2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Imprecise survival signature approximation using interval predictor models

1st Jasper Behrensdorf
Institute for Risk and Reliability
Leibniz University Hannover
Hannover, Germany
0000-0001-9628-7250

2nd Matteo Broggi
Institute for Risk and Reliability
Leibniz University Hannover
Hannover, Germany
0000-0002-3683-3907

3rd Michael Beer
Institute for Risk and Reliability
Leibniz University Hannover
Hannover, Germany
0000-0002-0611-0345

***Abstract*—This paper presents a novel technique for the approximation of the survival signature for very large systems. In recent years, the survival signature has seen promising applications for the reliability analysis of critical infrastructures. It outperforms traditional techniques by allowing for complex modelling of dependencies, common causes of failures and imprecision. However, as an inherently combinatorial method, the survival signature suffers greatly from the curse of dimensionality. Computation for very large systems, as needed for critical infrastructures, is mostly infeasible. New advancements have applied Monte Carlo simulation to approximate the signature instead of performing a full evaluation. This allows for significantly larger systems to be considered. Unfortunately, these approaches will also quickly reach their limits with growing network size and complexity. In this work, instead of approximating the full survival signature, we will strategically select key values of the signature to accurately approximate it using a surrogate radial basis function network. This surrogate model is then extended to an interval predictor model (IPM) to account for the uncertainty in the prediction of the remaining unknown values. In contrast to standard models, IPMs return an interval bounding the survival signature entry. The resulting imprecise survival signature is then fed into the reliability analysis, yielding upper and lower bounds on the reliability of the system. This new method provides a significant reduction in numerical effort enabling the analysis of larger systems where the required computational demand was previously prohibitive.**

***Index Terms*—survival signature, interval predictor models, radial basis function networks**

I. Introduction

The reliability analysis of critical infrastructure systems, such as electrical, gas and water distribution systems, traffic networks and communication networks is of paramount importance to the safety and development of modern societies. Our increasing dependence on the availability of these systems only escalates this fact. The analysis and assessment of reliability comes with increasingly higher computational demand due to the growing size and complexity of the infrastructure systems. In addition, phenomena such as dependencies inside or between these networks can have adverse effects and must not be disregarded [1]. Unfortunately, this is where traditional approaches such as fault tree analysis or reliability block diagrams quickly reach their limits. A modern development in system analysis built to circumvent these drawbacks is the survival signature [2].

The survival signature was introduced as a generalization of the system signature [3] to systems with multiple component types. It excels in particular where diverse effects such as dependencies, common causes of failure or imprecision need to be included in the reliability analysis. Through the separation of the system structure from probabilistic information about component failures it provides a flexible method where a variety of scenarios can be studied without unnecessary reevaluation of the system structure.

However, like the traditional methodologies, the survival signature suffers greatly from the curse of dimensionality. As a result, the numerical demand to compute the survival signature increases with non-polynomial growth with increasing network size and number of component types. As the number of components in critical infrastructure systems can quickly reach upwards of hundreds or thousands this computational demand to obtain the survival signature becomes prohibitive.

A number of promising solutions to this problem have been proposed over the recent years. These involve for example binary decision diagrams (BDD) [4], extended universal generating functions (UFG) [5] or Monte Carlo simulation [6], [7]. However, obtaining alternative system representations such as the BDD or UFG are non-trivial problems themselves, while standard Monte Carlo simulation will also quickly reach the limits of its feasibility for larger systems.

In this paper we propose a new method based on building an accurate surrogate model for the survival signature of large systems. In a first step, an adaptive strategy is employed to select which values of the survival signature to compute as data points for the surrogate model. These values are computed using the existing Monte Carlo method [6]. Then, a normalized radial basis function network is constructed from these data points to approximate the remaining values of the survival signature. Finally, the uncertainty resulting from the Monte Carlo simulation is used to extend the surrogate model to an interval predictor model (IPM). The uncertainty is propagated through the IPM which ultimately yields bounds on the survival signature.

The remainder of the paper is structured as follows. Section II presents the survival signature while Section III discusses the radial basis function based surrogate models. The newly proposed methodology is introduced in Section IV

followed by an application of the technique to a numerical example in Section V. The paper closes with some concluding remarks in Section VI.

II. Survival Signature

Consider a system with K different component types and m_k components of each type $k \in \{1, 2, \ldots, K\}$. The survival signature condenses the structural information of the system into a probability that the system is working for l_k out of m_k components working per type as

$$\Phi(l_1, \ldots, l_K) = \left[\prod_{k=1}^{K} \binom{m_k}{l_k}^{-1}\right] \times \sum_{\mathbf{x} \in S_{l_1,\ldots,l_K}} \varphi(\mathbf{x}), \quad (1)$$

where $S_{l_1,\ldots,l_K}$ denotes the set of all state vectors of the system with $l_1 \ldots, l_K$ working components and $\varphi(\mathbf{x})$ is the structure function which evaluates to 1 if the system is working for a given state vector and 0 if it is not.

A state vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the representation of a distinct system state where $x_i = 0$ indicates a component is failed and $x_i = 1$ indicates a working component for all $i \in \{1, 2, \ldots, n\}$ components of the system and $n = \sum_{k=1}^{K} m_k$.

The survival signature is defined for coherent systems, for which it is monotonically nondecreasing. The monotone behavior can be expressed as

$$\begin{aligned} \Phi(l_1^a, \ldots, l_K^a) \leq \Phi(l_1^b, \ldots, l_K^b), \text{ if} \\ l_k^a \leq l_k^b, \ \forall k \in \{1, \ldots, K\}, \end{aligned} \quad (2)$$

where the superscripts a and b refer to any two inputs of the survival signature. This monotonicity property is exploited later to design a monotone radial basis function network as a surrogate model.

The survival function, that is the reliability of the system at time t, is defined as

$$P(T_s > t) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \ P\left(\bigcap_{k=1}^{K} \left\{C_t^k = l_k\right\}\right), \quad (3)$$

where C_t^k denotes the number of components of type k functioning at time t. For known failure time distributions with cumulative distribution functions (CDF) $F_k(t)$ the probabilistic part of the survival function can be analytically computed as

$$P\left(\bigcap_{k=1}^{K} \left\{C_t^k = l_k\right\}\right) = \prod_{k=1}^{K} \left(\binom{m_k}{l_k} [F_k(t)]^{m_k - l_k} [1 - F_k(t)]^{l_k}\right). \quad (4)$$

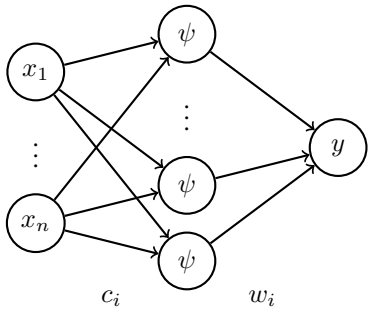

Fig. 1. Graphical representation of a radial basis function neural network. From left to right: input layer, hidden layer, linear output layer.

III. Monotone Radial Basis Function Networks

Radial basis function (RBF) networks [8] are a very simple type of artificial neural network with only one hidden layer, where the activation functions are radial basis functions, and a linear output layer. Radial basis functions in $\mathbb{R}^d$ are defined with respect to a distance function between the input and a fixed point called a center. This distance is usually the Euclidean distance defined as

$$||x - c|| = \sqrt{\sum_{j=1}^{d} \frac{(x_j - c_j)^2}{2\sigma_j^2}}, \quad (5)$$

where $x \in \mathbb{R}^d$ is the input vector and the σ_j are normalization constants controlling the spread of the radial basis function in each dimension. The most common radial basis function is the Gaussian function

$$\psi(||x - c||) = e^{-||x-c||^2}. \quad (6)$$

Using the basis functions we can define the RBF network as

$$y(x) = \sum_{i=1}^{N} w_i \psi(||x - c_i||), \quad (7)$$

where N is the number of neurons, c_i is the center point associated with neuron i and w_i is the weight of that neuron in the linear output layer. Refer to Fig. 1 for a simple graphical representation of an RBF network.

The output of the radial basis function network can be normalized as

$$y(x) = \frac{\sum_{i=1}^{N} w_i \psi(||x - c_i||)}{\sum_{i=1}^{N} \psi(||x - c_i||)}. \quad (8)$$

Normalized radial basis function (NRBF) networks have shown to improve extrapolation of the network when compared to regular RBF networks [9]. For the remainder of this paper, the basis functions used will be normalized as

$$\hat{\psi}_i(x) = \frac{\psi(||x - c_i||)}{\sum_{\hat{i}=1}^{N} \psi(||x - c_i||)}. \quad (9)$$

In this work, we restrict the center locations of the NRBF network to a regular uniform grid over the entire input domain. This allows for easier selection of the σ_j shape parameters as

well as greatly simplifying the monotonicity constraints. In the next step, the NRBF network will be used as a surrogate model of the survival signature. Enforcing monotonicity of the surrogate model greatly improves its performance since the survival signature itself is a monotone function.

Monotonicity of an NRBF network can be enforced through simple linear constraints on the weights. For centers located on a grid, only sequences of weights lying on a line in the direction of the d dimensions need to be monotonic for the NRBF network to be monotonic [10].

Assuming M data points x_j with associated function values y_j for $j \in \{1, \ldots, M\}$, the optimal weights w_i are determined as the solution to the following constrained least squares optimization problem

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & ||\Psi w - y||_2^2 \\ \text{subject to} \quad & w_p \leq w_q,\ p, q \in \{1, \ldots, N\}, \\ & \text{if } c_p^\alpha < c_q^\alpha, c_p^\beta = c_q^\beta, \alpha \neq \beta, \end{aligned} \tag{10}$$

where the superscripts $\alpha, \beta \in \{1, \ldots, d\}$ refer to the coordinates in dimensions α and β with $w = [w_1, \ldots, w_N]^T$ and $y = [y_1, \ldots, y_M]^T$. The matrix $\Psi \in \mathbb{R}^{M \times N}$ is defined as

$$\Psi_{ji} = \hat{\psi}_i(x_j). \tag{11}$$

This is a convex optimization problem which can be solved by standard solvers. Here we use the splitting cone solver (SCS) [11] in connection with JuMP [12].

IV. Approximation of the Survival Signature

This section presents the proposed methodology to the approximation of the survival signature for very large systems. Equation (1) clearly shows that the computation of the survival signature is a combinatorial problem which greatly increases in numerical demand with increasing number of components and component types. Therefore, analytical computation of the survival signature is infeasible for very large systems. In [6] we presented an approach to approximating the survival signature using percolation to exclude negligible parts of the signature and approximating the remaining entries with Monte Carlo simulation. While this enabled the analysis of larger systems, this method still requires a significant number of system evaluations and will quickly reach prohibitive computational demand. In this paper, we propose an alternative solution where the survival signature is approximated through a radial basis function network based on a few strategically selected values. In addition, the uncertainty in the survival signature entries resulting from the Monte Carlo approximation is propagated through an interval predictor model and reflected as bounds on the survival signature.

While other types of neural networks, such as feed forward networks, can be used to approximate the survival signature [7], RBF networks are explicitly chosen for their desirable properties regarding monotonicity constraints as well as the ability to be extended to interval predictor models.

A. Design of the NRBF network surrogate model

Similarly to the previously developed technique, this new approach starts by applying percolation theory to find the critical threshold f_c and exclude a significant portion of the survival signature where the probability that the system functions is negligible, i.e. close to zero [6].

Next, let Ω be the set of remaining survival signature entries $l_1, \ldots, l_K$ for $l_k \in \{0, 1, \ldots, m_k\}$ and $k \in \{1, \ldots, K\}$ satisfying the condition

$$\sum_{k=1}^{K} l_k \geq (1 - f_c) \sum_{k=1}^{K} m_k. \tag{12}$$

We start by selecting M initial entries from Ω and calculate their survival signature values using Monte Carlo simulation. Note that values that require less than the number of Monte Carlo samples N_{MC} structure function evaluations are computed exactly. In order to have good coverage of the input space, we generate the M data points by creating a uniform grid between the lower and upper bounds of Ω and then select the nearest neighbors in Ω as the starting values. In addition, we create a uniform grid of centers c_i for $i = 1, \ldots, N$ spread over the entire domain of Ω. The spread of the radial basis function σ_j is chosen as half of the distance between two successive centers in the jth dimension. These have shown to be a good starting point for NRBF networks [10]. Then, the weights for the initial monotone NRBF approximation of the survival signature are obtained by solving the optimization problem (10).

In the next step, new values of the survival signature to be approximated are selected and used to improve the surrogate model. This is performed by means of a Taylor expansion-based adaptive design (TEAD) [13]. The TEAD is a combination of exploration and exploitation based on nearest neighbor distance and Taylor expansion. This allows to both adequately cover the input space as needed as well as concentrate samples where the highest changes in the survival signature are located to provide an accurate surrogate model using a small sample size. At each point a new sample is selected from Ω and added to the M data points. Then the weights are updated by again solving the optimization problem (10). The search for new points is aborted once the change in weights falls below a defined threshold twice in a row.

B. Extension to an interval predictor model

Once an adequate surrogate model for the survival signature is obtained, it is extended to an interval predictor model. In difference to regular models, where a single output value is returned for any inputs, interval predictor models return an interval where the value is predicted to fall [14]. The goal here is to propagate the uncertainty of using Monte Carlo simulation to approximate individual entries of the survival signature through the surrogate model. By simply fitting a NRBF network to some estimated values of the survival signature the uncertainty of the Monte Carlo simulation represented

by the coefficient of variation of the particular entries would be lost.

The IPM approach is simple. Instead of fitting a single surrogate model to the available data points we fit two models that act as upper and lower bounds. These models are built based on the points identified in the previous step. Both models use the exact same center locations c_i and their difference in output is only influenced by two different sets of weights $w_{\min}$ and $w_{\max}$. The bounds are defined as

$$\overline{y}(x, w_{\max}) = \Phi(x)^T w_{\max} \tag{13}$$

$$\underline{y}(x, w_{\min}) = \Phi(x)^T w_{\min}, \tag{14}$$

where $\Psi(x) = [\hat{\psi}_1(x), \ldots, \hat{\psi}_N(x)]$.

The spread of the IPM, i.e. the separation between its limits, is defined as

$$\delta_y(x, w_{\max}, w_{\min}) = \Phi(x)^T (w_{\max} - w_{\min}). \tag{15}$$

The parameters $w_{\max}$ and $w_{\min}$ are given by the following optimization problem

$$\begin{aligned} \underset{w_{\max}, w_{\min}}{\text{minimize}} \quad & \frac{1}{N} \sum_{j}^{N} \delta_y(x_j, w_{\max}, w_{\min}) \\ \text{subject to} \quad & \underline{y}(x_j, w_{\min}) \leq \underline{y}_j \\ & \overline{y}(x_j, w_{\max}) \geq \overline{y}_j \\ & w_{\min} \leq w_{\max}, \end{aligned} \tag{16}$$

where $(x_j, \underline{y}_j, \overline{y}_j)$ for $j \in \{1, \ldots, M\}$ are the data points. The bounds $\underline{y}_j$ and $\overline{y}_j$ are obtained as

$$\underline{y}_j = y_j - \gamma \cdot y_j \cdot cv_j \tag{17}$$

$$\overline{y}_j = y_j + \gamma \cdot y_j \cdot cv_j, \tag{18}$$

where cv_j is the coefficient of variation of the Monte Carlo approximation for y_j. The parameter $\gamma \geq 1$ can be adapted to obtain more conservative bounds. In addition to the constraints in (16) we also invoke the monotonicity constraints of (10) on $w_{\max}$ and $w_{\min}$ in order to ensure monotonicity of the IPM bounds.

Once the IPM is obtained by solving the optimization problem (16), the bounds on the survival signature can be fed into (3) to estimate the bounds of the imprecise survival function. Since bounds of the IPM are strictly monotone functions, it is sufficient to evaluate the survival function for the upper and lower bound. If the CDFs of the component failure times are also imprecise, the bounds on the reliability can be obtained by applying the vertex method [15].

V. Numerical Example

In this section we apply the proposed technique to a numerical example. The example used is that of the Berlin S- and U-Bahn system as presented in [6], see Fig. 2. The entire network consists of 306 nodes and 350 edges. The nodes are separated into two types based on their degree, with type 1 consisting of nodes with degree of two or less, while type 2 contains all nodes with a degree larger than two. As a result, there are 245 nodes of type 1 and 61 of type 2. The full survival signature of this network has 15 252 entries. Let the network efficiency be defined as

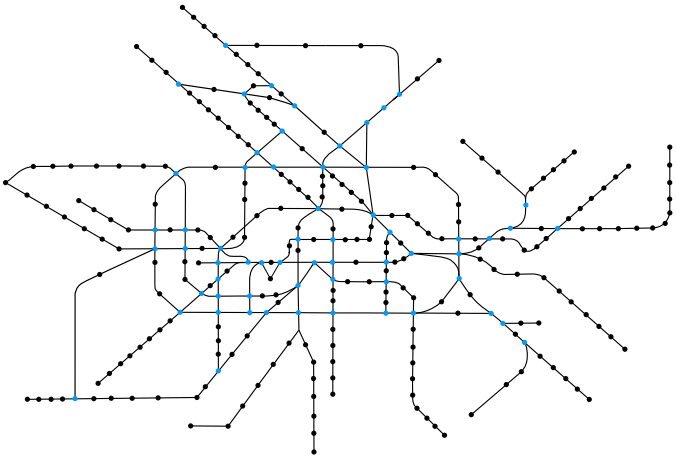

Fig. 2. Topology of the Berlin metro system with 306 nodes. Nodes highlighted in blue represent stations with more than two connections. Adapted from [17].

$$E(G) = \frac{1}{n(n-1)} \sum_{u \neq v \in G} \frac{1}{d_{uv}}, \tag{19}$$

where G is the network consisting of n nodes and d_{uv} measures the length of the shortest path between nodes u and v [16]. Then, a structure function for the network analysis can be defined by the loss of efficiency for a given network state. Here we assume that the network fails to function once the loss of efficiency exceeds 50 % as given by

$$\varphi(\mathbf{x}) = \frac{E(G(\mathbf{x}))}{E(G)} < 0.5. \tag{20}$$

Computing the full analytical survival signature given this structure function is impossible and even the Monte Carlo simulation requires more than 27 hours using 10 000 samples per entry distributed among 64 cores [6].

The approximation of the survival signature starts by applying percolation to find the set of candidate points Ω. The critical threshold $f_c \approx 0.39550$ reduces Ω from 15 252 points down to 5673. 400 center points c_i are generated uniformly over the domain of Ω, of which 109 are excluded using (12), leaving 291 centers. Then, 25 initial points are selected from Ω. New points are chosen using the TEAD until the change in weights is less than 1×10^{-3} for two consecutive points. The procedure aborts after 110 new points have been selected. The survival signature entries for all points are approximated using 10 000 Monte Carlo samples. Fig. 3 shows the initial points as well as the new points chosen by the TEAD. The plot clearly presents how the new points selected by the adaptive technique are concentrated in the area of the survival signature with high relevance whereas only a few points are chosen where the survival signature is negligible. This ability to identify the important region of the signature allows to approximate it accurately using only a low number of evaluated entries.

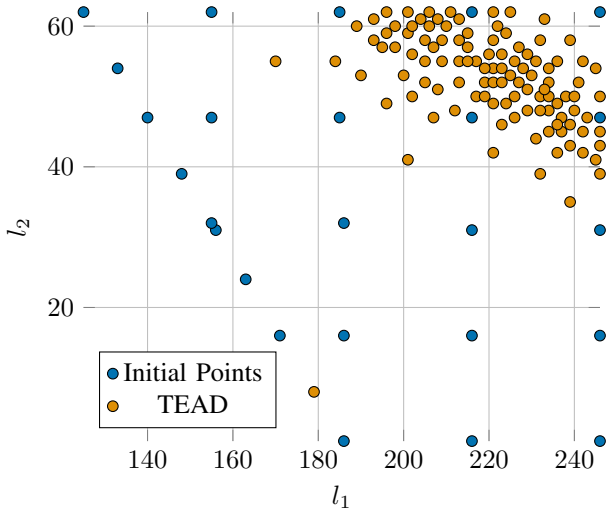

Fig. 3. Initial data points and adaptive points selected by the TEAD.

In the final step, the coefficient of variation resulting from the approximation is used to define the upper and lower bounds of the evaluated survival signature entries and the weights for the IPM are obtained from the optimization (16).

We assume the failure times for component types 1 are exponentially distributed with $\lambda = 0.25$ while the failures of type 2 components are exponentially distributed with $\lambda = 0.5$. These values are arbitrarily chosen for illustrative purposes and have no real world relevance. The bounds on the resulting survival function are compared to the reliability of the full Monte Carlo approximation in Fig. 4. For a closer look at the reliability bounds between $t = 0.25$ and $t = 0.5$ refer to Fig. 5. As evident from the plots, the IPM is able to accurately predict upper and lower bounds on the reliability. This accuracy is especially remarkable due to the fact that it only requires the evaluation of 135 out of the 5673 survival signature values necessary for the full Monte Carlo approach. This is a significant reduction in numerical effort and enables the analysis of even larger networks than using the Monte Carlo method alone.

VI. Conclusions

This paper presents a novel approach to the approximation of the survival signature using normalized radial basis function networks and interval predictor models as surrogate models. The NRBF networks are able to accurately predict the entire survival signature given only a small fraction of the evaluated signature entries. The uncertainty resulting from using Monte Carlo simulation to approximate individual entries is efficiently propagated through the IPM to yield bounds on the survival signature. An adaptive procedure to select the data points for the surrogate model ensures high accuracy while keeping the numerical demand low. The effectivity of the developed method was proven by applying it to the analysis of a large system and comparing the results with a previously presented technique. The code developed as part of this work will be added to the open source Julia package *SurvivalSignature.jl* [18].

Future effort will be focused on improving the NRBF fit. While the number of centers and parameters controlling the spread of the basis functions chosen in this work serve as a good basis, optimizing these parameters to the specific problem could proof beneficial. Furthermore, the method should be applied to even larger systems with more than two component types to investigate the scalability.

By extending this method to include imprecise probabilities for the failure time distributions of the components such as probability boxes, a fully imprecise survival function can be devised.

References

[1] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.

[2] F. P. A. Coolen and T. Coolen-Maturi, "Generalizing the Signature to Systems with Multiple Types of Components," in *Complex Systems and Dependability*, ser. Advances in Intelligent and Soft Computing, W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, and J. Kacprzyk, Eds. Springer, 2013, vol. 170, pp. 115–130.

[3] P. J. Boland and F. J. Samaniego, "The Signature of a Coherent System and Its Applications in Reliability," in *Mathematical Reliability: An Expository Perspective*, ser. International Series in Operations Research & Management Science, R. Soyer, T. A. Mazzuchi, and N. D. Singpurwalla, Eds. Springer, 2004, pp. 3–30.

[4] S. Reed, "An efficient algorithm for exact computation of system and survival signatures using binary decision diagrams," *Reliability Engineering & System Safety*, vol. 165, pp. 257–267, 2017-09-01.

[5] B.-H. Xu, D.-Z. Yang, C. Qian, Q. Feng, Y. Ren, Z.-L. Wang, and B. Sun, "A New Method for Computing Survival Signature Based on Extended Universal Generating Function," in *2019 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*, 2019, pp. 597–602.

[6] J. Behrensdorf, T.-E. Regenhardt, M. Broggi, and M. Beer, "Numerically efficient computation of the survival signature for the reliability analysis of large networks," *Reliability Engineering & System Safety*, vol. 216, p. 107935, 2021.

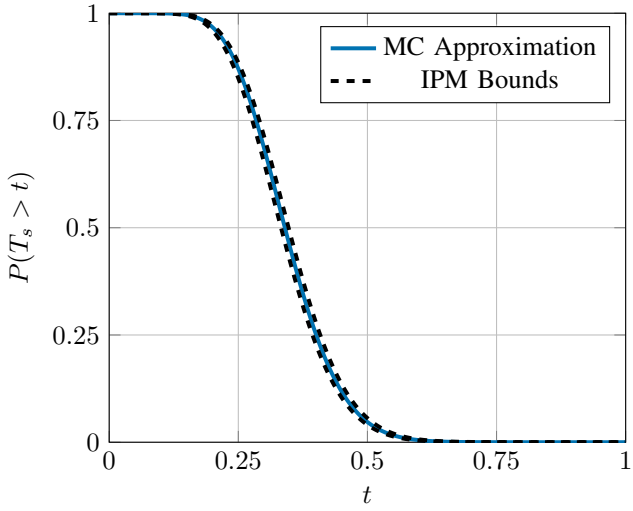

Fig. 4. Bounds of the IPM surrogate model compared to the NRBF approximation.

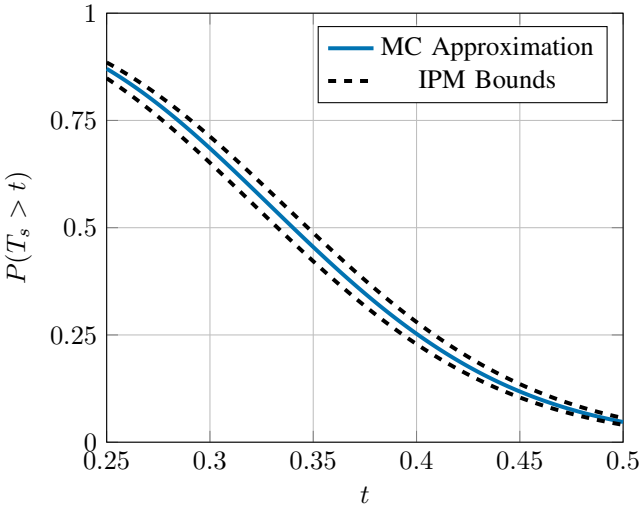

Fig. 5. Bounds of the IPM surrogate model compared to the NRBF approximation between $t = 0.25$ and $t = 0.5$.

[7] F. Di Maio, C. Pettorossi, and E. Zio, "Entropy-driven Monte Carlo simulation method for approximating the survival signature of complex infrastructures," *Reliability Engineering & System Safety*, vol. 231, p. 108982, 2023.

[8] D. S. Broomhead and D. Lowe, "Multivariable Functional Interpolation and Adaptive Networks," *Complex Syst.*, vol. 2, 1988.

[9] F. Heimes and B. van Heuveln, "The normalized radial basis function neural network," in *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*, vol. 2, 1998, pp. 1609–1614 vol.2.

[10] P. Hušek, "On Monotonic Radial Basis Function Networks," *IEEE Trans. Cybern.*, pp. 1–11, 2022.

[11] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding," *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 1042–1068, 2016.

[12] M. Lubin, O. Dowson, J. Dias Garcia, J. Huchette, B. Legat, and J. P. Vielma, "JuMP 1.0: Recent improvements to a modeling language for mathematical optimization," *Mathematical Programming Computation*, 2023.

[13] S. Mo, D. Lu, X. Shi, G. Zhang, M. Ye, J. Wu, and J. Wu, "A Taylor Expansion-Based Adaptive Design Strategy for Global Surrogate Modeling With Applications in Groundwater Modeling," *Water Resources Research*, vol. 53, no. 12, pp. 10 802–10 823, 2017.

[14] L. G. Crespo, D. P. Giesy, and S. P. Kenny, "Interval predictor models with a formal characterization of uncertainty and reliability," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 5991–5996.

[15] W. Dong and H. C. Shah, "Vertex method for computing functions of fuzzy variables," *Fuzzy Sets and Systems*, vol. 24, no. 1, pp. 65–78, 1987.

[16] V. Latora and M. Marchiori, "Efficient Behavior of Small-World Networks," *Physical Review Letters*, vol. 87, no. 19, p. 198701, 2001.

[17] J. Salomon, M. Broggi, S. Kruse, S. Weber, and M. Beer, "Resilience Decision-Making for Complex Systems," *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, vol. 6, no. 2, 2020.

[18] J. Behrensdorf, "Survivalsignature.jl," 2021, https://github.com/FriesischScott/SurvivalSignature.jl.