful and reliable information from forecasts associated with the R&R needs of existing systems. A secondary objective is to provide insights into what is highly valued data and to provide an indication where more clarity in the base data and associated information is needed. The R&R model objective is not intended to make the R&R decisions, to be a black-box for establishing values and trade-offs, or to replace logically correct reasoning to make decisions without human input.

Simply, the R&R model is intended to forecast funding requirements to maintain or re-establish the reliability functions for existing systems and facilities. The R&R model discussed herein is not intended to serve as a failure prediction model, risk prioritization tool, nor a day-to-day maintenance planning tool. The model is also not intended to be a risk prioritization tool that can be used in the absence of more formal systems and reliability engineering analysis, especially in the case of complex systems. Certainly key insights can be gained from the model's outputs. Rather, the model's outputs can provide reliable insights and these in turn can be used in a larger decision making process.

2.2 Deterministic Model as a Base

Developing a deterministic model should be the next step following validation of the problem statement. Groenendaal [2] and others have noted that the logic developed for a deterministic model is not necessarily appropriate for a probabilistic analysis and the direct use of the deterministic model may produce unrealistic results from the probabilistic model. The potential for these concerns is acknowledged here and should be considered by the model developer. However, a more detailed discussion of these concerns is beyond the scope of this paper.

A deterministic model should be developed and preferably validated in Microsoft Excel® as the preferred spreadsheet tool, since most commercially available probabilistic software readily use Excel. Input parameters for deterministic R&R models commonly include a complete list of assets, asset replacement values, estimates of remaining useful life of each asset, maintenance renewal and replacement strategies (frequencies), and some form of asset group or asset type. Primary data sources include the computerized maintenance management system (CMMS), financial records, and equipment manuals. Interviews with staff (subject matter experts) and some level of additional field data collection is also required.

An important step that is needed between the deterministic and probabilistic models is the development and/or confirmation of input parameter distributions. Generic probability distributions are available for different asset classes; a number of public and private sector organizations also have compiled distributions from previous projects and testing. However, it is critically important that the distributions be confirmed with the system owner based on actual experience and the specific operating environment. Review of reliability performance data, interviews with subject matter experts, and the use of statistical tools to validate the shape of modified or new distributions should be performed as best practices. A more detailed discussion follows in a later section of this paper.

Once the values used to define the input parameter distributions have been determined they should be used for a sensitivity analysis of the deterministic model. This analysis is completed to identify the inputs that the model is most sensitive. Typically, the first run of the sensitivity analysis is a quality control (QC) review of the model inputs before the probabilistic analysis. A review of the sensitivity analysis results will provide the ability to understand the impact each input has on the model and how that agrees qualitatively with the logic of the model, as well as compare it to the actual operating context.

There are a number of potential issues related to the input data. These include: completeness and accuracy of the asset registry; consistency with the basis of replacement asset values; the lack of or inconsistency of formal maintenance strategies for renewal and replacement; and the subjective nature of predicting estimated remaining useful life in complex systems (even field condition assessments do not correlate directly to this prediction). Without further discussion of the details of these potential issues, it should be apparent that deterministic approaches and associated point estimates are not as well suited as probabilistic models to account for uncertainties associated with input data.

Regardless of these limitations, the deterministic approach used as a base model is a highly effective tool for proofing and validating the overall logic, calculations, and links prior to performing the probabilistic analysis. An associated sensitivity analysis can be extremely helpful in framing where more highly valued information is needed.

Following the completion of probabilistic analysis, using Monte Carlo simulation, a statistical analysis of the outputs should be performed. The types and degree of analysis will vary depending on the type of analysis being performed. However, at a minimum this should include basics such as output data set mean, median, mode, probability density function plots, cumulative density plots, and comparative analysis with sensitivity analysis that was performed with the deterministic model.

2.3 Additional Considerations for Probabilistic Model

Selecting the most appropriate probability distributions to represent model input variables is fundamentally important in terms of developing a probabilistic forecast. Solomon, Sharpe, and Seachrist [3] have documented the issues and alternatives related to the development and selection of probability distributions for R&R forecasts. It is both a science and an art.

Three major challenges to establishing useful probability distributions include the nature of the data (whether continuous or discrete scales have been used), the underlying quality of the available reference data, and the potential lack of meaningful lifetime data. Several techniques beyond classic statistical analysis are available to address these issues in a more practical manner. Spetzler and Stael von Holstein [4] developed a methodology for developing a cumulative distribution function (CDF), which remains heavily relied upon. Efron and Tibshirani [5] are credited with creating the bootstrapping method which is used to determine how well data set fits with an assumed data set characteristic. Keelin and Powley [6] advocate the use of quantiles, namely the quantile density function, as yet another way of prescribing a probability density function. The authors have used all three of these techniques to support the development of probability distribution in R&R models.

Issues related to underlying R&R strategies, especially related to repairable assets, is another major issue that has to be considered. However, this issue can be addressed by a probabilistic approach whereas it confounds a deterministic one.

Essentially the issue comes down to what is meant by repair, or renewal, and how these actions change the overall makeup of a given system and affect the system behavior differently due to different repair approaches. For example, repair of a repairable system may not mean replacing any part or asset; repair may be simply lubrication, cleaning, adjusting settings, or re-calibrating the system or subsystem. In other cases, repair or renewal of a given asset within a system may improve or degrade the performance of other assets and require a modification to their R&R frequencies. Therefore a system of assets, especially repairable assets, should not be treated over-simplistically since their behavior is dependent, not independent, on other assets, parts, and maintenance activities within the same system.

The practical aspects related to the development of the R&R model include the manner by which to establish "arrival times" for maintenance activities, the magnitude of costs associated with those activities (some will be capital and some will be annual operations and maintenance (O&M) budgets), and how to address the interdependencies within the same system or subsystem.

3 DISCUSSION

For a first generation model, the insights gained from the development of the forecast can be just as valuable as the projected numbers themselves. A number of insights can be gained from the forecast model development ranging from key data management issues, such as the additional asset attributes in a CMMS, to financial planning insights, such as inconsistencies in asset valuation information.

Probabilistic analyses are a highly effective tool for structuring problems and gaining insights about key forecast data inputs. When done properly, they improve a decision maker's understanding of risks, business value drivers, and the sensitivities of key decisions. They also provide an understanding of the relative importance and interdependencies of key variables, and in turn the value of both acquiring additional information and the potential areas for business process improvements.

Probabilistic analysis provides the decision maker with a better understanding and sensitivity of the portfolio as a whole. For example, this approach provides insight into whether a large number of lower value assets (such as actuators) create more uncertainty, and in turn risk, than a small number of higher values assets (such as high service pumps).

The understanding of uncertainty at the asset and portfolio level provides the opportunity to identify mitigation strategies to manage this risk (procurement strategies, warehousing strategies, O&M strategies). In a number of cases, the authors have found that organizations can maintain or improve system reliability more cost effectively by choosing a management strategy for risk mitigation rather than a more invasive renewal or replacement activity.

Additionally, staff engagement and the capture of institutional knowledge is an ancillary benefit of probabilistic analysis. The forecast development process for a probabilistic forecast requires more staff input than with deterministic approaches, primarily in the form of a better understanding of maintenance strategies and failure data. In the verification phase, the review of the R&R forecast model outputs provide a meaningful framework to engage operations and maintenance staff in a dialogue on R&R practices, maintenance practices, asset data collection and the importance of a CMMS. Staff is typically more engaged in the output results from a probabilistic analysis than a deterministic analysis because a probabilistic approach better reflects the uncertainties that they contend with each day.

The following sections provide additional insights from the authors' practical experience utilizing probabilistic analyses for forecasting R&R funding demands.

3.1 Evaluation at the System or Subsystem Level

As a best practice, reliability prediction, and therefore R&R forecasting, should begin at the system level. As the system becomes more closely defined it can be extended to more detailed levels such a sub-systems and individual components. In principle, eventually it is necessary to extend the analysis to individual parts; however, at the lower levels of analysis there is greater uncertainty inherent in predicting the reliability of the whole system. In practice, it is important to remember that many system failures are not caused by the failure of parts and that not all part failures cause system failures [7].

3.2 System Data Requirements

A meaningful amount of data is required for any forecast. However, most organizations collect too much data, and the cost of this data comes at high incremental costs. By accounting for uncertainty, the probabilistic analysis creates the best understanding by decision makers of what is highly valued data and allows data collection to be focused in these areas.

Highly valued data comes in either the form of collecting more system or asset specific information related to attributes such as size, capacity, horsepower, or most recent maintenance activity. It also includes the need to better assess remaining useful life by some form of condition assessment.

For systems or assets that are not driving the R&R forecast, a secondary advantage of a probabilistic approach over a purely deterministic approach is the ability to more confidently use a reasonable range of default parameters and distributions rather that spending valuable resources on collecting or analyzing data which has little impact on decision making. Focusing more resources on the assets that most impact the magnitude and sensitivity of the forecast and spending less resources on those that least impact decision making is key.

Again, focus more resources on the assets that impact the magnitude and sensitivity of the forecast most, and less resources on those that impact decision making the least.

3.3 Asset Replacement Values

Asset replacement values are critically important to the development of a meaningful R&R model. Also known as replacement asset value (RAV), it is also a key component of benchmarking of best practices by such organizations as the Society for Maintenance and Reliability Professionals (SMRP). Gulati [8] cites inconsistencies related to RAV across different organizations as a primary detriment in performing meaningful benchmarking. The same can be said as it relates to R&R forecasting. These inconsistencies are also significant within many organizations where financial, engineering, and O&M divisions calculate and track asset values in very different ways and in different databases.

For developing well defined and accurate asset replacement values, a meaningful amount of cross-functional time should be allocated. In most cases, internal data will not be either complete or robust enough, and outside sources will need to be used to complete the dataset. The approaches and methods associated with developing RAV is beyond the scope of this paper.

3.4 Incorporating Criticality and Redundancy

The assessment of criticality of subsystems and individual assets is performed as a post-process on the model outputs. It is not part of the base R&R forecast model for a number of reasons. The primary reason is that the evaluation of complex systems is not necessarily intuitive. Complex systems, especially those that are designed as critical systems, often utilize different techniques for the system to achieve sufficient levels of fault tolerance or fault avoidance. These techniques may include, but are not limited to selecting higher reliability components, over-sizing certain components, operating components at lower stress conditions, and using one of several different forms of redundancy. Inherently, fault tolerance allows a complex system to perform despite the presence of an internal component failure and, therefore, confounds casual assessments of the critical nature of any one component. The model, in its base form, provides results related to which components or subsystems are most important from a financial perspective without prematurely or incorrectly assigning priority. This supports the justification to direct resources toward a more comprehensive assessment of complex systems, and minimizes the potential effects that increase or decrease component prioritization based on intuition.

For similar reasons, the potential presence of redundancy is not included as a prioritization factor in the base R&R model forecast. The presence of redundancy, or alternative means of production, is a feature of the operating context which must be considered in detail [9]. And when component failure indication equipment is not utilized, regular inspection and testing of duplicate components in necessary to verify that designed redundancy actually exists in service [10]. As a matter of best practice, the assessment of redundancy in complex systems should not be an informal process. Redundancy should be confirmed prior to adjusting R&R priorities.

Criticality and redundancy are aspects of reliability management that are based on reliability predictions. Reliability predictions must be an iterative process, since management objectives and risk must be balanced [11]. The first-generation R&R forecast model is intended to assess financial funding requirements based on current maintenance and reliability practices, and is not intended in its initial form to be a reliability prediction tool. Therefore, the model output is intended to be a driver for directing resources where more detailed, and iterative reliability analysis is justified to offset potential investments and associated risks.

4 COMMUNICATION OF R&R MODEL RESULTS

Effective communication of probabilistic analysis is cited by risk and reliability experts as one of the most difficult aspects of any forecast. Gegerenzer and Kahneman have documented that all types of people, including technical professionals, cannot intuitively understand statistics including the base meaning of probabilities [12, 13].

Solomon and Sharpe have outlined a number of best practices for communicating risk and reliability information to decision-makers [14, 15]. A communication plan should be developed before the modeler or analyst attempts to communicate the R&R forecast results to decision makers. This is consistent when attempting to communicate any type of technical data to end users. A basic outline is normally sufficient, and minimum (primary) considerations should include a written description of the target audience, the experience of the audience in the subject matter (here, forecasting and probabilistic methods), key communication objectives, and a list of supporting graphics to communicate the forecast.

Additional considerations depend on the nature of the referenced primary considerations. These include the type of presentation forum, the presenters of the information, the type of graphics, and the format, colors, and word choices of the presentation follow establishing the four key items listed above.

An overarching recommendation is to simplify the message to the extent possible. This includes using a finite number of visuals, avoiding detailed explanation of the process and its limitations, and to generally let the receiving audience ask questions at their level of interest. A summary of the five key visual presentations tools are recommended in the following subsections.

4.1 Time Series Plot of Funding Needs using a Box-and-Whisker Plot

A time series plot displays observations on the y-axis against equally spaced time intervals on the x-axis. For financial analysis, monetary values are represented on the y-axis and time (in years) on the x-axis. The major advantage of time series plots is that they allow the user to evaluate patterns and behavior in forecast data over time.

When performing a probabilistic analysis, the use of a box-and-whisker plot (sometimes called simply a box plot) is used to complement the presentation of the financial data in any given year. More specifically, a box-and-whisker plot is an aerial view of a histogram. The median value is synonymous in the presentation with the point forecast associated with a simple deterministic time series plot. Normally the 25th and 75th quartile are depicted as the "box" while the most extreme values, typically the 5th and 95th projections define the ends of the "whiskers." Exhibit 1 provide an example R&R forecast time series plot use a box-and-whisker plot.

The box-and-whisker plot allows for a visual representation of the risk associated with the forecast in one graphic. The greater width of the box in any given year implies more uncertainty or potential risk associated with the forecast. Boxes in any given year that are skewed in either direction from the median value also depict the potential for more upside or downside in any given year.

4.2 Renewal & Replacement Frequency Table

This table is most useful to O&M staff in understanding the R&R frequencies that are driving the base model. For ease of visualization, renewal events are plotted in orange and replacement events plotted in green across the entire planning timeline for each facility. By examining the table, O&M staff can appreciate potential lengthening or contracting R&R frequencies. In the case of multiple facilities, it can be quickly and visually determined if different R&R strategies are being used at different facilities for the same type of asset or subsystem.

4.3 Asset List for Targeted Capital Cycle

The forecast model should provide a list of assets anticipated by the probabilistic analysis for any given time frame, which is most commonly a capital improvement planning cycle. The use of a probabilistic forecast allows the user to expand (or contract) the subsystems or assets by the relative risk desired, i.e., 50% certainty or 75% certainty over any given time period. This fully captures the power of the

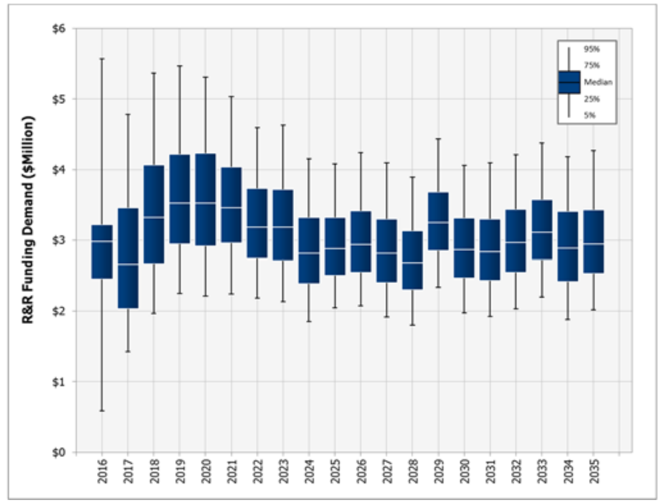

Figure 1 – Sample R&R Forecast

probabilistic analysis by allowing for comparative analysis based on the potential uncertainty around the timing of R&R events, rather than simply a fixed point estimate at a specific time interval.

4.4 Tornado Diagram

Tornado diagrams are modified versions of a bar chart. They are a classic tool used to communicate the results of a sensitivity analysis--providing end users a quick overview of the critical systems or assets and the relative risk involved with each. For an R&R model, the tornado diagram quickly shows which subsystems and assets create the greatest risk to the total funding demand over the course of the evaluated lifecycle. Tornado diagrams are also a powerful quality control tool relative to the deterministic model prior to performing the probabilistic forecast.

Exhibit 2 presents a sample tornado diagram, showing the input parameters that create the greatest variation from the base total R&R cumulative funding demand for an individual facility. The range presented in the sensitivity analysis results

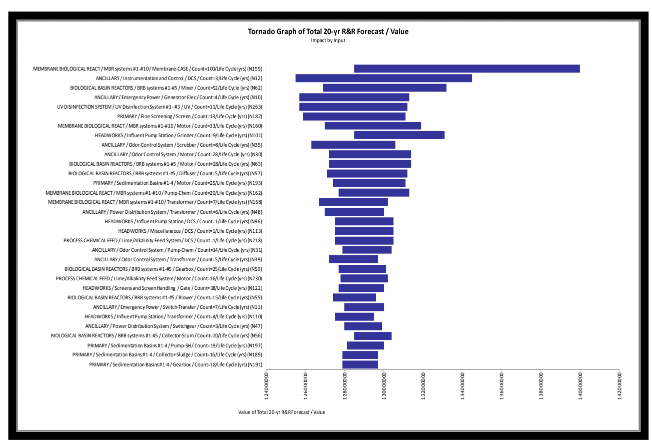

Figure 2 – Sample Tornado Diagram

represent the minimum and maximum potential impact derived from each input parameter, ordered from most to least important, often resulting in a tornado shape – hence the name of the sensitivity analysis results chart "tornado diagram."

4.5 Table of Historical Spending versus Forecasted Funding Needs

Another graphic that is a compelling tool is a summary of historical spending versus future forecasted funding needs based on the probabilistic model. This is summarized for both capital funding and for annual maintenance budgets. Normally this is best shown as a 3-year or 5–year average and it is important to decide on the quartile that will be represented from the probabilistic forecasts.

This comparison is powerful to not only provide context of the future funding demand forecast but to also understand historical R&R expenditures compared to future spending needs to visualize what changes needs to occur in funding, if any. The information also drives the understanding of maintenance strategies and the consequential impact of R&R funding demands; deferred maintenance certainly will be depicted clearly in the forecast of future R&R funding needs.

REFERENCES

1. International Organization for Standardization (ISO), *ISO 31000-2009: Risk management — Principles and Guidelines*, www.iso.org, 2009.
2. Huybert Groenendaal, "You got @RISK, now what? Building technically sound simulation models". Proceedings from The Risk Conference. Las Vegas, NV, 2013.
3. JD Solomon, Adam Sharpe, and Steven Seachrist, "The Leading Edge: Gwinnett County DWR'S Approach to Infrastructure Renewal and Replacement". Proceedings from the AWWA ACE Conference, 2014.
4. Carl S. Spetzler and C.S. Stael von Holstein, "Probability Encoding in Decision Analysis". Management Science, Volume 22(3), 1975.
5. B. Efron and R.J. Tibshiranio, *An Introduction to the Bootstrap*, New York, Chapman and Hall, 1993.
6. Thomas W. Keelin and Bradford W. Powley, "Quantile-Parameterized Distributions". Decision Analysis, 8(3):206-219, 2011.
7. Patrick O'Connor and Andre Kleyner, *Practical Reliability, Fifth Edition*, West Sussex UK, John Wiley & Sons. 2012, p. 134-135.
8. Gulati, Ramesh. 2012. Maintenance and Reliability Best Practices, 2nd Edition. Industrial Press, New York.
9. Moubray, John. 1992. Reliability Centered Maintenance, Second Edition; New York: Industrial Press Inc. p 208.
10. Bazovsky, Igor. 1961. Reliability Theory and Practice, Dover Publications, p. 170.
11. O'Connor, Patrick, and Kleyner, Andrew. 2012. Practical Reliability, Fifth Edition. West Sussex UK: John Wiley & Sons. p 143.
12. Gigerenzer, Gerd. 2002. Calculated Risks: How to Know When Numbers Deceive You. Simon & Schuster.
13. Kahneman, Daniel. 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux.
14. Solomon, JD, and Sharpe, Adam. 2013. "Regional Planning Using @RISK, PrecisionTree and Evolver, Part 2: Communicating to Decision Makers". Proceedings from The Risk Conference. Las Vegas, NV.
15. Solomon, JD, and Sharpe, Adam. 2015. "Tailoring Analytical and Communication Approaches for High Stakes Capital Infrastructure Projects". Proceedings from The Risk Conference. Chicago, IL.

BIOGRAPHIES

J.D. Solomon, PE, CRE, CMRP
CH2M
3120 Highwoods Boulevard
Suite 214
Raleigh, NC 27604

e-mail: jd.solomon@ch2m.com

JD Solomon is a Vice President with CH2M. He serves as a senior consultant focusing on maintenance & reliability, asset management, financial management, strategic decision-making, and master planning. He is a Certified Reliability Engineer (CRE), Certified Maintenance and Reliability Professional (CMRP), is certified in Lean Management, and is a Six Sigma Black Belt. JD has a Professional Certificate in Strategic Decision and Risk Management from Stanford, an MBA from the University of South Carolina, and a BS Civil Engineering from NC State. Governor Pat McCrory recently appointed JD to serve on the State's Blue Ribbon Commission to Study the Building and Infrastructure Needs of North Carolina.

Adam Sharpe
CH2M
3120 Highwoods Boulevard
Suite 214
Raleigh, NC 27604

e-mail: adam.sharpe@ch2m.com

Adam Sharpe is a project manager based in CH2M's Raleigh office. He focuses on value based decision making and risk management related to long-range planning, asset management, and strategic capital investment decisions. He is experienced in advanced data analytics and probabilistic modeling. Adam is a graduate of SUNY Cortland and NC State University, and has a professional certificate in Strategic Decision and Risk Management from Stanford University.

This peer-reviewed paper appears in the proceedings of the International Reliability and Maintenance Symposium (RAMS 2016)