Determination of Recovery Time Distributions for Water Supply Systems for Level 2 PRA

Miika Wallius
Teollisuuden Voima Oyj, Eurajoki, Finland

Abstract: This paper presents an in-depth analysis of how recovery time distributions are determined for the water supply systems at Olkiluoto 1 and Olkiluoto 2 boiling water reactors (BWR) at the Olkiluoto nuclear power plant (NPP). The paper begins with an introduction to the importance of recovery times in the context of the Level 2 probabilistic risk assessment (PRA). The methodology for determining recovery times is detailed, including the processing of initial values from Level 1 PRA, and the use of the Latin Hypercube Sampling (LHS) method for the formation of random sample. The random sample is used for the calculation of log-normal distribution parameters, which are then used in actual Level 2 PRA model. While the log-normal distribution generally provides a good fit, it is not perfect, with some exceptions mainly due to the presence of difficult-to-recover maintenance packages, fires, and seismic events which distort the random sample to be less log-normal. The concept of dependencies and system availability derived from recovery time analysis are also discussed, since they are crucial for determining recovery times in the level 2 PRA model. However, these parameters do not influence the calculation of log-normal parameters and were not analyzed further. The paper concludes that while the log-normal distribution is a useful model, it has its limitations and suggests further research could explore other distribution models or goodness-of-fit measures and analyze dependencies and system availability derived from the recovery time analysis further.

Keywords: probabilistic risk assessment (PRA), level 2, recovery time, log-normal distribution

1. INTRODUCTION

In probabilistic risk assessment (PRA), accurately determining recovery times for plant damage states (PDSs) is essential for accurate accident progression analysis in a Level 2 PRA. This paper describes the methodology for determining the parameters of log-normal distributions that are used to estimate the recovery times of three water supply systems, specifically the containment spray (CS) system, the low-pressure reactor core spray (LPRCS) system, and the auxiliary feedwater (AFW) system. The focus is on Olkiluoto 1 and Olkiluoto 2 boiling water reactors (BWR) at the Olkiluoto nuclear power plant (NPP).

The entire procedure for computing the parameters of the log-normal distribution is depicted in figure 1. This paper will provide more comprehensive instructions in the subsequent sections. In short, the procedure starts by identifying the 100 most significant minimal cut sets (MCSs) for each PDS from a Level 1 PRA. The aim is to find fault combinations within the PDS MCS that could prevent the safety function of any of the three water supply systems. If a fault combination is found, a recovery time is determined for it. Once each MCS of the PDS has been processed, all unique recovery times are collected in a table, and their cumulative share of the total frequency of all MCS of the PDS is calculated. Using this method, we create an initial value matrix for each system in each PDS.

From the initial values we create a random sample using the Latin Hypercube Sampling (LHS) method, ensuring that the sample represents the actual variation of the data set. This sample is then used to calculate the parameters of the log-normal distribution.

The accuracy of the log-normal distribution is then evaluated by comparing the created random sample to the values produced by the log-normal distribution.

In the conclusion, the paper addresses the following two research questions:

- RQ1 How can the recovery times for different systems in different PDSs in a nuclear power plant be determined?
- RQ2 How well does the log-normal distribution model the recovery times in different PDSs?

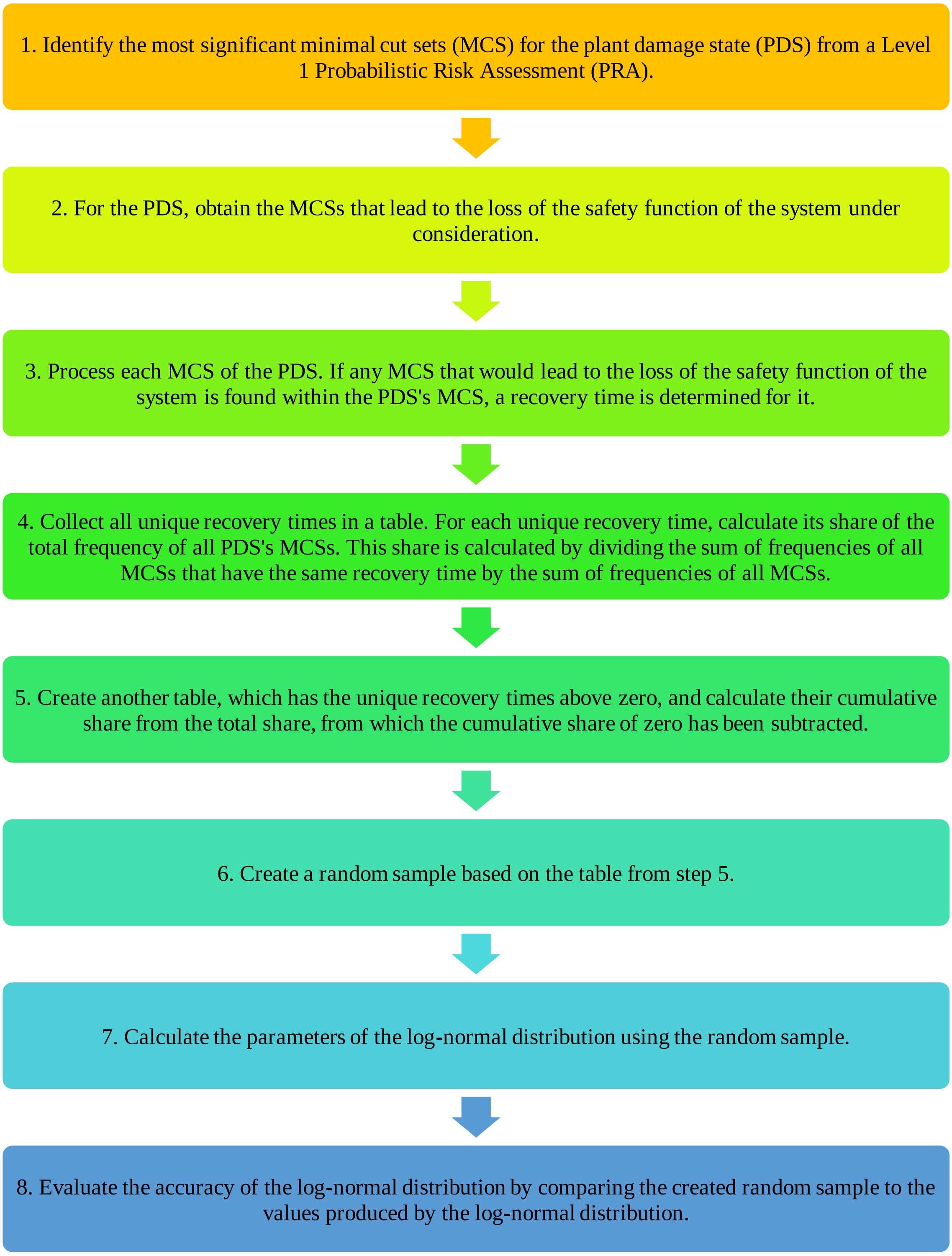

Figure 1. The entire procedure for computing the parameters of the log-normal distribution for one system in one PDS.

2. DETERMINATION OF RECOVERY TIMES

2.1. Analyzing the MCSs to determine recovery times

For each PDS, MCSs that lead to core damage are obtained from the Level 1 PRA model. The 100 most significant MCSs are considered in the analysis, as they cover the largest part of each PDS's core damage frequency, thus providing a sufficiently representative picture of the most significant combinations of faults leading to core damage or core melt. Of the 13 PDSs, nine were considered in the analysis. These are underlined in table 1.

Table 1. Abbreviations and descriptions for each PDS.

Number	Abbreviation for the PDS	Description, where the words explaining the abbreviations are **bolded**
1	CBP	A large **containment by-pass** path exists before core damage (low power operation modes only)
2	RCO	**Reactivity control** is totally lost. The insertion of control rods fails followed by unsuccessful boration.
3	ROP	Very early **reactor overpressurization** prevents core cooling
4	COP	Very early **containment overpressurization** destroys pipe works thus preventing core cooling.
5	HPL	An early core melt at **high primary pressure**, initiated by a loss-of-coolant accident (**LOCA**).
6	HPT	An early core melt at **high primary pressure**, initiated by a **transient**.
7	LPL	An early core melt at **low primary pressure**, initiated by a **LOCA**.
8	LPT	An early core melt at **low primary pressure**, initiated by a **transient**.
9	RHL	A late core melt caused by the depletion of auxiliary feedwater or **loss of residual heat removal**, initiated by a **LOCA**.
10	RHT	A late core melt caused by the **loss of residual heat removal**, initiated by a **transient**.
11	VLL	Successful very late venting without core damage or very late core damage caused by **very late leakage** or rupture of containment due to unsuccessful venting.
12	FCF	**Fuel cladding failure** terminated with boration or control rod insertion.
13	ACH	Transient initiated core melt begins, when **water reserves have been used from the high-pressure alternate coolant injection system** (ACIS) and depressurization of reactor coolant system (RCS) has been unsuccessful.

For each PDS, MCSs that lead to the loss of the safety function of the three water supply systems (CS, LPRCS and AFW) are also obtained.

The next step is to search for combinations of faults affecting the operation of the water supply systems, i.e., the MCSs of the water supply systems, within the MCSs of the PDSs. The MCSs of the PDSs are examined one by one, and if a combination of faults within the MCS of the PDS is found that disables the required number of subsystems for the execution of the safety function, this MCS of the PDS is considered in the determination of recovery times.

The recovery time is listed for the basic event of the MCS of the PDS with the shortest recovery time which recovers the system's safety function. In the case of common cause failures (CCF), the recovery time of a single basic event is multiplied by two. The rationale behind this is, that if we simply used the recovery time of a single basic event for the CCF, it might underestimate the complexity and coordination required to address multiple simultaneous failures. In contrast, multiplying the repair time by the number of basic events included in the CCF could overestimate the time needed, as some recovery activities might overlap or be streamlined. Therefore, multiplying the recovery time by two provides a more balanced estimate.

The recovery time for each basic event is obtained from the input data of the PRA model. If the recovery time has not been defined for some basic event, an appropriate recovery time is estimated for it. For example, in the case of a severe seismic event, a long recovery time of 48 hours is used, as it is assumed that the systems most likely cannot be recovered during such event. However, if the MCS of the PDS does not affect the execution of the safety functions of the systems, a recovery time of zero is listed for the last basic event of its MCS. An example of this procedure is shown in figure 2.

MCS number	Frequency	Name	Description	Recovery time [h]
1	3.10E-08	INT-S2	Small coolant leak	
		322P001D1AZ-BCD	3x CCF pump does not start	8
		322P001..H	Maintenance package	
2	1.60E-08	INT-S1	Medium coolant leak	
		213V225B2AZ-ABCD	4x CCF valve does not open	0

Figure 2. Example: Identifying recovery times for the CS system in an example PDS's MCSs. The fault combination outlined in yellow causes CS system's safety function to fail, and one outlined in black does not.

In figure 2, within the MCS of the PDS marked in blue, there is a MCS that causes CS's safety function to fail (outlined in yellow). To restore the CS system's safety function, one train out of four must be restored. In this example case, it has been estimated that the CCF of three pumps is faster to restore than the maintenance package, so the recovery time (8 h) is listed for the CCF. Inside the MCS marked in red, there is no MCS that causes the CS system's safety functions to fail, so zero is recorded for the last event (outlined in black). In our analysis, this process is done using an automated script.

Once each MCS of the PDS has been processed, all unique recovery times are collected in a table, and their cumulative share of the total frequency of all MCSs of the PDS is calculated. The cumulative share represents the proportion of the total frequency corresponding to each recovery time. This allows us to create an initial value matrix, show in table 2.

Table 2. Example: Initial value matrix of recovery values for CS system in LPL PDS.

Recovery time [h]	Cumulative share
0	0.4261
0.17	0.4263
0.5	0.5734
2	0.7124
6	0.8223
8	0.8746
10	0.8804
18	0.8822
20	1

2.2. Other parameters derived from the recovery time analysis

By applying the analysis shown in previous chapter, dependencies can also be calculated. Dependencies mean what is the probability that two systems fail due to the same fault. When this happens, both systems will have the same recovery time. The dependency between two systems is the sum of the frequencies of the PDS MCSs affecting two systems divided by the sum of the frequencies of all PDS MCSs. For example, in the case of LPL, there is a 97% dependency between the CS and LPRCS systems, a 99% dependency between the LPRCS and AFW systems, and a 97% dependency between the CS and AFW systems. Thus, in the LPL PDS, if any of these systems fail due to some fault combination, the other systems are also most likely out of order.

The probability of availability for each system in each PDS can also be calculated based on initial values. It is simply the cumulative proportion of zero recovery time. For example, the availability for the CS system, visible in table 2, is approximately 43% in the LPL PDS.

Dependencies and availabilities need to be considered in the level 2 PRA model when determining recovery times for the systems, but they do not affect or are considered in the calculation of the log-normal parameters.

3. PARAMETERS OF THE LOG-NORMAL DISTRIBUTION

3.1. Processing of initial values

Since a recovery time of zero is not considered from the values of table 2, we calculate what is the cumulative share of other recovery times from the total share, from which the cumulative share of zero has been subtracted. This value *C* can be calculated using equation (1) as follows:

$$C=\frac{c\text{-}c_0}{1\text{-}c_0} \tag{1}$$

where c is the cumulative share of the value to be calculated from the set of recovery times, which also includes zero, and
c_0 is the cumulative share of zero from this set.

Using equation (1) for all recovery times above zero in table 2, we can form table 3.

Table 3. Example: The cumulative share of the LPL PDS's recovery times as a proportion of the total share, after subtracting the cumulative share that equals zero.

Recovery time [h]	Cumulative share
0.17	0.0004
0.5	0.2566
2	0.4988
6	0.6904
8	0.7815
10	0.7916
18	0.7948
20	1

3.2. Creation of a random sample

Using the values calculated by equation (1) from table 3, a random sample is formed using the LHS method, which ensures that the random sample represents the actual variation of the data set. However, in this case, the random sample is not shuffled as it normally would be, since shuffling does not affect the calculation of log-normal parameters.

The random number p is calculated using the LHS method with equation (2) as follows [1]:

$$p=\frac{i\text{-}1+Rnd}{n} \tag{2}$$

where i is the index of the random sample and it ranges from 1 to n,
n is the number of random samples, and
Rnd is a random number in the range [0, 1].

Using equation (2), a random number p is drawn on the first round from the range [0, 1/n], on the second round from the range [1/n, 2/n], …, and on the last round from the range [1-1/n, 1]. In comparison, a random number p is drawn on each round from the range [0, 1] in the more common uniform random sampling method.

When the random number p is compared to the cumulative proportions in table 3 in the manner indicated in table 4, it can be determined what return time has to be used in the calculation of the inverse cumulative distribution function (ICDF), also known as the quantile function, for the exponential distribution. The exponential distribution is used here because the recovery times are not assumed to be constant, as in the initial values, but they vary due to various reasons, such as the type of components affected, the cause of the fault, and the repair method, etc. The value x [h], i.e. the recovery time given by the exponential distribution's quantile function with probability Rnd, is calculated as follows [2]:

$$x=\frac{ln\,(1-Rnd)}{1/\beta} \quad (3)$$

where *Rnd* is a random number in the range [0, 1], and
β is the recovery time [h].

Table 4. Example: Choosing β based on p and example values from table 3.

Comparing p to the cumulative share	β [h]
$0 < p \leq 0.0004$	0.17
$0.0004 < p \leq 0.2566$	0.5
$0.2566 < p \leq 0.4988$	2
$0.4988 < p \leq 0.6904$	6
$0.6904 < p \leq 0.7815$	8
$0.7815 < p \leq 0.7916$	10
$0.7916 < p \leq 0.7948$	18
$0.7948 < p \leq 1$	20

3.3. Calculation of log-normal distribution parameters

Next, the parameters of the log-normal distribution, i.e. the mean μ, variance σ^2, error factor *EF*, expected value *E*, and standard deviation σ, are to be calculated from the random sample formed in the previous chapter. They can be calculated using equations (4), (5), (6), (7), and (8) as follows [3, 4]:

$$\mu=\frac{\sum_{i=1}^{n} ln(x_i)}{n} \quad (4)$$

$$\sigma^2=\frac{\sum_{i=1}^{n}(ln(x_i)\,-\mu)^2}{n-1} \quad (5)$$

$$\sigma=\sqrt{\sigma^2} \quad (6)$$

$$E=e^{\mu+\frac{\sigma^2}{2}} \quad (7)$$

$$EF=e^{1.64485\cdot\sigma} \quad (8)$$

where i is the index of the random sample,
x_i is the recovery time of the i-th random sample, and
n is the number of random samples.

In table 5 are the example results for CS system in LPL PDS. These results are calculated from the random sample formed using the values from table 4.

Table 5. Example: Parameters calculated from a random sample (n = 100000) formed using the values from table 4.

Parameter	Value
Standard deviation σ	1.86
Mean μ	0.59
Expected value *E* [h]	10.11
Error factor *EF*	21.22

4. EVALUATION OF THE RESULTS

This section presents how the accuracy of the log-normal distribution can be evaluated. This is done by comparing the created random sample to the values produced by the log-normal distribution.

First, the values of the random sample are statistically divided into 5-minute (approximately 0.08 h) sections for 48 hours. Then we calculate how many data points are in the range (0, 0.08], how many data points are in the range (0.08, 0.17], etc. In this way, we can estimate the statistical probability of each interval. For example, if there are a total of 100000 iteration rounds and there are 5476 values in the range (0, 0.08], the statistical probability *s* is calculated as follows:

$$s(0.08\ h)=\frac{5476}{100000}=5.48E\text{-}2$$

The statistical value is then compared to the numerical integral (10) of the log-normal density function (9). [3]

$$lognpdf(x)=\frac{e^{\frac{-1}{2}\cdot\left(\frac{ln(x)-\mu}{\sigma}\right)^2}}{x\cdot\sqrt{2\cdot\pi}\cdot\sigma} \tag{9}$$

where x is the recovery time [h],
μ is the mean, and
σ is the standard deviation.

$$lognp(x_i)\approx(x_i\text{-}x_{i\text{-}1})\cdot lognpdf(x_i) \tag{10}$$

Using equation (10), a numerical integral can be calculated for the desired interval. This calculation is repeated for all intervals up to 48 hours, which covers most of the random sample. As an example, the values for the first hour would look like the ones in table 6. In it, *lognp(x)* is calculated using the values from table 5.

Table 6. Example: Statistical probability *s(x)* and probability *lognp(x)*, which is calculated through the numerical integral of the density function of the log-normal distribution.

x [h]	*lognp*(*x*)	*s*(*x*)
0.00	0.00E+00	0.00E+00
0.08	5.46E-02	5.48E-02
0.17	4.72E-02	4.74E-02
0.25	4.07E-02	4.15E-02
0.33	3.55E-02	3.67E-02
0.42	3.15E-02	3.26E-02
0.50	2.82E-02	2.95E-02
0.58	2.55E-02	2.64E-02
0.67	2.33E-02	2.38E-02
0.75	2.13E-02	2.17E-02
0.83	1.97E-02	1.98E-02
0.92	1.83E-02	1.79E-02
1.00	1.70E-02	1.69E-02

From table 6, it can be seen that the statistical probability *s(x)* and probability *lognp(x)* are almost the same. However, for a more precise estimate, the R-squared (R^2) value **Error! Reference source not found.** can be calculated. It represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Its value can vary for non-linear regressions in the range [-∞, 1] and for a linear regression in the range [0, 1].

$$R^2=1-\frac{\sum_{i=1}^{n}\left(s(x_i)-lognp(x_i)\right)^2}{\sum_{i=1}^{n}(s(x_i)-\bar{s})^2}$$

where $\bar{s}$ is the mean of the statistically recorded probabilities.

Using equation **Error! Reference source not found.** for the example values in table 6, the R^2 value is 0.94, which shows that the recovery times of LPL PDS follow the log-normal distribution excellently. For the entire 48-hour set, the R^2 value was 0.99. When this is visualized in figure 3, the similarity is even more apparent. Only the first 24 hours are plotted in the figure so that the beginning of the curve is more visible.

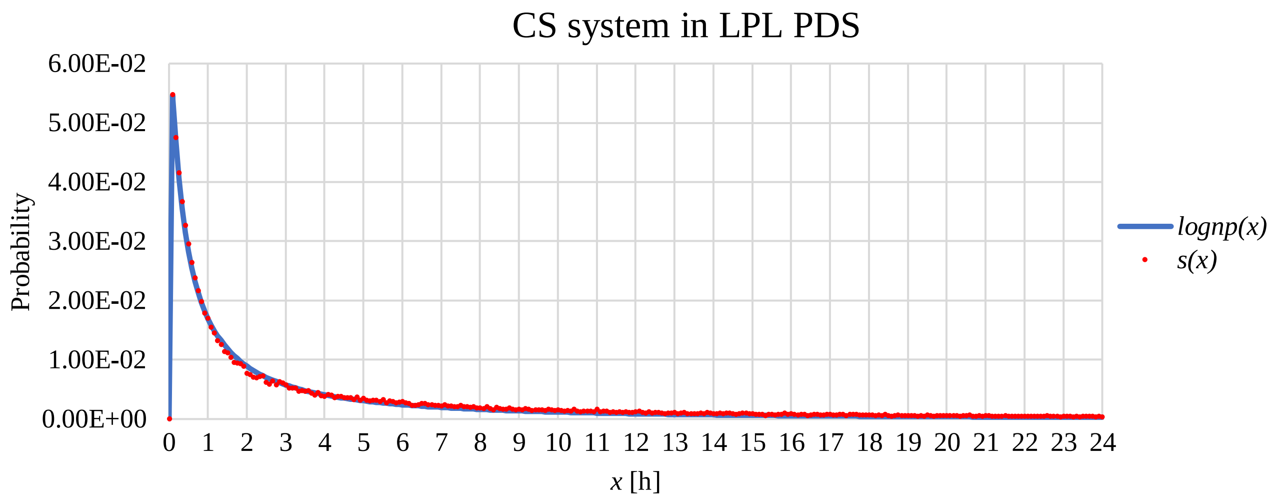

Figure 3. Example plot for the CS system in the LPL PDS.

We can also see that the data does not always follow log-normal distribution perfectly, as evidenced by the suboptimal R^2 value of 0.74 for the CS system in the RHT PDS, shown in figure 4. The difference is mainly due to what kind of MCSs there are in the RHT PDS and what is their proportion of the PDS's total frequency. The RHT's MCSs contain many difficult-to-recover maintenance packages and fires, which have a long recovery time. The relatively large proportion of these long recovery times distort the random sample to be less log-normal.

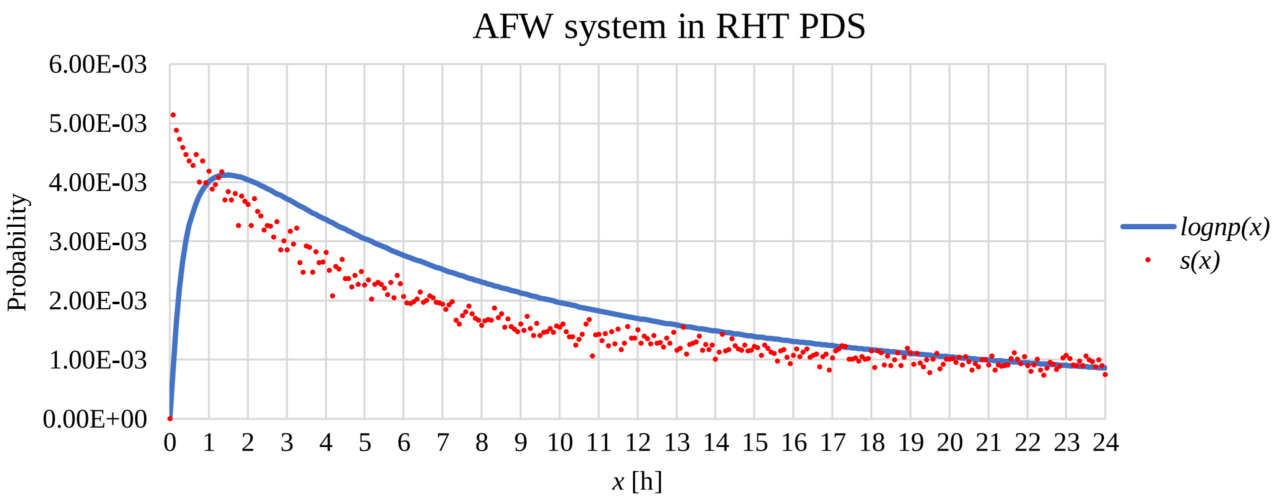

Figure 4. Example plot for the AFW system in the RHT PDS.

Even worse R^2 value is for the LPRCS system in the RCO PDS, shown in figure 4. It is -0.07, which indicates that the data does not follow log-normal distribution at all. This is because in the RCO plant failure state, there are many MCSs that include seismic events. For these, a recovery time of 48 hours is used, and since there are hardly any other recovery times observed, the random sample distorts to completely non-log-normal. The same phenomenon was observed with the ROP PDS, but in the context of fire events.

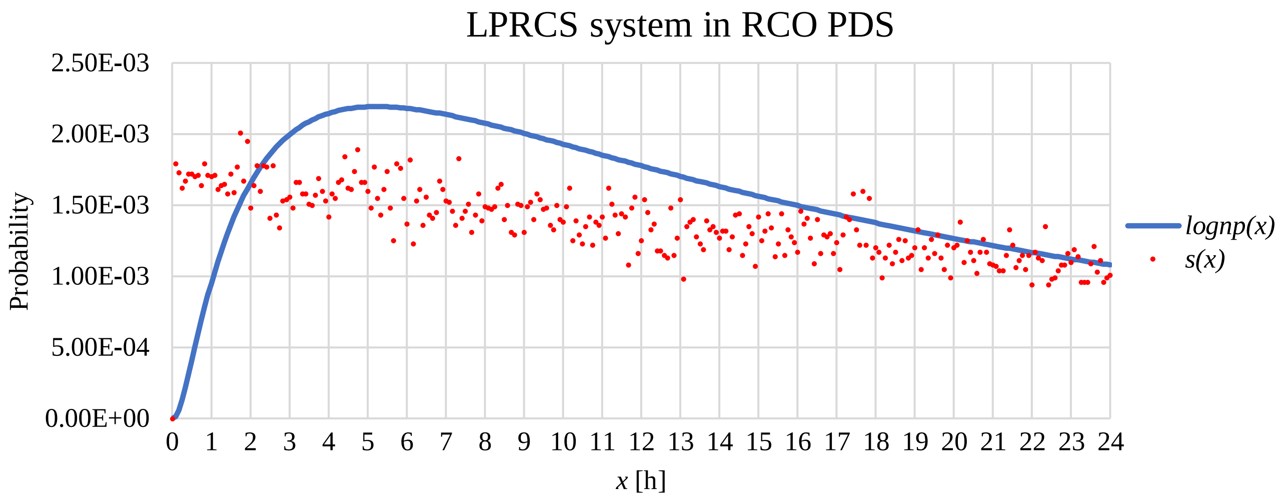

Figure 5. Example plot for the LPRCS system in the RCO PDS.

Three water supply systems were analyzed across nine different PDSs, resulting in a total of 27 figures. To conserve space, only three of these figures are included in this paper. Regarding the fits,

- 16 of them had an R^2 value above 0.9 (indicating a good fit),
- four had an R^2 value less than 0.9 but greater than 0 (indicating a poor fit),
- four had an R^2 value less than 0 (indicating that the data was completely non-log-normal), and
- three had an R^2 value of NaN (indicating that the system is always available in the PDS).

The R^2 values vary due to the types of MCSs present in each PDS and their respective recovery times. MCSs with long recovery times, such as those involving difficult-to-recover maintenance packages, fires, seismic events and other hazards, tend to distort the distribution, leading to lower or even negative R^2 values.

The drawback of the R^2 value is that it can be misleading with non-linear models. Even though we can calculate R^2 for non-linear regression, it doesn't provide a clear interpretation in terms of the proportion of variance explained by the predictors. This can lead to an overestimation of the goodness of fit indicated by the R^2 value. Despite this drawback, the R^2 value, when used together with visual inspection, provides a quick and easy-to-understand measure of the overall fit of the model, and that is why it is used in this evaluation.

It should be noted that when dealing with log-normal distributions, it's often appropriate to use some goodness-of-fit measures other than R^2 value, that might be more suitable for this type of distribution, such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), as they take into account both the goodness of fit of the model and the complexity of the model. However, applying these methods would require testing multiple different models, which is outside the scope of this paper. Instead, this paper focuses on providing a comprehensive understanding of the basic statistical properties and applications of the log-normal distribution.

5. CONCLUSION

In response to RQ1, this paper has shown that recovery times for different systems in different PDS in a NPP can be determined through a detailed analysis of MCSs obtained from the Level 1 PRA model. The process involves identifying combinations of faults affecting the operation of the water supply systems within the MCSs of the PDSs. Once each MCS of the PDS has been processed, all unique recovery times are collected and their cumulative share of the frequency of all MCSs of the PDS is calculated.

As for RQ2, this paper has demonstrated that the log-normal distribution models the recovery times in different PDS quite well, but not perfectly. The goodness of fit of the log-normal distribution was evaluated by comparing the created random sample to the values produced by the log-normal distribution. The R-squared (R^2) value was used together with visual inspection as a measure of the overall fit of the model. For most of the data, the R^2 value was close to 1, indicating an excellent fit. However, there were some exceptions, where the R^2 values were suboptimal or even negative, indicating a poor fit. The difference between fits is due to what kind of MCSs there are in the PDS and what is their proportion of the PDS's total frequency. MCSs that contain events with long recovery times, such as difficult-to-recover maintenance packages, fires, and seismic events, distort the random sample to be less log-normal.

Chapter 2.2 discusses the concept of dependencies and system availability derived from recovery time analysis. Dependency refers to the likelihood of two systems failing due to the same fault. System availability, on the other hand, is calculated as the cumulative proportion of zero recovery time for each system in each PDS. These parameters are crucial when determining recovery times in the level 2 PRA model. However, they do not influence the calculation of log-normal parameters, so they were not analyzed further in this paper.

In conclusion, while the log-normal distribution provides a useful model for recovery times in different PDS, it is not without its limitations. It is important to consider the specific characteristics of each PDS and the types of faults that can occur when applying this model. Further research could explore other goodness-of-fit measures, such as AIC or BIC, to find distribution models that might provide a better fit for the recovery time data in the cases where the data does not follow the log-normal distribution. In addition, the above-mentioned dependencies and system availability could be analyzed further.

Acknowledgements

I appreciate TVO and especially my supervisor Jenna Levo for allowing me the time during work hours to complete this paper. Their understanding and flexibility have been very helpful in this work.

References

[1] Cheng, J., & Druzdzel, M. (2000). Latin Hypercube Sampling in Bayesian Networks. Proceedings of FLAIRS-2000. 287-292.
[2] Joram, S. (2022). Proof: Quantile function of the exponential distribution. The Book of Statistical Proofs. Available at: https://statproofbook.github.io/P/exp-qf.html (Accessed: 31 May 2024). DOI: 10.5281/zenodo.4305949.
[3] Evans, M., Hastings, N., & Peacock, B. (1993). Statistical Distributions (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.
[4] Britton, P., Mohammad, A. H., & Robert, R. (2023). Uncertainty Estimation Cheat Sheet for Probabilistic Risk Assessment. Available at: https://ntrs.nasa.gov/api/citations/20170012415/downloads/20170012415.pdf (Accessed: 17 October 2023).