Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2023) 158-164

ECCOMAS
Proceedia

UNCECOMP 2023
5th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Science and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Athens, Greece, 12-14 June 2023

LOCALIZED APPROXIMATIONS BY POLYNOMIAL CHAOS EXPANSIONS FOR HIGHLY NON-LINEAR FUNCTIONS

Lukáš Novák[1], **Michael D. Shields**[2], **Václav Sadílek**[1], **Miroslav Vořechovský**[1]

[1]Brno University of Technology
Brno, Czech Republic
e-mail: {novak.l, sadilek.v, vorechovsky.m}@fce.vutbr.cz
[2]Johns Hopkins University
Baltimore, USA
e-mail: michael.shields@jhu.edu

Abstract.

The paper presents a strategy for adaptive sequential decomposition of the design domain and for localized surrogate modeling via polynomial chaos expansion. The proposed method combines adaptivity of localized surrogate modeling and sequential sampling. The iteration process of the decomposition and sequential sampling balances between exploitation of the surrogate model and exploration of the design domain. The proposed methodology is especially beneficial for highly non-linear functions, which are extremely challenging for and approximation by polynomial chaos expansion. The obtained numerical results confirm its superiority over a single global surrogate model and a recently proposed technique based on similar concepts.

Keywords: Polynomial Chaos Expansion, Adaptive Sampling, Sequential Sampling.

ISSN:2623-3339
Peer-review under responsibility of the organizing committee of UNCECOMP 2023.
doi: 10.7712/120223.10331.19858

1 INTRODUCTION

It is typically highly computationally demanding to create a single global polynomial chaos expansion (PCE) approximation of a mathematical model with high local non-linearity. Computational requirements grow rapidly with both the maximum polynomial order p of PCE and the dimension of a given stochastic model. Unfortunately, in the case of functions with localized non-linearities, the use of high order p is necessary despite the fact that a major part of the function could have been approximated with significantly lower p. A possible solution to this problem is a selection of a sparse set of basis functions by the best-model selection algorithm, which can dramatically reduce the number of basis functions and thus the number of evaluations of the original mathematical model.

Another solution, investigated in the present paper, is to divide the whole design domain into sub-domains in which approximation by a low-order PCE suffices. This contribution presents a novel algorithm for the iterative identification of the sub-domain containing significant non-linearity and approximation of the whole original function by local PCEs. The proposed algorithm is based on the concept of active learning governed by a recently proposed variance-based adaptive sequential sampling [8]. Active learning for PCE can greatly improve an accuracy of the surrogate model by sampling in area of interest [10, 14], here we employ it to identify sub-domains containing significant non-linearities. Each sub-domain has its own PCE surrogate model and thus it is possible to effectively reduce p and associated size of experimental design.

2 POLYNOMIAL CHAOS EXPANSION

The PCE represents the quantity of interest (QoI) Y obtained as a result of the original mathematical model $\mathcal{M}(X)$, as a polynomial expansion of an another random variable ξ called a germ with a known distribution. A set of polynomials, orthogonal with respect to the probability distribution of the germ, are used as a basis of the Hilbert space of all real-valued random variables of finite variance. The orthogonality condition for all $j \neq k$ is given by the inner product of the Hilbert space defined for any two functions ψ_j and ψ_k with respect to the weight function p_ξ (probability density function of ξ) as:

$$\langle \psi_j, \psi_k \rangle = \int \psi_j(\xi)\psi_k(\xi)p_\xi(\xi)\,\mathrm{d}\xi = 0. \tag{1}$$

Orthogonal polynomials ψ corresponding to common probability distributions p_ξ can be chosen according to Wiener-Askey scheme [13] or numerically constructed in case of arbitrary probability distributions [12]. In the case of $\boldsymbol{X}$ and $\boldsymbol{\xi}$ being vectors containing M random variables, the polynomial $\Psi(\boldsymbol{\xi})$ is multivariate and it is built up as a tensor product of univariate orthogonal polynomials.

The QoI $Y = \mathcal{M}(\boldsymbol{X})$, can then be represented as [3]:

$$Y = \mathcal{M}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^M} \beta_{\boldsymbol{\alpha}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \tag{2}$$

where $\boldsymbol{\alpha} \in \mathbb{N}^M$ is a set of integers called the multi-index, $\beta_{\boldsymbol{\alpha}}$ are deterministic coefficients and $\Psi_{\boldsymbol{\alpha}}$ are multivariate orthogonal polynomials. For practical computation, PCE expressed in Eq. (2) must be truncated to a finite number of terms P. The truncation is commonly achieved by retaining only terms whose total degree $|\boldsymbol{\alpha}|$ is less than or equal to a given p.

From a statistical point of view, truncated PCE is a simple linear regression model with intercept. Therefore, it is possible to use ordinary least square (OLS) regression as simple

non-intrusive solution. In order to use OLS for $\boldsymbol{\beta}$ estimation, it is necessary to first obtain n_{sim} realizations of the $\boldsymbol{X}$ and the corresponding results of the original mathematical model $\mathcal{Y}$, together called the experimental design (ED). Then, the vector of deterministic coefficients $\boldsymbol{\beta}$ is calculated using data matrix $\boldsymbol{\Psi}$ as

$$\boldsymbol{\beta} = \left(\boldsymbol{\Psi}^T \boldsymbol{\Psi}\right)^{-1} \boldsymbol{\Psi}^T \mathcal{Y}. \tag{3}$$

The optimal size of ED is clearly affected by the number of terms P dependent on M and p. Therefore, it is typically useful to find sparse solution using advanced model selection algorithms such as Least Angle Regression (LAR) [2, 1], orthogonal matching pursuit [11] or Bayesian compressive sensing [4] to find an optimal set of basis functions further denoted by $\mathcal{A}$.

Once the PCE is constructed, the first statistical moment (the mean value) is simply the first deterministic coefficient of the expansion $\mu_Y = \left\langle Y^1 \right\rangle = \beta_{\mathbf{0}}$. The second raw statistical moment, $\left\langle Y^2 \right\rangle$, can be obtained as

$$\begin{aligned} \left\langle Y^2 \right\rangle &= \int \left[\sum_{\boldsymbol{\alpha} \in \mathcal{A}} \beta_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}} \left(\boldsymbol{\xi}\right) \right]^2 p_{\boldsymbol{\xi}} \left(\boldsymbol{\xi}\right) \, \mathrm{d}\boldsymbol{\xi} = \sum_{\boldsymbol{\alpha}_1 \in \mathcal{A}} \sum_{\boldsymbol{\alpha}_2 \in \mathcal{A}} \beta_{\boldsymbol{\alpha}_1} \beta_{\boldsymbol{\alpha}_2} \int \Psi_{\boldsymbol{\alpha}_1} \left(\boldsymbol{\xi}\right) \Psi_{\boldsymbol{\alpha}_2} \left(\boldsymbol{\xi}\right) p_{\boldsymbol{\xi}} \left(\boldsymbol{\xi}\right) \, \mathrm{d}\boldsymbol{\xi} \\ &= \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \beta_{\boldsymbol{\alpha}}^2 \int \Psi_{\boldsymbol{\alpha}} \left(\boldsymbol{\xi}\right)^2 p_{\boldsymbol{\xi}} \left(\boldsymbol{\xi}\right) \, \mathrm{d}\boldsymbol{\xi} = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \beta_{\boldsymbol{\alpha}}^2 \left\langle \Psi_{\boldsymbol{\alpha}}, \Psi_{\boldsymbol{\alpha}} \right\rangle . \end{aligned} \tag{4}$$

Considering the orthonormality of the polynomials, it is possible to obtain the variance $\sigma_Y^2 = \left\langle Y^2 \right\rangle - \mu_Y^2$ as the sum of all squared deterministic coefficients except the intercept (which represents the mean value). Note that the computation of higher statistical central moments – specifically skewness γ_Y (3^{rd} moment) and kurtosis κ_Y (4^{th} moment) – are more complicated since they require triple and quad products in Eq. 4. These can be obtained analytically only for certain polynomial families, e.g. formulas for Hermite and Legendre polynomials (and their combination) can be found in [7].

3 ACTIVE LEARNING-BASED DOMAIN ADAPTIVE LOCALIZED PCE

The core of the proposed approach further referenced as Domain Adaptive Localized PCE (DAL-PCE) is a sequential decomposition of the input random space $\mathcal{D}$ for the construction of local approximations. The proposed methodology aims to iterativelly decompose the design domain $\mathcal{D}$ into $n_{\mathcal{D}}$ smaller non-overlapping sub-domains $\mathcal{D}_i \subset \mathcal{D}$ and to create localized piecewise low-order PCEs that are valid only in individual sub-domains $\mathcal{D}_i$. The constructed set of low-order PCEs can approximate highly non-linear functions with significantly lower ED in comparison to a single global PCE. The whole iterative process contains the following main steps after the construction of the initial global PCE and its first splitting into two parts:

1. identification of sub-domain associated with the highest variance density;
2. splitting of the identified sub-domain;
3. extension of the experimental design in the identified sub-domain;
4. constructions of local PCE for a new sub-domain.

The significant advantage in comparison to the existing methods is an active learning feature which identifies a single sub-domain, which is further processed for construction of local PCE.

An active learning is employed in these two specific tasks: (i) identification of an important sub-domain, that is, a domain that is either large compared to other sub-domains or that is associated with a high local variance (see Eq. 8); and (ii) identification of the best positions for additional samples extending the current ED in the selected sub-domain (see Eq. 5). Each of these steps must be based on a criterion that balances *exploration* of the input random space with *exploitation* of the surrogate model, which in our case is in the form of a PCE. The Θ-criterion for adaptive sequential sampling [8] is employed for both steps:

$$\Theta(\boldsymbol{\xi}^{(\mathrm{c})}) \equiv \Theta^c = \underbrace{\sqrt{\sigma^2_{\mathcal{A}}(\boldsymbol{\xi}^{(\mathrm{c})}) \cdot \sigma^2_{\mathcal{A}}(\boldsymbol{\xi}^{(\mathrm{s})})}}_{\text{ave variance density}} \underbrace{l^M_{\mathrm{c,s}}}_{\text{vol.}}. \tag{5}$$

The *exploration* aspect is maintained by accounting for the distance $l_{\mathrm{c,s}}$ between a candidate $\boldsymbol{\xi}^{(\mathrm{c})}$ and its nearest neighboring realization from the existing ED, $\boldsymbol{\xi}^{(\mathrm{s})}$ as

$$l_{\mathrm{c,s}} = \sqrt{\sum_{i=1}^{M} |\xi_i^{(\mathrm{c})} - \xi_i^{(\mathrm{s})}|^2}. \tag{6}$$

The *exploitation* component aims to sample points in regions with the greatest contributions to the total variance of the QoI σ_Y^2, i.e. at points with the highest *variance density* (see Eq. 4) defined as

$$\sigma^2_{\mathcal{A}}(\boldsymbol{\xi}) = \Big[\sum_{\boldsymbol{\alpha}\in\mathcal{A},\, \boldsymbol{\alpha}\neq\mathbf{0}} \beta_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \Big]^2 p_{\xi}(\boldsymbol{\xi}). \tag{7}$$

The selection process to identify the most important sub-domain for further processing is governed by extending the Θ-criterion as follows

$$\Theta_i = \underbrace{\mathcal{W}_i \cdot \exp(Q_i^2)}_{\text{weight of subdomain}} \cdot \underbrace{\sqrt{\sigma^2_{\mathcal{A}_i}(\boldsymbol{\xi}^{(c)}) \cdot \sigma^2_{\mathcal{A}_i}(\boldsymbol{\xi}^{(s)})}\, l^M_{c,s}}_{\Theta^c \text{ in } i\text{th subdomain}}. \tag{8}$$

This extended criterion aims to identify sub-domains of the input random space associated with the maximum value of Θ^c, while simultaneously accounting for the size of each subdomain and the accuracy of the existing local PCE. Θ^c is calculated for a rich pool of screening global candidates, while the volume of each sub-domain $\mathcal{W}_i$ and the leave-one-out error Q_i^2 are calculated for existing sub-domains. When the approximation is perfect ($Q_i^2 = 0$) the true volume of the sub-domain is used. Meanwhile, a poor approximation with $Q_i^2 \gg 0$ leads to increased volume.

The complete surrogate model is assembled from the $n_{\mathcal{D}}$ local PCEs as

$$Y \approx \sum_{i=0}^{n_{\mathcal{D}}} \sum_{\boldsymbol{\alpha}_i\in\mathcal{A}_i} \beta_{\boldsymbol{\alpha}_i} \Psi_{\boldsymbol{\alpha}_i}(\boldsymbol{\xi}) \mathbb{1}_{\mathcal{D}_i}(\boldsymbol{\xi}), \tag{9}$$

where $\mathbb{1}_{\mathcal{D}_i}(\boldsymbol{\xi})$ represents indicator function, i.e. $\mathbb{1}_{\mathcal{D}_i}(\boldsymbol{\xi}) = 1$ only if $\boldsymbol{\xi} \in \mathcal{D}_i$ and $\mathbb{1}_{\mathcal{D}_i}(\boldsymbol{\xi}) = 0$ otherwise. In other words, to approximate the original model at any point, it suffices to determine the one relevant sub-domain and use the corresponding local PCE. Notet that, each local PCE has its own set of basis functions $\mathcal{A}_i$ and corresponding coefficients $\beta_{\boldsymbol{\alpha}_i}$, which can be obtained by any model-selection algorithm. In this paper the OLS and LAR algorithms are employed, but generally any non-intrusive technique can be used.

4 NUMERICAL EXAMPLE

The PCE is constructed using the UQPy package [9] and for the comparison, the original implementation of the recently proposed stochastic spectral embedding (SSE) [6] is used from the UQLab package [5]. To compare both methods with a single global PCE, the relative mean squared errors ϵ are calculated for all three approximations $\tilde{f}$ on a validation set containing a large pool of 10^6 integration points generated by crude Monte Carlo according to:

$$\epsilon(\boldsymbol{X}) \coloneqq \frac{\mathbb{E}\Big[\big(f(\boldsymbol{X}) - \tilde{f}(\boldsymbol{X})\big)^2\Big]}{\mathbb{D}\Big[f(\boldsymbol{X})\Big]}, \tag{10}$$

where $\mathbb{E}[]$ and $\mathbb{D}[]$ are the mean value and variance operators, respectively. To show representative results of the proposed DAL-PCE algorithm, the calculations were repeated 100 times, and the same settings of the algorithm were selected as follows: maximum local polynomial degree $p = 2$, minimum number of samples for local PCE construction $n_{\text{sim}} = 1.5P$ and $\boldsymbol{\beta}$ are obtained by LAR and OLS algorithm. Minimum number of samples in sub-domains required to justify an expansions for SSE was set identically to DAL-PCE and polynomial order is adaptively selected in the range $p \in [2, 6]$. Since the SSE is not a sequential approach, the presented results were obtained for 10 discrete sample sets of increasing size.

The numerical example is a simple 1D function [6] that is difficult to approximate with PCE due to the third, highly nonlinear "exp" term:

$$f(X) = -X + 0.1\sin(30X) + \exp(-(50(X - 0.65))^2), \quad X \sim \mathcal{U}[0, 1]. \tag{11}$$

The poor performance of a single global PCE learned from 200 samples is depicted by the blue line in Fig. 1c where it is clear that a single global PCE is not able to accurately approximate the function even for a high number of samples and high maximum polynomial order adaptively increased as $p \in [5, 25]$. This function was originally developed to demonstrate the efficiency of the SSE based on domain decomposition and thus it is a natural choice for comparison of the proposed method and SSE.

Fig. 1a-b show a typical realization of the DAL-PCE where the algorithm sequentially decomposes the domain and adds additional samples to the ED. Specifically shown are the 4th and 11th iterations. The boundaries of sub-domains are represented by blue vertical lines and red dots show the positions of samples in the ED. Once the algorithm discovers the highly nonlinear region (the steep peak caused by exp), it it took seven iterations for the algorithm to divide the sub-domain containing the important peak and add samples to sufficiently describe the original model and create local low-order PCEs.

Of course, these figures show only one realization of the algorithm and the decomposition is dependent on the initial ED. Therefore, it is necessary to repeat the algorithm many times with random initial ED to assess convergence. Fig. 1d shows convergence of the error ϵ from 100 repeated trials. The single global PCE is unable to accurately approximate the original function even when using high p and thus the ϵ does not converge, as expected. Both methods based on domain decomposition (DAL-PCE and SSE) achieve great accuracy already for 200 samples. However, the DAL-PCE consistently has 1–2 orders of magnitude higher accuracy than SSE for the given number of samples. Although active learning might lead to lower accuracy (higher ϵ) initially (for small $n_{\text{sim}} = 10$–20) as it is dominated by exploration, it rapidly improves once it identifies important features and begins to favor exploitation.

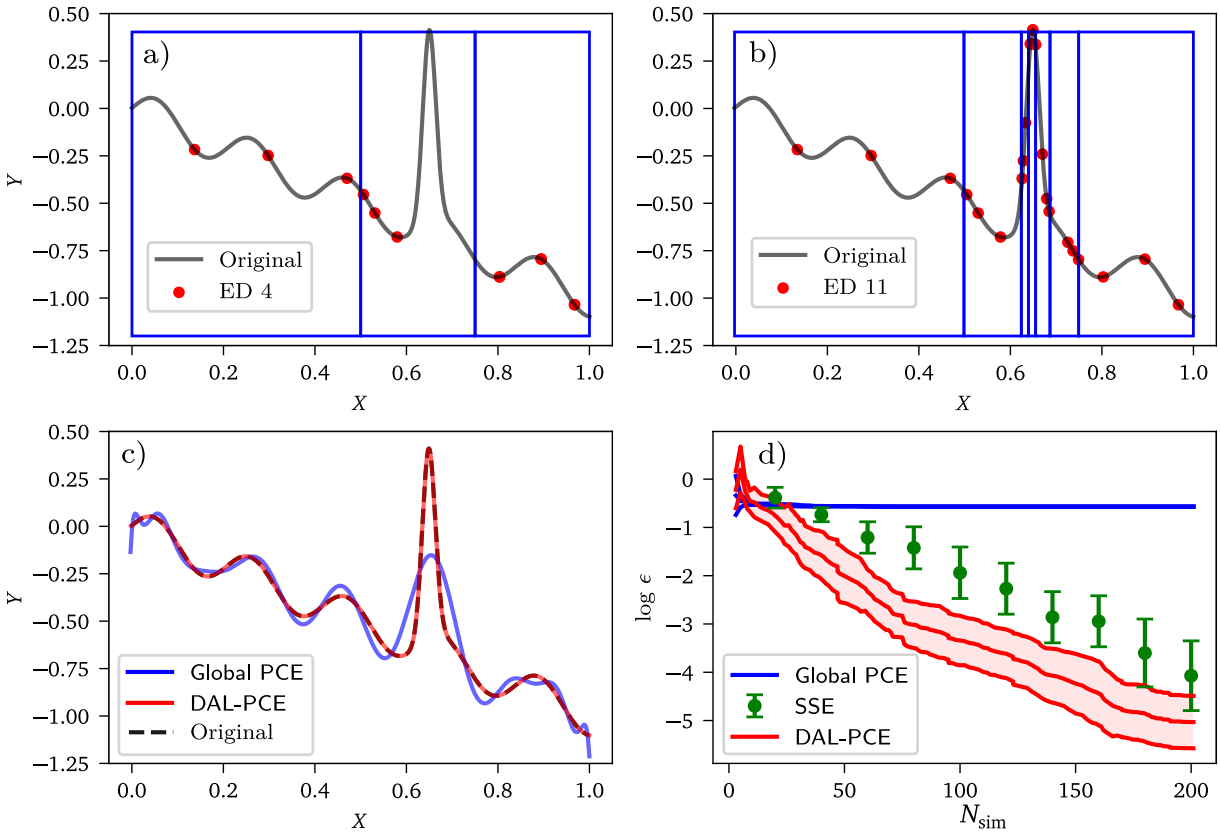

Figure 1: (a), (b) The adapted domain and ED before (iteration 4) and after (iteration 11) exploration and discovery of the exponential part of the mathematical model. (c) Final surrogate models from global PCE and DAL-PCE. (d) Convergence plot comparing the mean square error for global PCE SSE, and DAL-PCE. The convergence plots for Global PCE and DAL-PCE show continuous mean value $\pm\sigma$ intervals from 100 repeated trials, while those for SSE are plotted for several discrete ED sizes.

5 CONCLUSIONS

The paper presents a novel methodology for a decomposition of the design domain and construction of localized PCEs based on active learning. The proposed DAL-PCE method leads to very efficient construction of approximation of highly non-linear functions thanks to (i) sequential decomposition of the design domain and (ii) adaptive sequential sampling in the subdomains. Both tasks are accomplished by employing a generalized version of the recently proposed criterion, which balances between exploration and exploitation within an active learning setting. The DAL-PCE was compared to a recently proposed SSE developed for similar problems and it clearly shows the benefits of active learning in the task of a decomposition as well as in extending of ED. Further work will be focused on application of the DAL-PCE to more complicated examples with higher number of input random variables and examples with discontinuities.

ACKNOWLEDGMENT

The first author would like to express his thanks for the support provided by the Technology Agency of the Czech Republic under Project No. TM04000012. The third and last authors acknowledge the support provided by the Czech Science Foundation under project No. 22-06684K.

REFERENCES

[1] Géraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, 2011.

[2] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–451, 2004.

[3] Roger G. Ghanem and Pol D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer New York, 1991.

[4] Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008.

[5] Stefano Marelli and Bruno Sudret. UQLab: A framework for uncertainty quantification in Matlab. In *Vulnerability, Uncertainty, and Risk*, pages 2554–2563, 2014.

[6] Stefano Marelli, P.-R. Wagner, Christos Lataniotis, and Bruno Sudret. Stochastic Spectral Embedding. *International Journal for Uncertainty Quantification*, 11(2):25–47, 2021.

[7] Lukáš Novák. On distribution-based global sensitivity analysis by polynomial chaos expansion. *Computers & Structures*, 267:106808, 2022.

[8] Lukáš Novák, Miroslav Vořechovský, Václav Sadílek, and Michael D. Shields. Variance-based adaptive sequential sampling for polynomial chaos expansion. *Computer Methods in Applied Mechanics and Engineering*, 386:114105, 2021.

[9] Audrey Olivier, Dimitrios Giovanis, BS Aakash, Mohit Chauhan, Lohit Vandanapu, and Michael D Shields. Uqpy: A general purpose python package and development environment for uncertainty quantification. *Journal of Computational Science*, 47:101204, 2020.

[10] Mishal Thapa, Sameer B. Mulani, and Robert W. Walters. Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling. *Computer methods in Applied Mechanics and Engineering*, 360:112759, 2020.

[11] Joel A. Tropp and Anna C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

[12] Xiaoliang Wan and George Em Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*, 209(2):617–642, 2005.

[13] Dongbin Xiu and George Em Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[14] Jian Zhang, Weijie Gong, Xinxin Yue, Maolin Shi, and Lei Chen. Efficient reliability analysis using prediction-oriented active sparse polynomial chaos expansion. *Reliability Engineering & System Safety*, 228:108749, 2022.