14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

TIME-DEPENDENT RELIABILITY ANALYSIS FRAMEWORK COMBINING ACTIVE LEARNING AND ENSEMBLE OF SURROGATES

Oscar Sanchez Jimenez[1] **and Younes Aoues**[2]

[1] 1 INSA Rouen Normandie, Normandie Univ, LMN UR 3828,
F-76000 Rouen, France,
oscar.sanchezjimenez@insa-rouen.fr

[2] 1 INSA Rouen Normandie, Normandie Univ, LMN UR 3828,
F-76000 Rouen, France,
younes.aoues@insa-rouen.fr

Key words: Time-variant reliability, Surrogate Models, Reliability

Abstract. We present some advances in the application of ensemble of surrogate models to the study of time-dependent reliability problems. We introduce a framework combining limit state functions with an appropriate temporal discretization and the training of instantaneous surrogates mapping the limit state function. This framework is an extension of the traditional active learning strategy widely used in classical (time-invariant) reliability. Our main result is the implementation of ensembles for this type of problem, and we compare our results with Monte-Carlo simulations and with a single-surrogate strategy based on the same active learning, time-discretization principle.

1 INTRODUCTION

Time-dependent reliability problems remain one of the challenges in contemporary structural, process and machine design and analysis[16]. In the past few decades, reliability criteria have been adopted in engineering practice as a rigorous framework to consider variability and uncertainty, and to study performance and failure in a systematic and formal fashion. Yet, as more complex and detailed models are studied, the traditional reliability theory is incapable of yielding useful results. This complexity comes in the form of:

1) Nonlinearities,

2) Non-gaussian random variables and stochastic processes,

3) Time-varying effects which translate in non-stationary behavior from the probabilistic perspective,

4) Large computational models for the evaluation of solutions, often requiring a high computational cost for the model evaluation, as well as high dimensionality.

In practice, time-dependent reliability analysis is useful in a large variety of design scenarios. One example is that of degradation processes, such as corrosion, fatigue, crack propagation, ageing and biodegradation. Another example is the dynamic response of systems under non-stationary excitations, such as rotating machinery in highly variable environments.

To address the computational burden of realistic models, surrogate-model-based data-driven techniques have been proposed in the recent past ([12, 6, 11, 8]). The Active Learning strategy, in which surrogate models are trained to fit the limit state function in a particular region of interest, has proven successful in reducing computation time in many contexts. However, most of the results in this domain concern time-invariant limit states. While some results exist in the literature addressing the time-dependent generalization, several opportunities remain to be explored.

In this work, we propose a framework to apply surrogate-model-based reliability analysis to time-varying problems, from a non-intrusive, i.e. "black-box", approach. The framework relies on time-discretization and active learning principles to handle the computational complexity of the problem, and it exploits the idea of ensemble of surrogate models to further improvements in the requirements of functional evaluations for the estimation of the probability of failure in a closed interval. The ensemble is composed of Kriging and artificial neural network surrogates. We contrast our ensemble method with Monte-Carlo simulation and with a single surrogate strategy for time-varying problems, the so-called AK-SYS-t method ([1]).

2 Formulation of time-dependent reliability problems

We consider a time-dependent reliability problem described by a limit state function (LSF) in the form:

$$G\left(\mathbf{X}, t\right) = R\left(\boldsymbol{\omega}, t\right) - S\left(\mathbf{X}, t\right), \tag{1}$$

where $\mathbf{R}\left(\mathbf{X}, t\right)$ is the available strength, or resistance of the system under analysis; $\mathbf{S}\left(\mathbf{X}, \boldsymbol{\omega}, t\right)$ is a response quantity; $\mathbf{x}$ is a vector of stochastic inputs, potentially including random variables and stochastic processes; as usual t denotes time. This type of problem is studied on a closed interval $t \in [t_{initial}, t_{final}]$, usually delimiting the design life of the underlying system. The failure associated to this limit state is denoted by convention as the event $G \leq 0$.

The total probability of failure for the LSF in 1 can be written as:

$$P_f^t = \mathbb{P}\left[G\left(\mathbf{X}, \tau\right) \leq 0, \exists \tau \in [t_{initial}, t_{final}]\right], \tag{2}$$

where $\mathbb{P}$ denotes the probability of the event inside the brackets. This probability is computed over the entire life of the system, the interval $[t_{initial}, t_{final}]$. An alternative,

so-called instantaneous probability of failure can be computed by fixing time:

$$P_f^i\left(t_i\right) = \mathbb{P}\left[G\left(\mathbf{X}, \tau\right) \leq 0, \tau = t_i\right], \tag{3}$$

this probability does not take into account the event at previous instants, and so it can not be related in a straightforward manner to the total probability of failure. The interpretation of the latter is the likelihood of the LSF to observe a failure event at any given instant.

An important aspect of the numerical resolution of problems like 1, is the time-discretization of the problem. Different strategies have been proposed to handle the time-variability, as summarized in [19]: from the application of Rice formula with different integration schemes to parallel-system-based methods such as the Phi2 method ([3]) or the JUR/FORM variant ([10]).

A more recent branch of methods ([1]) use the formulation of system reliability, but with a serial interpretation for consecutive time-instants upon a given discretization of the form:

$$t_n = t_{initial} + n\Delta t, n = 0, 1, \ldots, N-1 \tag{4}$$

where Δt is a carefully selected time-step and N is the number of temporal sampling points. This type of discretization facilitates the application of surrogate models to approximate the instantaneous limit state function, as fixing time renders G a random variable that can be handled with the same techniques and surrogates as time-independent problems. The uniform discretization, however, has been shown to be very inefficient for many problems of interest, and we here rely on adaptive, non-uniform temporal discretizations based on [15].

3 Active Learning in time-variant reliability

In the context of surrogate-model-based reliability, active learning is a strategy of sequential optimal experimental design. It aims at exploring the input space in a computationally efficient fashion by selecting regions of interest based on the predicted values of the LSF. An important, pioneering work is [7], where the now classic AK-MCS method is introduced. More recently, the review [14] provides an extensive and detailed review of the subject.

After an initial, often space-filling and small, experimental design has been created, surrogate models are trained to fit the LSF with this training set. Then, the surrogate is used to predict the values of the LSF in the entire input space, and a learning function is used to select the next training point to be selected from this input space. The learning function is computed directly from the prediction of the surrogate, and usually combines the value of the predicted LSF as well as the uncertainty. The most promising point will either reduce the uncertainty in regions of interest, capitalize on regions with low predicted values of the LSF, or some combination of both. We highlight the fact that,

for reliability problems, the essential region of interest is the limit state surface (or surfaces), $G(\mathbf{X}, t) = 0$, which delimits the failure and safety regions.

Let us assume an adequate temporal discretization 4 and a LSF such as 1. An Active Learning algorithm requires: a) a surrogate model; b) a learning function; c) a reliability estimation method. The general algorithm of this branch of methods can be summarized as follows:

1. An initial experimental design is sampled and evaluated, $\mathbf{X}_{ED}$,$Y_{ED} = G(\mathbf{X}_{ED}, t)$, forming $D = \{\mathbf{X}_{ED}, Y_{ED}\}$, the initial training set;
2. A surrogate (or surrogates) is trained to fit D, $\hat{G}(\mathbf{X}, t) \approx G(\mathbf{X}, t)$;
3. The surrogate is used to generalize G outside of D, by predicting its value on the entire input space $\mathbf{X}$, yielding predictions μ_i and uncertainties e_i at each point $\mathbf{x}_i$;
4. The learning function is used to compute $\mathbf{x}_{new}$, the most promising new point ;
5. The point $\mathbf{x}_{new}$ is evaluated in the true LSF: $y_{new} = G(\mathbf{x}_{new}, t)$, and the training set D is enriched, $D \leftarrow \{D \cup \{\mathbf{x}_{new}, y_{new}\}\}$;
6. Some stopping criterion is evaluated to determine whether the algorithm can halt: if yes, the trained surrogates are used to compute the probabilities of failure with the selected estimation method; if no, steps 2 to 5 are repeated.

A popular learning function in reliability is the U-function ([7]):

$$U(\mathbf{x}) = \frac{|\mu_G(\mathbf{x})|}{\sigma_G(\mathbf{x})}, \tag{5}$$

where $\mu_G(\mathbf{x})$ denotes the surrogate prediction of G at candidate point $\mathbf{x}$, and $\sigma_G(\mathbf{x})$ denotes the prediction uncertainty of the surrogate at $\mathbf{x}$. The learning criterion is to select as $\mathbf{x}_{new}$ the candidate point that minimizes U. The points that minimize this learning function will be points that either have large prediction uncertainty, large $\sigma_G(\mathbf{x})$, or points near the failure boundary, $|\mu_G(\mathbf{x})|$ close to 0.

To adapt the learning function 5 to time-varying problems, the AK-SYS-t ([1]) method proposes the application of the composite learning function criterion from AK-SYS ([9]) to the temporal discretization introduced with 4. This new learning function reads:

$$U_s(\mathbf{x}) = \frac{|\mu_G(\mathbf{x}, t_s)|}{\sigma_G(\mathbf{x}, t_s)} \tag{6}$$

$$t_s = \underset{\tau \in \{t_n\}}{\text{ArgMin}} \left[\mu_G(\mathbf{x}, \tau)\right], \tag{7}$$

for each point sampled from the input space, $\mathbf{x}_i, i = 1, \ldots, N_S$, the LSF and uncertainty are calculated at each instant of the discretization $\{t_n\}, n = 0, \ldots, N-1$; of the N "instantaneous" LSF predictions that correspond to each $\mathbf{x}_i$, the one with the smallest predicted value is selected as representative, which corresponds to t_s. The learning criterion is then applied to U_s, and consequently, the new training point $\mathbf{x}_{new}$ is the one that minimizes U on its smallest predicted instantaneous limit state.

Popular combinations of the active learning algorithm in reliability consist of a Kriging surrogate, the U learning function with stopping criterion Min $[U] > 2$, and Monte-Carlo Simulation for reliability estimation. This combination is effective, easy to implement and versatile, but it is not the most efficient (See: [14]). In particular, the condition Min $[U] > 2$ is often conservative and leads to the evaluation of many enrichment points after the desired probability of failure has converged. Other proposed stopping criteria consider the variation of the total probability of failure between iterations, or in the case of FORM-based reliability estimations, the variation of the reliability index; the maximum number of iterations, or budget, is another possibility. As we will show later, the use of ensembles of surrogates allows for the formulation of a new type of stopping criterion: ensemble agreement or consensus.

4 Ensemble of surrogate models

An ensemble of surrogates can be understood as a set of M surrogates approximating the same underlying function G, accompanied by some decision-making or "ensemble" rule that dictates how predictions and uncertainties from each surrogate are combined to yield the ensemble prediction and uncertainty. Given M surrogates $\hat{G}^{(M)}$, with respective estimation uncertainty at x_i: e_i^M, some weighted average can be used to give:

$$\hat{G}^{(E)}(x_i) = \sum_{j=1}^{M} w\left(e_i^j\right) \hat{G}^{(j)}(x_i), \tag{8}$$

where $w(\cdot)$ is some weight function, and the superscript E denotes the ensemble prediction. Some examples for the choice of w include:

1. Best surrogate approach: every weight is set to zero, except for the weight of the surrogate with the smallest prediction uncertainty;
2. Weighted average approach: the weights are inversely proportional to the prediction uncertainty, and chosen to be normalized over the number of surrogates;
3. Hierarchical approach: the weights are fixed, and assigned based on a goodness-of-fit hierarchy.

Among the possible surrogates, we distinguish the two utilized in this article: Kriging and Artificial Neural Networks. We use the Kriging surrogate formulation presented in [5] and [13], where the Kriging surrogate has the form:

$$G\left(\mathbf{X}\right) = \beta^T f\left(\mathbf{X}\right) + \sigma Z\left(t\right), \tag{9}$$

the terms on the right are respectively the trend and a standard normal stationary Gaussian Process multiplied by a standard deviation (effectively a hyper-parameter). This surrogate is convenient, flexible, and has the advantage of providing an intrinsic measure of prediction uncertainty.

The Artificial Neural Network we have selected is a shallow Feedforward Neural Network. It can be understood as an universal approximator of the form (see: [18]):

$$G\left(\mathbf{X}\right) = H_{N_L}\left(U_{NL}\left(\mathbf{X}\right)\right), \tag{10}$$

with N_L being the output layer, and U_{NL} being the the output of each neuron in the hidden layer, or:

$$H_{i,j}\left(U_{i,j}\right) = \rho\left(U_{i,j}\right); \begin{cases} i \in N_L & \text{neuronnumber} \\ j \in N_k & \text{layernumber} \end{cases} \tag{11}$$

$$U_{i,j} = \sum_{p=1}^{N_{i-1}} Y_p^{i-1} W_{p,j}^i, \tag{12}$$

with W being an importance or weight of the p-th neuron on the i-th layer onto the j-th neuron in layer I, and similarly Y is the output of the corresponding neuron in the previous layer. The function ρ is the so-called activation function of the neuron. The FNN architechture when used as a surrogate is simple and easy to train, and can be enhanced by the implementation of numerous types of activation functions and tuning algorithms. A major drawback of FFNNs is the lack of an embedded measure of prediction uncertainty. We here rely on the approach proposed in [17], consisting on 5-fold cross-validation and the jackknife estimate for ANN uncertainty.

5 Proposed framework

Our proposed framework consist of the following elements:

1. A time-discretization scheme: this can be some uniform temporal grid, or some adapted temporal discretization (here, we use the latter as described in [15])

2. An ensemble definition: this includes the ensemble composition rule, the types of surrogate models to be used and their parametrization

3. One or more stopping criterion.

In our experiments, we will explore the Kriging-FFNN combination (only two surrogates at each instant), detailing the selected parametrization in each case. Our composition rule is the "best surrogate" for each $\mathbf{x}_i$ and t_i. We keep the U-function from AK-SYS-t. We impose a budget stopping criterion, and we introduce the surrogate consensus stopping criterion: the algorithm stops if the independent prediction of both surrogates over the probability of failure is sufficiently close to the ensemble prediction. A similar stopping criterion is used in [17] for time-invariant reliability problems.

Our framework includes the AK-SYS-t method, taking an uniform discretization in 1, selecting an ensemble of 1 member (Kriging) in 2, using a single stopping criterion based on the minimum of U. Given this particularity and the similar features of both approaches, in the examples we will contrast the performance and behavior of our ensemble with that of our implementation of AK-SYS-t. Among the metrics of interest, we can cite the number of iterations, the accuracy, the computation time, and the enrichment points selected with each approach.

6 Examples

6.1 1D example

Our first example is a nonlinear function with a single stochastic dimension, used as a text function in [11], [12] and [2]:

$$G\left(X,t\right)=0.014-\frac{1}{X^{2}+4}\sin\left[2.5X\right]\cos\left[\left(t+0.4\right)^{2}\right] \tag{13}$$

where $X\ \rightarrow\ \mathscr{N}\left(10,1\right)$. The interval of interest is $[t_0=1,t_f=2.5]$. The following specifications have been used:

- The Kriging surrogate has a Gaussian separable correlation, it is Ordinary Kriging;
- The FFNN has a Radial Basis activation function, trained using the Bayesian Regularization algorithm; it has 1 hidden layer of 5 neurons;
- The total budget for this problem is 100 functional evaluations outside of the initial experimental design, this initial experimental design has size 10;
- The AK-SYS-t Kriging has the same settings as the ensemble Kriging, its stopping criterion is also 100 functional evaluations or U;
- The ensemble is stopped by 3 conditions: budget, ensemble consensus, convergence of the total probability of failure.

We first report the results of a single, typical run, which will allow us to analyze the behavior of both AK-SYS-t and Ensemble. We then report some statistics of multiple runs, to account for the stochastic nature of both methods.

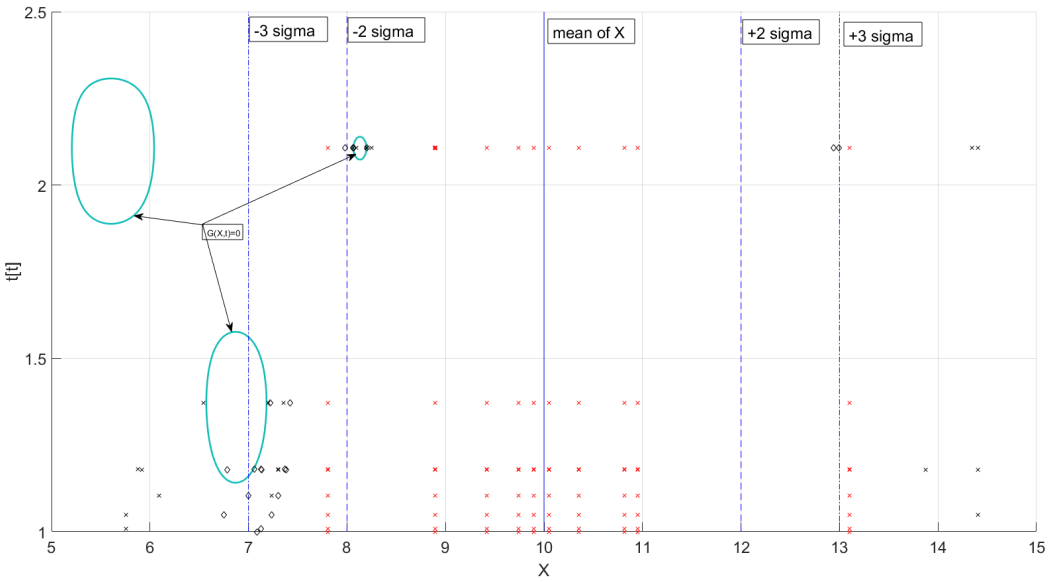

Figure 1: Comparison of AK-SYS-t and Ensemble: true LSF and enrichment points; red-initial experimental design; black circle-AK-SYS-t enrichment; black diamond-Ensemble

In 1, we see the contour plot of the LSF 13: the horizontal, where the horizontal axis is the input dimension and the vertical axis is time. The blue stripes denote, respectively, the mean value of the input, 2 and 3 sigmas away from the mean, giving a measure of the probability mass of the input. The blue closed curves represent failure regions, and their proximity to the blue horizontal lines gives an intuition of how much each region contributes to the total probability of failure. In red, we have the initial experimental design at each one of the time-instants of the discretization. Finally, the black circles denote the enrichment points from the AK-SYS-t algorithm, and the black diamonds denote the enrichment points from the ensemble.

From 1 we can make a few observations: the exploration of the ensemble method seems better guided, as evidenced by the amount of exploration points to the right of the mean (3 for the ensemble, 5 for AK-SYS-t). For this run, both methods seem to explore the smaller failure region near $X = 8$, but the ensemble method seems to cover the larger failure region intersected by $X = 7$ with more precision, as AK-SYS-t continues to add exploration points into the X regions past this failure zone.

In the following table, we present selected statistics to compare 10 runs of the ensemble method and our AK-SYS-t implementation. We have selected to show the median, maximum and minimum values of the percent error with respect to the MCS reference solution, as well as the number of functional evaluations (NFE) discounting the experimental design. The AK-SYS-t method achieves a good level of accuracy with a low

number of functional evaluations; our ensemble method has a lower median error, with some realizations achieving virtually no error with respect to the MCS reference, and with the same maximal error as AK-SYS-t. Of the 10 runs, 1 stopped due to ensemble "consensus", 9 stopped due to the criterion on the minimum value of U, the condition on the maximum number of functional evaluations was never triggered. The run triggering the ensemble consensus had the lower NFE, at 14 iterations, with a percent error of 3% ($P_f^t = 1.13 \times 10^{-1}$).

Our results suggest that the ensemble method, as implemented for this article, is more variable than the basic AK-SYS-t algorithm. Indeed, two extreme behaviors have been showcased during our experiments: instances in which the method converges to the reference probability of failure with less than 0.5% error, and cases in which the method halts at 9% error; similarly, the ensemble can converge with as little as 14 iterations, but some runs result in as high as 54. Consistent with our previous observations, it is apparent that the addition of the FFNN enables the method to explore regions of interest more effectively. However, the sensitivity of the method to the randomness inherent to the ANN training/fitting needs to be further studied.

Method	Median % error	Max. % error	Min. % error	Median NFE	Max. NFE	Min. NFE
AK-SYS-t	4	9	1	24.5	41	17
Ensemble	1.5	9	0	34	54	14

Table 1: Comparison of ensemble method with AK-SYS-t for 10 independent runs

To conclude with this example, a typical convergence plot of the ensemble is shown in 2. For each iteration, the probability of failure is computed with each surrogate as well as the ensemble. The computed probabilities of failure are normalized by the reference MCS value: $\frac{P_{f,M}^t}{P_{f,MCS}^t}$ where M stands for either Kriging, ANN or ensemble surrogate. We see that, initially, the FFNN helps guide the Kriging surrogate. The ensemble prediction (green) is closer to the Kriging prediction (red). After the 20th iteration, the FFNN has an erratic behavior, and subsequently converges along the rest of the ensemble towards the true probability of failure. We highlight that the ensemble prediction, despite being dominated by the Kriging surrogate, is in fact always some combination of Kriging and FFNN: our "best surrogate" criterion is not just instantaneous (the best surrogate at a given t), it is also local, which means that for each input point $\mathbf{x}_i$, the best surrogate is selected for the prediction. From the perspective of MCS simulation, this means that at any given instant, a fraction of the sampled points is always best approximated by the FFNN.

6.2 5D example with discretized stochastic process input

The second example is a more complex mathematical test case presented in [4]. The limit state function has the form:

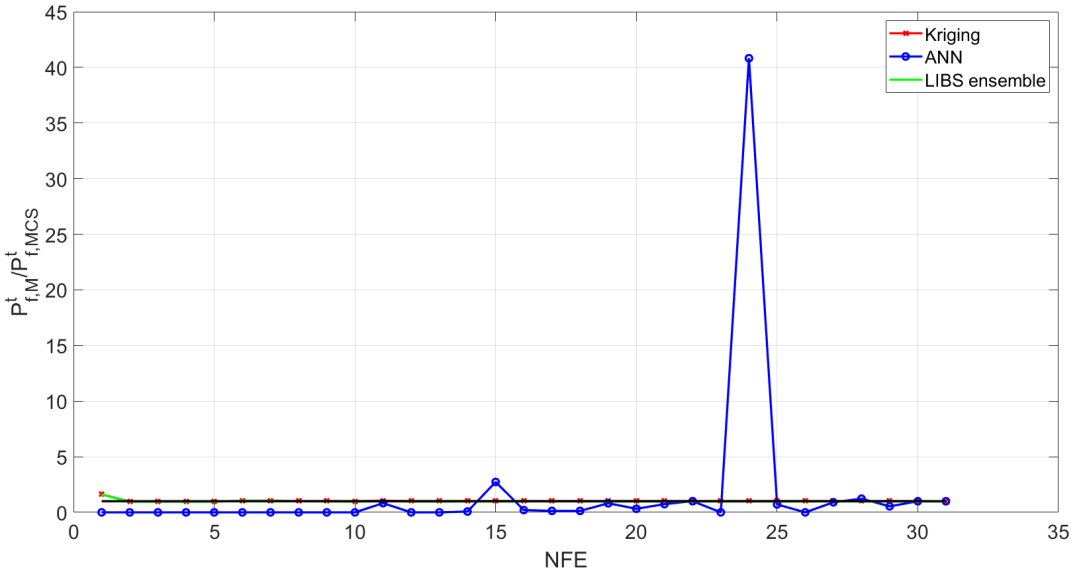

Figure 2: Convergence of normalized predicted total probability of failure with respect to the number of functional evaluations (NFE); ensemble and each surrogate are shown.

$$G\left(X,t\right) = {X_1}^2 X_2 - 5X_1(1-Y).*t + (X_2+1)t^2) - 20; \tag{14}$$
$$Y\left(t\right) = \sum_{i=1}^{3} \lambda_i \phi_i\left(t\right) X_{2+i}$$

where X_1, X_2 are normal random variables with mean 3.5 and standard deviation 0.25; the Gaussian stationary input $Y\left(t\right)$ has been discretized using a K-L expansion with 3 terms with the independent random variables X_3, X_4, X_5 having zero mean and unit standard deviation. The interval of interest is $[0, 1]$. The correlation function of the load has the form $R_{YY}\left(t_1, t_2\right) = \exp\left[-\left|t_2 - t_1\right|^2\right]$.

The reference probability of failure is $P_f^t = 0.3074$, with coefficient of variation below 0.005, using Monte-Carlo simulation with an uniform time discretization of 11 time nodes. While the reference reports 10^6 samples for the MCS benchmark, we have found that 1×10^5 yields a reasonable approximation with a satisfactory coefficient of variation (below 0.05). Similarly, using the K-L discretization scheme, the input can be expressed in terms of 3 (not 4) random variables and the corresponding eigenfunctions and eigenvalues, ϕ_i and λ_i^2. The variance percentage for this representation is above 99.95%.

For this problem, as before, the configuration of the Kriging surrogate in the AK-SYS-t method and in the Ensemble are identical, as well as the size of the experimental

Method	Median % error	Max. % error	Min. % error	Median NFE	Max. NFE	Min. NFE
AK-SYS-t	0	1	0	100	100	100
Ensemble	0	2	0	53	100	0

Table 2: Comparison of ensemble method with AK-SYS-t for 20 independent runs

design, $ED = 20$. Ordinary Kriging with a Gaussian correlation has been used, this time without the assumption of separability. For the FFNN, a shallow architecture of 1 layer with 5 Neurons has been used; a Poslin activation function (similar to the popular ReLU) has been selected based on preliminary performance analysis. Bayesian regularization Back-propagation has been selected as the learning function, with input and output Min-Max normalization.

The stopping criteria for the ensemble are identical as in the previous example. First, a function evaluation budget of 100 has been put in place. Second, the traditional condition on the minimum of the U function being larger than 2 has been maintained. Finally, the consensus criterion remains the same, given by $\max\left[\left|P^t_{f,K} - P^t_{f,e}\right| P^t_{f,ANN} - P^t_{f,e}\right] < 0.01$, where $P^t_{f,K}$ is the Kriging-sub-ensemble prediction of the total probability of failure, $P^t_{f,ANN}$ is the FFNN/ANN-sub-ensemble prediction, and $P^t_{f,e}$ is the ensemble prediction.

In 2, we compile the results of the AK-SYS-t method and the ensemble, this time for 20 independent runs each. The errors have been rounded to the nearest integer, as the discrepancies were found to be very small on each case. In both cases, the median error is 0. The more remarkable feature of this comparison is in the number of functional evaluations. For AK-SYS-t, every run triggered the maximum evaluation budget, 100. The ensemble, on the other hand, triggered the consensus stopping criterion 14 times out of 20. In one of the runs, the consensus was immediate: zero NFE, $P^t_{f,e} = 3.08 \times 10^{-1}$, with approximately zero error.

This example illustrates an effect of the U function in certain scenarios, already remarked by [14]: in some circumstances, the stopping criterion $\text{Min}\,[U] > 2$ can be extremely conservative, as one can readily see that the underlying probability of failure has already converged to the right value, but the condition is not triggered. The ensemble consensus appears to be an effective alternative, which can potentially result in important savings in terms of functional evaluations. The advantage of the "secondary" surrogate (the FFNN) in this case has been threefold: 1) it helps in the exploration of some regions of interest; 2) it enhances the numerical precision in some of the time nodes already dominated by the Kriging surrogate; 3) it enables the establishment of a stopping mechanism based on surrogate consensus.

We stress that the feasibility of our ensemble method, as with any surrogate-based method, heavily depends on the good tuning of the underlying surrogates. This is a task that requires case-by-case analysis and testing. In the case of the Kriging surrogate, the tuning is mostly confined to the selection of a form of Kriging (Ordinary, Universal,

PC-K); in the case of the FFNN, the tuning becomes trickier, as one must select architecture (number of layers and number of neurons per layer), activation function, training algorithm, and regularization schemes and potentially re-training schemes.

Conclusion

We have presented a framework combining the ideas of ensemble of surrogate models and the branch of Active-Learning-Based reliability methods that rely on temporal discretization. Our framework permits the combination of multiple surrogates to approximate the limit state function, and to set several combination rules for the computation of the prediction and uncertainty of the ensemble. The framework has been combined with traditional time discretization schemes, as well as more recent techniques that provide adaptive non-uniform temporal discretization. The feasibility of our approach has been demonstrated with two mathematical examples with distinct properties in terms of temporal variability, stochastic complexity and dimension.

The proposed framework has been contrasted with raw Monte-Carlo simulation and with a more recent Active-learning-based time-variant reliability method: AK-SYS-t. For the two selected examples, the advantages of the ensemble of surrogates have been illustrated. The use of multiple surrogates provides improved exploration behavior. The numerical precision appears to be enhanced ,in some situations, by the addition of a secondary surrogate (in our case, a Feedforward Neural Networ). Most interestingly, the addition of a second surrogate allows the definition of a new type of stopping criterion based on the agreement between the surrogates. While, in principle, this type of stopping criterion has already been applied in the literature, to our knowledge this is the first instance in which such criterion is discussed in the context of time-variant reliability.

Future work in this framework would involve the consideration of more diverse surrogates. Additionally, the exploration of new ensemble composition rules remain to be tested. A more extensive study of the behavior of shallow neural networks in time-variant reliability appears to be pertinent. Finally, more complex problems, including true black-box functions, remain to be tackled by our proposal.

REFERENCES

[1] Morteza Ahmadivala, Cécile Mattrand, Nicolas Gayton, André Orcesi, and Thierry Yalamas. Ak-sys-t: New time-dependent reliability method based on kriging metamodeling. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 7(4):04021038, 2021.

[2] Morteza Ahmadivalaa, Cécile Mattrandb, Nicolas Gaytonb, Antoine Dumasa, Thierry Yalamas, and André Orcesi. Application of AK-SYS method for time-dependent reliability analysis. In *Congres Français de Mécanique*, 2019.

[3] C. Andrieu-Renaud, Bruno Sudret, and Maurice Lemaire. The PHI2 method: a way to compute time-variant reliability. *Reliability Engineering & System Safety*, 84(1):75–86, 2004.

[4] Weiqi Du, Yuanxin Luo, and Yongqin Wang. Time-variant reliability analysis using the parallel subset simulation. *Reliability Engineering & System Safety*, 182:250–257, 2019.

[5] Vincent Dubourg. *Adaptive surrogate models for reliability analysis and reliability-based design optimization*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2011.

[6] Benjamin Echard, Nicolas Gayton, and Maurice Lemaire. Ak-mcs: an active learning reliability method combining kriging and monte carlo simulation. *Structural safety*, 33(2):145–154, 2011.

[7] Benjamin Echard, Nicolas Gayton, and Maurice Lemaire. Ak-mcs: an active learning reliability method combining kriging and monte carlo simulation. *Structural safety*, 33(2):145–154, 2011.

[8] William Fauriat and Nicolas Gayton. Ak-sys: an adaptation of the ak-mcs method for system reliability. *Reliability Engineering & System Safety*, 123:137–144, 2014.

[9] William Fauriat and Nicolas Gayton. Ak-sys: an adaptation of the ak-mcs method for system reliability. *Reliability Engineering & System Safety*, 123:137–144, 2014.

[10] Zhen Hu and Xiaoping Du. Time-dependent reliability analysis with joint upcrossing rates. *Structural and Multidisciplinary Optimization*, 48(5):893–907, June 2013.

[11] Zhen Hu and Xiaoping Du. Mixed efficient global optimization for time-dependent reliability analysis. *Journal of Mechanical Design*, 137(5):051401, 2015.

[12] Zhen Hu and Sankaran Mahadevan. A single-loop kriging surrogate modeling for time-dependent reliability analysis. *Journal of Mechanical Design*, 138(6):061406, 2016.

[13] C. Lataniotis, D. Wicaksono, S. Marelli, and B. Sudret. UQLab user manual – Kriging (Gaussian process modeling). Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2022. Report UQLab-V2.0-105.

[14] Maliki Moustapha, Stefano Marelli, and Bruno Sudret. Active learning for structural reliability: Survey, general framework and benchmark. *Structural Safety*, 96:102174, 2022.

[15] Sebastien Leveneur Oscar Sanchez Jimenez, Younes Aoues. A black-box time-discretization scheme for time-dependent reliabilityproblems. In *Proceedings of the 8th International Conference on System Reliability and Safety*, 2024.

[16] Rüdiger Rackwitz. Reliability analysis— a review and some perspectives. *Structural safety*, 23(4):365–395, 2001.

[17] Chao Ren, Younes Aoues, Didier Lemosse, and Eduardo Souza De Cursi. Ensemble of surrogates combining kriging and artificial neural networks for reliability analysis with local goodness measurement. *Structural Safety*, 96:102186, 2022.

[18] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1):43–62, 1997.

[19] Bohua Zhang, Weigang Wang, Yanlin Wang, Yueru Li, and Chun-Qing Li. A critical review on methods for time-dependent structural reliability. *Sustainable and Resilient Infrastructure*, 9(2):91–106, May 2023.