DE GRUYTER
OPEN

Dynamic risk assessment method – a proposal for assessing risk in water supply system

Scott Fortmann-Roe, Ryszarda Iwanejko*, Włodzimierz Wójcik

Cracow University of Technology, Poland
Faculty of Environmental Engineering
Institute of Water Supply and Environmental Protection

*Corresponding author's e-mail: riw@vistula.wis.pk.edu.pl

Keywords: risk, system dynamics, water treatment, waterborne pathogens.

Abstract: System Dynamics is methodology for modeling and analyzing complex systems. Such systems can be characterized by interconnectedness and feedback. Applying risk assessment to the results of System Dynamics models is a challenge. Though in some cases the resulting time series data generated by a simulation may appear approximately random at a specific scale, there is often a high-degree of auto-correlation within the data series due to the deterministic nature of generation and feedback loops inherent in the system. This paper presents proposed Dynamic Risk Assessment Method (DRAM) that allows for the estimation of risk for system dynamics data series that appear to be approximately random. DRAM is based on standard risk assessment methods and is simple both to calculate and apply. In this article, the proposed method is applied to determine the risk connected with hypothetical costs of illness stemming from water supply system contamination with *Cryptosporidium*.

Introduction

The accurate assessment of risks in order to evaluate their severity and to develop appropriate responses and compensations is a key function of managers in a range of fields. Nevertheless, it is often difficult to integrate risk assessment into the tools that managers use. Here we propose a new risk assessment method applicable to the results of System Dynamics simulations. This form of simulation has been applied to a range of business management and public policy cases (Arquitt and Johnstone 2008, Dangerfield and Roberts 2000, Dangerfield 1999). As a case study, in this section we will look at the results from a model designed to estimate infection levels of *Cryptosporidiosis* within communities. *Cryptosporidium* and similar waterborne pathogens such as *Giardia* have caused a large number of disease outbreaks. For instance, in 1993, a failure of treatment in Milwaukee, USA led to over 400 000 people falling ill and 69 deaths. It is conservatively estimated that the direct costs (ignoring legal and capital investment costs) of that outbreak were $96.2 million dollars as for 1993 Corso et al. (2003). If the frequency of extreme weather events increases as predicted by global climate change, the risks from waterborne pathogens such as *Cryptosporidium* are expected to grow and the need for accurate methods to assess such risks will become increasingly important.

Procedure

The proposed method requires the execution of the following stages:
I. Obtain time-based data series through simulation.
II. Specify a number of thresholds and costs associated with violating given thresholds. Both upper and lower bound thresholds may be specified.
III. Determine the probability distribution that best models the data by one of two ways:
a. User specification.
b. Kolmogorov-Smirnov Goodness of Fit tests.
IV. Apply an autocorrelation analysis of the data. Locate the smallest separation interval at which the autocorrelation falls below a user-specified parameter. Divide the total number of data-points by this separation to obtain the number of approximately independent data-points.
V. For each threshold, calculate the probability of the threshold being violated during the simulation period based on the probability distribution and number of independent observations.
VI. Calculate risk by multiplying probability of violation by cost for each threshold and integrating to arrive at total risk. Risks for time periods other than those used in the analysis may be determined by linear extrapolation.

VII. Compare risk to acceptable levels and, if necessary, develop compensations to reduce risk.

Discussion

We will now explore at the steps of the DRAM procedure in more detail. For evaluation, the procedure was implemented using the software package Simgua (Simgua, v.2.8.2.) This implementation will be used to illustrate the use of DRAM. Fig. 1 is an illustration of Simgua's DRAM interface.

A simulation of a sample model will be used to illustrate the procedure. This model was the result of earlier research and predicts the level of infection of *Cryptosporidium*, a waterborne pathogen, within a community (Fortmann-Roe and Wójcik 2009). The variable that will be focused on in this analysis is the percent of the population that is infected with the pathogen.
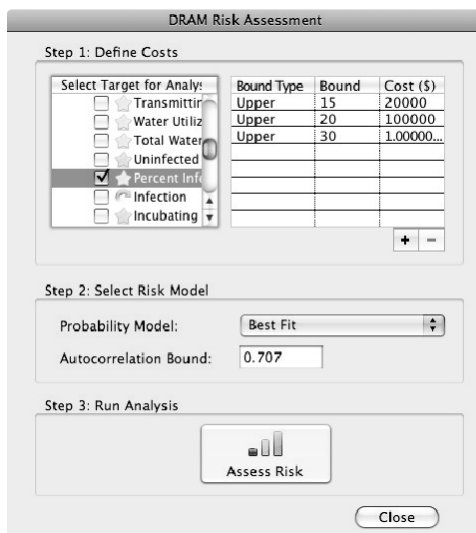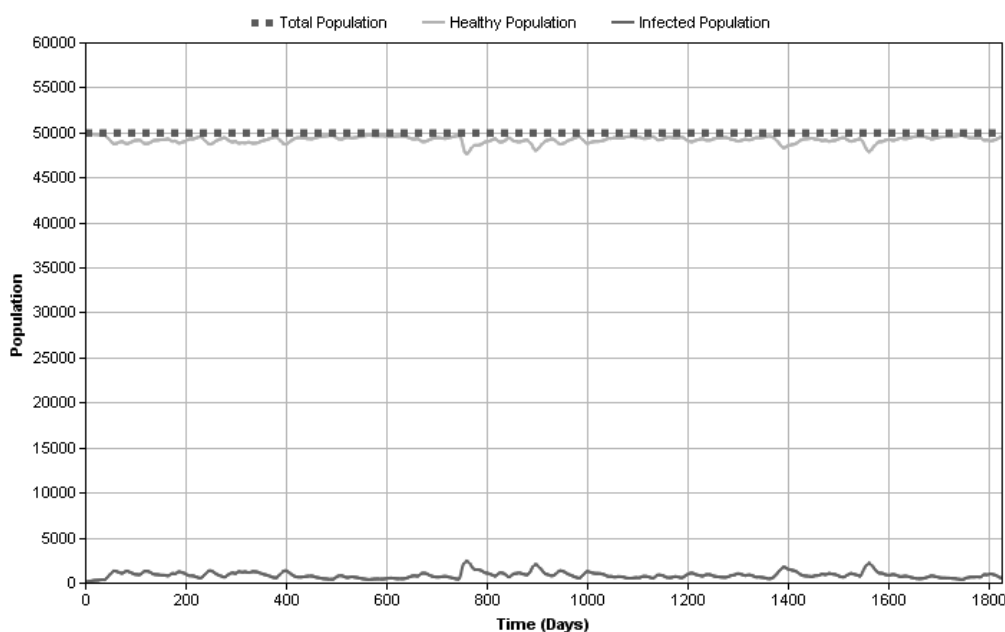

Fig. 1. A sample DRAM user-interface

Step 1

The infection model was implemented in Simgua and predicted infection levels for a specific community and set of conditions could be obtained by running the simulation. Figure 2 is an example of a portion of the simulation results from the model. It is important to note that the level of infected population includes all those infected with the disease. This value includes those with only mild symptoms in addition to asymptomatic individuals (approximately 65% of the number of infected).

Step 2

Thresholds based on the level of population infected and the associated costs to society were entered into the DRAM interface. Studies indicate that costs due to infection increased exponentially in relation to infection levels. All thresholds were upper bound thresholds given the nature of the problem. For other problems, such as the temperature or pressure of a boiler, one could conceivably have both upper and lower bound thresholds.

Step 3

The Simgua implementation of DRAM includes three probability models: Normal, Lognormal and Gamma Distributions. The user may specify which model to use or they may let the algorithm choose the best-fit distribution based on Kolmogorov-Smirnov Goodness of Fit tests. The test works by comparing the theoretical probability distribution for a dataset against the actual Cumulative Distribution Function (CDF). Figure 3 contains two Kolmogorov-Smirnov Goodness of Fit diagrams for the infection level data and the Normal and Lognormal distributions. Clearly, the Lognormal Distribution better approximates the observed data, as the two curves are visibly closer together.

The KS parameter is determined by taking the maximum distance between the theoretical and empirical distributions. The results of this test are then reported as a hypothesis test where
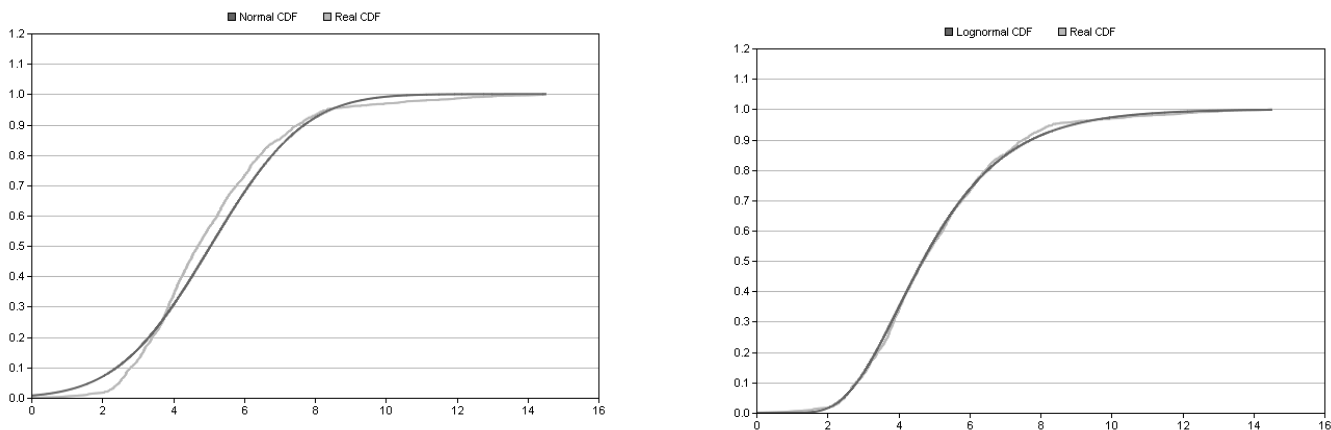

Fig. 2. Simulated infection level data

Fig. 3. Two Kolmogorov-Smirnov Goodness of Fit Tests

the null-hypothesis entails accepting the distribution, and the alternative hypothesis is rejecting it. *Alpha* is the probability of incorrectly rejecting the distribution. The current implementation of DRAM displays the results of this test for *alphas* of 20%, 10%, 5% and 1%. When choosing the best distribution automatically, DRAM selects the distribution that results in the smallest KS value when the test is applied to the time series data.

Step 4

The fundamental problem encountered when applying random number statistical theory to dynamic simulation results is that the simulation results are not (in the vast majority of cases) random variables. There is often a high level of autocorrelation between sequential data points in the resulting time series. Thus the sequential observations are not strictly independent. To compensate for this fact, the DRAM algorithm analyzes the autocorrelation within the dataset to approximate the total number of independent observations.

First, a correlogram (autocorrelation plot) for the data series is constructed. The correlogram plots the autocorrelation within the dataset for different offsets. The algorithm then locates the first place at which the autocorrelation falls below a set threshold (recommended to be 0.707, the point at which R^2 is 0.5 indicating that the majority of an observation is no longer explained by the previous observations). The offset at this point is recorded. The total number of data-points is then divided by this critical offset. The result of this division is the number of approximately independent data points. Figure 4 illustrates such a correlogram for the disease model.

The success of the method is seen in the way the number of independent observations changes as we change the simulation time step. The simulation time step is a parameter of system dynamics models which is used by numerical solvers. The smaller the time step, the more accurate the resulting simulation but the longer it takes to complete the simulation. Theoretically, reducing the time step does not change the probability that thresholds will be exceeded (assuming it is already at a relatively fine-grained level and the level on randomness in the model is not based on the time step) thus
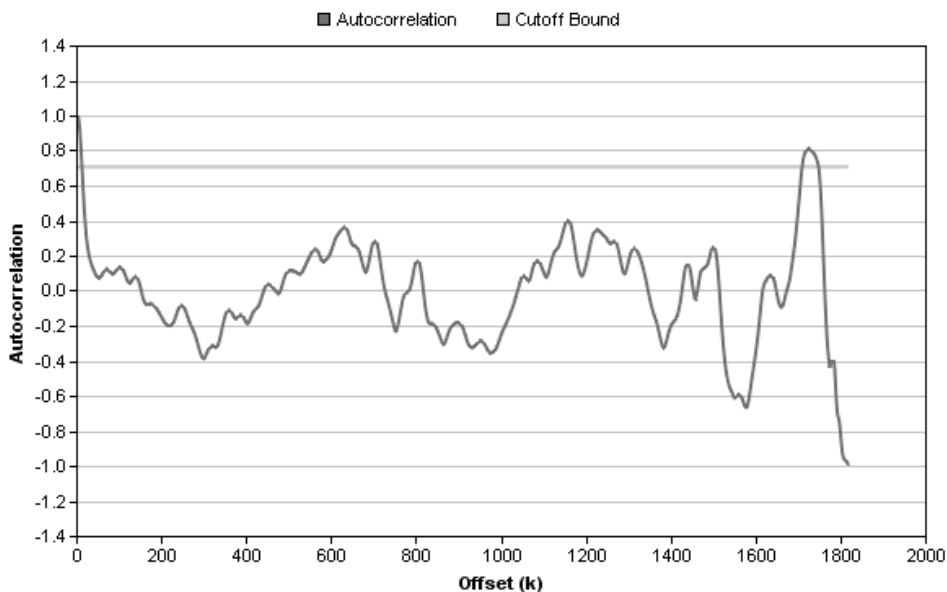

Fig. 4. Sample correlogram; the critical offset in this analysis is found to be 10

the number of independent data-points should remain constant. Cutting the time step in half (while holding constant the duration of the simulation) results in a doubling the number of observations. In our tests for the infection model, however, the number of independent observations calculated by DRAM only changed by approximately 0.5% when the number of real observations was doubled.

Step 5

For a random variable, the probability of a lower bound threshold not being violated or crossed in one trial is denoted q_l and is determined with Equation 1.

$$q_l = P(X \geq x) = 1 - CDF(x) \qquad (1)$$

For an upper bound threshold, the probability of the threshold being violated is denoted q_u and is defined in Equation 2.

$$q_u = P(X < x) = CDF(x) \qquad (2)$$

Where CDF is the Cumulative Probability Distribution with the parameters calculated from the time-series data and x is the user-specified threshold. The probability of the threshold not being violated in n trials is denoted Q and is calculated using Equation 3.

$$Q = q^n \qquad (3)$$

Where q is either q_l or q_u depending on the threshold type, n is the number of independent events in the time series. Thus the probability of a threshold being violated in n trials is denoted P and is calculated using Equation 4.

$$P = 1 - Q = 1 - q^n \qquad (4)$$

Step 6

The probability of violation is calculated separately for each threshold. Combined risk is calculated using Equation 5.

$$R = \sum_i P_i C_i \qquad (5)$$

Where P_i is the probability of the ith threshold being violated and C_i is the cost of such a violation. This total risk that is calculated is for the simulation duration. For different durations the risk may be scaled linearly.

Figure 5 illustrates example results from a DRAM analysis for the disease simulation. Hypothetical costs were assumed for testing the algorithm. Calculated risk for the simulated community of 50 000 individuals over the course of 5 years is $R = \$ 9\,123.33$. This cost may be scaled linearly for different time periods resulting in a yearly risk of: $R = \$ 1\,824.66$.
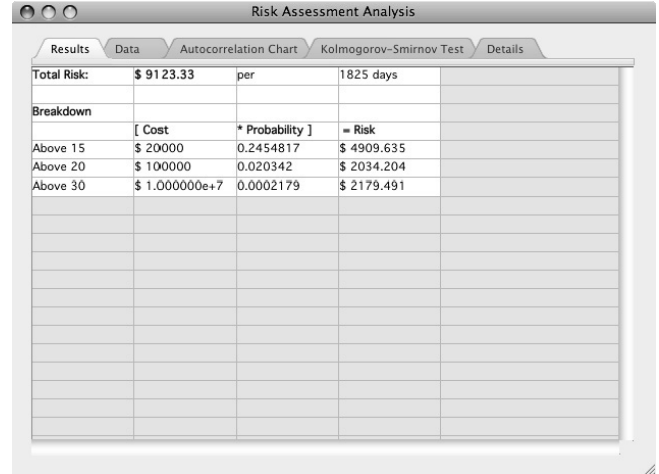

Fig. 5. Sample results from DRAM analysis

Step 7

To assess the potential risks, the value of R should be compared to the levels of acceptable risk (RA), tolerated risk (RT), and unacceptable risk (RN). The adoption of such benchmarks is one of the most difficult stages of risk analysis. For example, Rak (Rak 2009) proposed acceptable risk (RA) below 10^{-9}, tolerated risk (RT) between 10^{-9} and 10^{-6}, and unacceptable risk (RN) above 10^{-6}. The risk must be constantly monitored. In addition, if the costs of risk mitigation are not excessive, mitigation methods should be employed. The means of mitigation of the risks involved may include, for example, increasing levels of water disinfection or a change of the water source.

Conclusions

The method of estimating risk is based on the simulation of the number of people infected during a period of time and at the cost of treating the illness. The longer period of time used for the simulation, the more reliable the results obtained. The cost of treatment and loss productivities depend on the level of the population that is affected. Risk assessment requires the adoption of benchmarks such as the levels of acceptable, tolerable, and unacceptable risks. The proposed method of risk assessment may be useful for assessing risks in data series generated through system dynamics simulation techniques.

References

[1] Arquitt, S. & Johnstone, R. (2008). Use of system dynamics modelling in design of an environmental restoration banking institution, *Ecological Economics* 65, pp. 63–75.

[2] Dangerfield, B.C. & Roberts, C.A. (2000). A strategic evaluation of capacity retirements in the steel industry, *Journal of the Operational Research Society,* 51, pp. 53–60.

[3] Dangerfield, B. (1999). System dynamics applications to European health care issues, *Journal of the Operational Research Society,* 50, pp. 345–353.

[4]   Corso, P., Kramer, M., Blair, K., Addiss, D., Davis, J. & Haddix, A. (2003). Cost of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin, *Emerging Infectious Diseases,* 9, pp. 426–431.

[5]   Fortmann-Roe, S. & Wójcik, W. (2009). Simulation and management: applications to the optimum Cryptosporidium level disinfection. In: *Current issues in the water treatment and distribution,* pp. 331–340. (in Polish)

[6]   Simgua version 2.8.2. Downloaded from (http://simgua.com (11.03.2015)). Source code for this reference implementation is available from the authors upon request.

[7]   Rak, J.R. (2009). *Safe municipal water – Risk assessment in water supply system*, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 2009. (in Polish)

Metoda dynamicznego szacowania ryzyka
– propozycja oceny ryzyka w systemie zaopatrzenia wody

Dynamika Systemów jest metodologią modelowania i analizy złożonych systemów. Taki złożony system może być charakteryzowany przez wzajemne proste połączenia elementów oraz istniejące sprzężenia zwrotne. Przeprowadzenie oceny ryzyka przy modelowaniu Systemów Dynamicznych jest trudnym wyzwaniem. Chociaż w niektórych przypadkach, uzyskane za pomocą symulacji serie wyników mogą się wydawać przypadkowe, to jednak często istnieje wysoki stopień autokorelacji między tymi seriami wynikający z istnienia powiązań zwrotnych w systemie. Artykuł przedstawia propozycje Metody Dynamicznego Szacowania Ryzyka (MDSR), która pozwala oceniać ryzyko związane z hipotetycznymi kosztami zachorowań wywołanymi skażeniem systemu zaopatrzenia w wodę przez *Cryptosporidium*.