

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

Miika Wallius

Teollisuuden Voima Oyj
Olkiluoto, Eurajoki, Finland
miika.wallius@tvo.fi

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 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.

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).

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 [1], ensuring that the sample represents the

actual variation of the data set. Each sample is calculated using the inverse cumulative distribution function (ICDF) for the exponential distribution [2]. This sample is then used to calculate the parameters of the log-normal distribution, as demonstrated in [3] and [4].

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.

The reason for using this method is because Level 1 PRA and Level 2 PRA are not directly connected via any interface in the Olkiluoto 1 and 2 PRA model.

This paper is a summary of the full paper published in 7th International Conference on Probabilistic Safety Assessment and Management & Asian Symposium on Risk Assessment and Management (PSAM17&ASRAM2024) [5]. In it 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 thus were not included in this summary. The full paper also includes the whole calculation process in more detail with all the necessary equations and examples.

2 EVALUATION OF THE RESULTS

Accuracy of the log-normal distribution can be evaluated by comparing the created random sample to the values produced by the log-normal distribution. The R-squared (R^2) is used for this purpose. 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 $[-\infty, 1]$ and for a linear regression in the range $[0, 1]$.

LPL (an early core melt at low primary pressure, initiated by a LOCA) PDS follows the log-normal distribution excellently, as the R^2 value was 0.99. When this is visualized in figure 1, 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. In each figure $s(x)$ represents the number of random samples in each 5-minute section, and $\log np(x)$ represents the numerical integral of the log-normal density function.

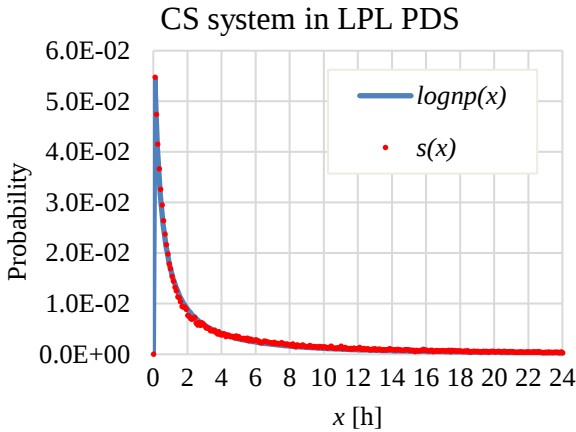


Figure 1: Example plot for the CS system in the LPL PDS.

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 (reactivity control is totally lost) PDS, shown in figure 2. The difference is mainly due to what kind of MCSs there are in the RHT (a late core melt caused by the loss of residual heat removal, initiated by a transient) PDS and what is their proportion of the PDS's total frequency. 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.

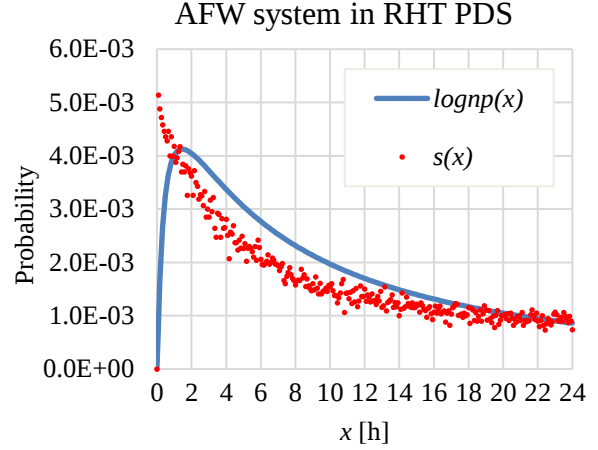


Figure 2: Example plot for the AFW system in the RHT PDS.

Even worse R^2 value is for the LPRCS system in the RCO (reactivity control is totally lost) PDS, shown in figure 2. 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 (very early reactor overpressurization prevents core cooling) PDS, but in the context of fire events.

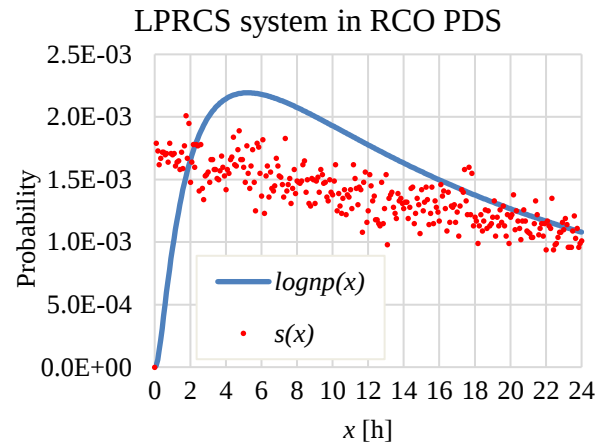


Figure 3: Example plot for the LPRCS system in the RCO PDS.

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.

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.

3 CONCLUSION

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.

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 cases where the data does not follow the log-normal distribution.

REFERENCES

- [1] 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).
- [5] Wallius, M. (2024) Determination of Recovery Time Distributions for Water Supply Systems for Level 2 PRA. Available at: <https://iapsam.org/PSAM17/program/Papers/PSAM17&ASRAM2024-1405.pdf>. (Accessed: 19 May 2025).