SURROGATE MODELS BASED ON RADIAL BASIS FUNCTION FOR UNCERTAINTY QUANTIFICATION IN NONLINEAR AND TIME-DEPENDENT PROBLEMS.

IAGO FREITAS DE ALMEIDA

DOCTORAL THESIS IN STRUCTURES AND CIVIL CONSTRUCTION

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

FACULTY OF TECHNOLOGY

UNIVERSITY OF BRASÍLIA

UNIVERSITY OF BRASÍLIA
FACULTY OF TECHNOLOGY
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

SURROGATE MODELS BASED ON RADIAL BASIS FUNCTION FOR UNCERTAINTY QUANTIFICATION IN NONLINEAR AND TIME-DEPENDENT PROBLEMS

IAGO FREITAS DE ALMEIDA

SUPERVISOR: FRANCISCO EVANGELISTA JUNIOR, PhD.

DOCTORAL THESIS IN STRUCTURES AND CIVIL CONSTRUCTION

PUBLICATION: E.TD-01A/23
BRASÍLIA/DF: 16th JANUARY - 2023

UNIVERSITY OF BRASÍLIA

FACULTY OF TECHNOLOGY

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

SURROGATE MODELS BASED ON RADIAL BASIS FUNCTION FOR UNCERTAINTY QUANTIFICATION IN NONLINEAR AND TIME-DEPENDENT PROBLEMS

IAGO FREITAS DE ALMEIDA

DOCTORAL THESIS SUBMITTED TO THE DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING OF THE FACULTY OF TECHNOLOGY OF THE UNIVERSITY OF BRASÍLIA AS PART OF THE NECESSARY REQUIREMENTS TO OBTAIN THE DOCTOR DEGREE IN STRUCTURES AND CIVIL CONSTRUCTION.

APPROVED BY:

____

Prof. Francisco Evangelista Junior, PhD. (ENC-UnB)
(Supervisor)

____

Prof. Adriano Todorovic Fabro, PhD. (ENM-UnB)
(Internal Examiner)

____

Prof. Eduardo Toledo de Lima Junior, DSc. (UFAL)
(External Examiner)

____

Prof. Evandro Parente Junior, DSc. (UFC)
(External Examiner)

BRASÍLIA/DF, 16th JANUARY – 2023

FICHA CATALOGRÁFICA

ALMEIDA, IAGO FREITAS DE Surrogate Models Based on Radial Basis Function for Uncertainty Quantification in Nonlinear and Time-Dependent Problems [Distrito Federal] 2023. xxii, 171 p., 210 x 297 mm (ENC/FT/UnB, Doutor, Estruturas e Construção Civil, 2023). Tese de Doutorado – Universidade de Brasília. Faculdade de Tecnologia. Departamento de Engenharia Civil e Ambiental. 1. Surrogate models 2. Radial basis function 3. Uncertainty quantification 4. Time-dependent problems 5. Non-linearity I. ENC/FT/UnB II. Título (Doutor)

REFERÊNCIA BIBLIOGRÁFICA

ALMEIDA, I. F. (2023). Surrogate models based on radial basis function for uncertainty quantification in nonlinear and time-dependent problems. Tese de Doutorado em Estruturas e Construção Civil, Publicação E.TD-01A/23, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 171p.

CESSÃO DE DIREITOS

AUTOR: Iago Freitas de Almeida
TÍTULO: Surrogate models based on radial basis function for uncertainty quantification in nonlinear and time-dependent problems.

GRAU: Doutor ANO: 2023

Iago Freitas de Almeida
iago.freitas@hotmail.com

AGRADECIMENTOS

A vida é uma benção que é dada a cada um de nós para podermos crescer moralmente e espiritualmente. A cada fase de nossa vida, temos a oportunidade de buscar aprender coisas novas, principalmente voltadas para o bem, que consiste em nosso planejamento. Neste momento, chego ao final de uma longa caminhada que começou em 2017. A etapa do doutorado consistiu em uma de minhas metas acadêmicas que me possibilitou quebrar barreiras e alcançar conhecimentos que jamais imaginei que teria contato na engenharia estrutural. Primeiramente, agradeço a Deus por me permitir acordar todos os dias e buscar meus sonhos. Ele mais que ninguém sabe o quanto eu lutei para chegar tão longe. Principalmente nos momentos difíceis, como por exemplo, na época do início da pandemia de Covid-19.

Agradeço também a minha família, em especial aos meus pais, Luiz Junior e Ely por me apoiarem, mesmo distantes e morando em outra cidade, me permitiram alcançar os meus objetivos. Agradeço aos meus irmãos Igor e Raul, aos meus afilhados Inara e Luiz Neto, que de certa forma, me permitiram ter um momento de alegria durante as festividades de final de ano. Agradeço também aos amigos espirituais que me auxiliaram e continuam a me auxiliar em todas as etapas da minha vida, principalmente aqueles que me auxiliaram nos momentos em que pensei em desistir do doutorado.

Ao meu orientador, pela orientação objetiva, segura e competente. A todos os professores do Programa de Pós-graduação em Estruturas e Construção Civil que de alguma forma contribuíram para este trabalho. A todos os professores de todas as etapas da minha vida até o presente momento que me incentivaram a continuar na área acadêmica.

Aos amigos: Alexandre, Álvaro, Bruna, Carlos, Gelson, Izabel, Jéssica Siqueira, Nelson, Pedro, Thiago, Veridiana, Victor, Wallas e Yuri.

A Universidade de Brasília (UnB) e a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio e a pelo auxílio financeiro.

“Eu, porém, vos digo: Amai aos vossos inimigos, e orai pelos que vos perseguem; para que vos torneis filhos do vosso Pai que está nos céus; porque ele faz nascer o seu sol sobre maus e bons, e faz chover sobre justos e injustos. Pois, se amardes aos que vos amam, que recompensa tereis? não fazem os publicanos também o mesmo?”

(Mateus 5: 44 – 46)

RESUMO

MODELOS SUBSTITUTOS BASEADOS EM FUNÇÃO DE BASE RADIAL PARA QUANTIFICAÇÃO DE INCERTEZA EM PROBLEMAS NÃO LINEARES E DEPENDENTES DO TEMPO.

Autor: Iago Freitas de Almeida
Orientador: Francisco Evangelista Junior
Programa de Pós-graduação em Estruturas e Construção Civil
Brasília, janeiro de 2023

Esta tese propõe uma estratégia utilizando função de base radial (RBF) como modelos substitutos para quantificação de incerteza de problemas não lineares e dependentes do tempo. Os modelos substitutos com RBF são treinados para substituir as demoradas avaliações da integral que determina a taxa de liberação de energia dependente do tempo para um problema da mecânica da fratura dependente da idade e do tempo. Esses modelos substitutos também são utilizados para substituir as respostas de funções desempenho que consistem em problemas não-lineares selecionados da literatura. O desempenho dos modelos substitutos com RBF foi validado por meio de validação cruzada e por simulações de Monte Carlo que avaliam de forma exaustiva as integrais dos problemas dependentes do tempo e as funções não-lineares. Os resultados dos erros quadráticos médios e dos testes de ajuste Kolmogorov-Smirnov para as funções de densidade de probabilidade e funções de distribuição cumulativa preditas demostraram que os modelos substitutos com RBF alcançaram uma boa precisão para pequenos conjuntos de treinamento. Além disso, uma nova RBF de suporte compacto foi apresentada para processos de modelagem substituta. Os resultados preditos pelos modelos substitutos com a nova RBF demostraram uma boa eficiência para os problemas apresentados de quantificação de incerteza de funções não-lineares com conjuntos de treinamento pequenos. Os modelos substitutos com RBF apresentaram capacidades de generalização para prever milhares de dados desconhecidos obtidos de simulação de Monte Carlo e avaliar com sucesso os problemas de mecânica da fratura e das funções não-lineares.

Palavras-chave: modelos substitutos, função de base radial, quantificação de incerteza, problemas dependentes do tempo, não linearidade.

ABSTRACT

SURROGATE MODELS BASED ON RADIAL BASIS FUNCTION FOR UNCERTAINTY QUANTIFICATION IN NONLINEAR AND TIME-DEPENDENT PROBLEMS.

Author: Iago Freitas de Almeida
Supervisor: Francisco Evangelista Junior
Programa de Pós-graduação em Estruturas e Construção Civil
Brasilia, January 2023

This thesis proposes a strategy using radial basis function (RBF) as surrogate models for uncertainty quantification in nonlinear and time-dependent problems. Surrogate models with RBF are trained to replace the time-consuming evaluation of the integral that determines the time-dependent energy release rate for an age and time-dependent fracture mechanics problem. These surrogate models are also used to replace the output of performance functions that consist of nonlinear benchmark problems selected from the literature. The performance of the surrogate models with RBF was validated through cross-validation and by the Monte Carlo simulations that exhaustively evaluated the integrals of time-dependent problems and the nonlinear functions. The results of mean square errors and Kolmogorov-Smirnov goodness of fit tests for the predicted probability density functions and cumulative distribution functions showed that the RBF-based surrogate models had good accuracy for small training sets. Moreover, a new compact support RBF was presented for surrogate modeling processes. The results predicted by the surrogate models with the new RBF showed a good efficiency for the presented problems of uncertainty quantification of nonlinear functions with small training sets. Surrogate models with RBF showed generalization capabilities to predicted thousands of unknown data obtained from Monte Carlo simulation and successfully evaluated fracture mechanics and nonlinear functions problems.

Keywords: surrogate models, radial basis function, uncertainty quantification, time-dependent problems, non-linearity.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, NOMENCLATURES AND ABBREVIATIONS

a	Elliptical crack length
$(\boldsymbol{A}_m)^{-1}$	RBFs variance matrix
$b_o, b_1, \ldots, b_k$	Polynomial regression coefficients
b_w	Beam width
c	Half the crack length
$\overline{c_1}, \overline{c_2}$	Spring constants
d, A_1, A_2, p_0	Creep fit parameters
D_0	Time independent global geometric factor
e_o	Concrete plate thickness
e_r	Relative error
E	Modulus of elasticity of the material
f_c	Concrete compressive strength (MPa)
$f(x)$	Probability density function
$f_{MC}(x)$	MC^{REF} Probability density function
F, G_1, G_2, Q, H_p	Geometric functions used in the determination of K_I
F_1	Force amplitude
$F(x)$	Cumulative distribution function
$F_{MC}(x)$	Cumulative distribution function of MC^{REF}
$G(x)$	Performance function
G_I	Fracture Mode I energy release rate
G_I^R	Resistance function of G_I
G_I^S	Demand function of G_I
$h(\xi)$	Regressors of the surrogate models based on RBFs
h_w	Beam height
H_0, H_1	KS hypothesis statistical test

I	Moment of inertia
$\mathbf{I}_p$	$p \times p$ order identity matrix
$\mathbf{I}_m$	$m \times m$ order identify matrix
$\mathfrak{I}$	Training set
$J(t, t_c)$	Aging creep compliance function
J_o	Initial compliance
K_I	Stress intensity factor (fracture) in mode I
l_w	Beam span length
M_1, M_2, M_3	Elliptical crack parameters
$N(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ
m	Number of regressors to be used in the surrogate models
m_0	Oscillator body mass
M_b	Bending moment acting on the concrete plate
n_d	Number of LHS sampling intervals
p	Number of samples of the training set and/or number of samples from the Monte Carlo simulation for the validation set
P_f	Probability of failure of the structural element
$\mathbf{P}_m$	RBFs projection matrix
r_0	Displacement at which one of the springs yields
s_p	Concrete slump (cm)
S	Number of parts of the LOOCV cross validation
t	Actual time
t_c	Loading age
x, x_i	Random variable and/or input variable
y	Dependent variable
y_{MC}	Vector of MC^{REF} reference solutions
$\hat{y}$	Output response from the prediction of the surrogate models
$z(t)$	Maximum displacement response of the system

w	Uniformly distributed load
W_t	Length of the concrete plate

Greek alphabet symbols

β	Structural reliability index
β_o	size parameter of the flat top RBF
γ	Skewness or third statistical moment
γ_0	Coefficient for building μ of the G_I^S
ε	Normalized mean square error
ε_a	Square root of normalized error
$\varepsilon\ (\theta)$	Time dependent deformation
ξ	Independent variables in standard normal space or training samples used to build the surrogate models
ξ^c	Vector with central points (samples) or center of the RBFs
ρ	Support parameter or RBF length
θ	Elapsed time
κ	Shape parameter of the flat top RBF
μ	Expected or mean value
υ	Coefficient of Poisson
λ	Regularization parameter
σ	Population standard deviation
σ^2	Variance or second statistical moment
σ_b	Tensile stress due to the bending moment M_b
σ_t	Tensile stress in the concrete plate
σ_0	Constant loading stress
$\dot{\sigma}_t$	Temporal derivative of the loading function
ϕ_a	Angular slope that describes the position in front of the crack surface

ϕ_0, η_1, η_2	Regressed parameters of the creep function
Φ	Standard normal cumulative distribution function
χ	Kurtosis of fourth statistical moment
Ψ (ξ, ρ)	Radial basis functions with a given length ρ
$\mathbf{\Psi}$	RBF design matrix
$\boldsymbol{\omega}$	Vector with the regression coefficients of the surrogate models
ω_0	Oscillator parameter
Ω	Global domain

Abbreviations

ANN	Artificial Neural Networks
A-RBF	Adaptive RBF technique
CDF	Cumulative Distribution Function
CSRBF	Compactly Supported Radial Basis Function
DS+NN	Directional Sampling + Neural Networks
FORM	First Order Reliability Method
FOSM	First Order Second Moment
GA	Gaussian RBF
GCV	Generalized Cross-Validation
GEV	Generalized Extreme Value
GFEM	Generalized Finite Element Method
GSx	Greedly Sampling on the inputs
GSy	Greedly Sampling on the output
iGS	Improved Greedy Sampling on both inputs and output
JCSS	Joint Committee on Structural Safety
KRIG	Kriging

KS	Kolmogorov-Smirnov Test
LHS	Latin Hypercube Sampling
LOOCV	Leave-one-out cross-validation
MC	Monte Carlo
MC^{REF}	Reference Monte Carlo
NEXUM	Numerical, Experimental and Uncertainty Mechanics
PCE	Polynomial Chaos Expansion
PDF	Probability Density Function
PRESS	Predicted Residual Error Sum of Squares
RA	Reliability Analysis
RBF	Radial Basis Function
RBFN-RSM2	RBF Neural Network – Response Surface Method
RMSE	Root Mean Square Error
SORM	Second Order Reliability Method
SSRM	Sequential Surrogate Reliability Method
UQ	Uncertainty Quantification
WE-C2	Wendland-C2 RBF

1. INTRODUCTION

The advancement of technology in recent decades has enabled engineers to design, increasingly complex structures, such as oil platforms, airports, and bridges made of reinforced concrete or long-span steel bridges. These structures, as well as conventional structures, present uncertainties capable of triggering problems in the execution or even the collapse of the structure. These uncertainties are verified through the uncertainty quantification (UQ) process and with the use of stochastic methods widely published in the literature. The UQ process allows to obtain the entire variation spectrum of the random output variable instead of just obtaining the statistical moments. The UQ also allows obtaining the probability distribution of the variable values. In this way, through the UQ we can determine the probability density functions, the cumulative distribution functions and also the statistical moments. The uncertainties that may occur in structures may be related to the material, such as concrete, which, even though it is already a known material in structural engineering, may present uncertainties regarding the creep and relaxation phenomena. The UQ process also has a subdivision widely explored in the literature which consists of the quantitative structural reliability analysis. Structural reliability analysis is often used to obtain the probability of failure of the studied problems, being most of these techniques based on statistical moments.

The use of numerical simulation methods such as Monte Carlo simulation is a highly efficient tool to quantify uncertainty, or even structural reliability [1, 2, 3]. However, this classic simulation technique, which consists of generating a large number of random samples in numerical simulations, presents a prohibitive computational cost for situations in which the number of variables or dimensions of the problem is quite high [4, 1, 5]. The Monte Carlo simulation technique also consists of a technique that allows obtaining the probability distributions and failure probabilities of structural components. Other classic stochastic techniques from the literature that can be used in structural reliability analysis are: First Order Second Moment (FOSM) method; First Order Reliability Method (FORM) and the Second Order Reliability Method (SORM). These last three methods as well as the traditional Monte Carlo obtain the failure probability of a structural component based on limit state equations [2]. Thus, a performance function of the structural problem is necessary, which is unfeasible for many structural engineering problems in which it is not possible to define an equation explicitly that defines the problem in the molds adopted in structural reliability, and for these cases, experimental tests or exhaustive computer simulations by finite elements are used to determine the behavior of the phenomenon.

A viable solution considered in problems with uncertainty quantification (UQ) and also in structural reliability consists in the use of approximation techniques that emulate the behavior of a physical and/or mathematical system with the use of computationally cheaper functions to evaluate the output in which the characteristics and input variables of the structural component are given. These techniques, which are the surrogate models [6, 7, 8, 5] and the response surface models [9, 10, 11] are used to solve the physical system and thus deal with the computational cost of numerical simulations that often require thousands of evaluations. The response surface model has some restrictions, such as a larger number of coefficients for a larger training set. To cite some types of surrogate models, we have: Polynomial Chaos Expansion (PCE), Kriging (KRIG), Artificial Neural Networks (ANN) and Radial Basis Function (RBF). The last one will be addressed in this thesis.

Radial basis functions (RBFs) are a special class of functions initially developed by Hardy [12] to fit topographic contours [13]. These RBF functions are considered in interpolation processes [12, 14], and are also used in solving partial differential equations [15, 16, 17] and in approximating functions by regression [18, 19]. In the regression processes that are the focus of this thesis, several adaptive techniques have been used to improve the efficiency of the RBFs [20, 21, 19]. In addition, to cite some of the RBFs we have: Gaussian [22], Multiquadric [14], Wendland-C2 and Wendland-C4 [23]. When used in surrogate modeling, RBFs have a variety of advantages, such as a smaller number of samples to build the training set, even in situations where there is a high number of input variables, when compared to the response surface models. Thus, it is noted that the RBFs consist of a set of functions already known in the literature and that have several potentialities for their use. More details about RBFs and uncertainty quantification will be presented throughout this thesis.

The process of uncertainty quantification and stochastic analysis of structural elements is a field of research of the analysis group on Numerical, Experimental, and Uncertainty Mechanics (NEXUM). The most relevant publications of the group are: Evangelista Jr. and Afanador-García [24], Narváez and coauthors [25] and Evangelista Jr. and Almeida [5]. In the first publication mentioned [24], the authors applied surrogate modeling in viscoelastic media and in the last publication [5], the authors used surrogate models based on RBF to quantify the uncertainty in fracture mechanics. This last paper consists of one of the chapters of this thesis. The RBFs are considered in this thesis in regression models and are not related to neural networks.

1.1. MOTIVATION

The uncertainties are characteristics of any engineering area, and are also present in structural engineering, especially in the design, construction, and operation stages of the structures [2]. These uncertainties, which may be related to the material, the geometry of the structure, or the applied loading, are often not considered in the design of civil structures, such as buildings, bridges, concrete dams or wind turbines. However, the study of uncertainties is a broad topic that has several publications mainly in structures [26, 27, 5, 28].

In the process of uncertainty quantification of structural systems, classical simulation methods, such as Monte Carlo simulation, demand prohibitive computational time, and a viable alternative is the use of approximation methods that emulate the behavior of the structural system. These techniques are surrogate models and are used to approximate the structural problem with a lower number of numerical simulations than the classical Monte Carlo simulation. Consequently, the computational cost is drastically reduced, making it possible to obtain a solution to the problem quickly and efficiently. The surrogate model is first trained with the use of an initial training set. Then, a Monte Carlo dataset is predicted by the surrogate model in the UQ process. In order to quantify the uncertainty, several techniques with surrogate models have been highlighted, but in this thesis, we will consider the radial basis functions (RBFs) which, as mentioned above, present as an advantage, the possibility of using a smaller training set. Surrogate models based on RBFs are generally used in structural reliability processes, but in the quantification of uncertainty processes, it is still recent, especially processes that involve analysis of probability density functions. Thus, one of the objectives of this thesis is to analyze the uncertainty present in the cumulative distribution functions. The choose of cumulative distribution functions is because is possible to observe better tail errors and also a better visualization into structural reliability problems. The CDFs allow us to perform a qualitative analysis of the output response. The quantitative analysis of the results is generally given by methods based on distances from the CDFs (e.g., Kolmogorov-Smirnov test). The cumulative distribution functions also have a scarcity of publications focused on uncertainty quantification.

The construction of the training sets is another important point in the surrogate modeling process. The training sets are known inputs of the phenomenon or are built using random sampling and the determination of the number of samples considered for these sets is crucial for good training efficiency and good performance of the surrogate models. A training set with many samples may generate a good surrogate, but it demands a higher computational cost, as it requires a greater number of numerical evaluations of the physical model and can lead to overfitting. On the other hand, a training set with a few samples can cause a decrease in the quality of the predictions and consequently a bad generalization of the surrogate model. Thus, it is important to build a training set in which an adequate number of samples must be used to train the surrogate model. This becomes more important for the case in which the number of input variables in the physical and/or mathematical model is high, as one of the examples studied in this thesis presents ten random variables. The definition of sample size for training is given through a convergence test. The initial sample size that starts the convergence test is defined in two ways. The first way considers the initial sample size for the convergence test to be equal to the sample set size adopted in the examples in the literature. Then, the convergence test is performed to obtain the minimum sample size to be used by the surrogate model. In case of the examples do not present a minimum size based on the literature, we adopted the initial sample size for the convergence test according to the response surface methodology (second way).

Among the many problems that are studied in structural engineering, nonlinearity and time-dependent phenomena consist of situations that present variability in the output response [29, 30, 31]. This variability is related to several factors, such as the material properties, the characteristics of the applied loads, the geometry of the structural element, the possible conditions of execution and service of the structures, and even related to the aging of the material. Thus, in this thesis, it was decided to address the problems related to aging in viscoelastic media in fracture mechanics, since there are not many works with surrogate models and time-dependent uncertainty quantification. To determine the uncertainty of nonlinear phenomena, we chose to select performance functions from the literature. These functions, although consisting of simple examples, are reported in the literature in structural reliability analysis and are examples capable of being reproduced. The functions will be presented later. Another important point is that these functions are rarely studied in UQ processes, mainly regarding the behavior of probability density functions and cumulative distribution functions.

1.2. OBJECTIVES

1.2.1. General Objective

The thesis aims to develop efficient surrogate models based on radial basis functions (RBFs) to accurately quantify the uncertainty of nonlinear problems and age and time-dependent fracture mechanics. The surrogate models are trained to replace the physical and/or mathematical model that predicts the output response of the studied problems.

1.2.2. Specific Objectives

The specific objectives are:

- Develop surrogate models with very few samples for the training set. The training sets are built with Latin hypercube sampling.
- Build surrogate models with the minimization of normalized errors. The normalized errors are: the normalized mean square error, the square root of the normalized mean square error, and the root mean square error.
- Training and optimizing the parameters of the RBF. The training, parameter optimization and validation process is performed using the minimization of the predicted residual error sum of squares and also with the minimization of the generalized cross-validation.

- Develop a strategy with surrogate models based on RBF to quantify uncertainty in time-dependent problems. The uncertainty quantification process is performed for a problem of fracture mechanics in viscoelastic media.
- Develop a strategy with surrogate models to quantify uncertainty in nonlinear problems. In this step, a new radial basis compact support function is used in structural reliability processes derived from the uncertainty quantification process.
- Determining the parameters of the RBFs using a genetic algorithm and building regularized regression models. The shape parameters of the RBFs and the regularization parameter are determined with a genetic algorithm.
- Quantify the statistical moments of time-dependent problems using surrogate models based on RBF.

1.3. CONTRIBUTIONS

- Determination and quantification of the uncertainty of complex problems with multiple variables in problems in fracture mechanics in viscoelastic media with very few samples to build the training set. The analysis of fracture mechanics in viscoelastic media has limited publications for uncertainty quantification.
- Uncertainty quantification of nonlinear problems with strategies for building surrogate models based on RBF. Publications are limited mainly to uncertainty quantification of cumulative distribution functions.
- Development of a new compact support RBF function for uncertainty quantification processes. This contribution consists of an unprecedented RBF function that is one of the main contributions of the thesis.
- Determination of a strategy for surrogate models with regularized regression and optimization of the parameters of the RBFs with a genetic algorithm. Generally, RBFs are a little used with regularized regression in surrogate models and uncertainty quantification process.
- The state of art has limited models and techniques that use few samples to determine uncertainty.
- The use of RBFs in the construction of surrogate models for uncertainty quantification is limited in the literature. The thesis will contribute to the formulation and implementation of a methodology using RBF that provides flexibility in the response surface with a reduced number of samples.

1.4. SCIENTIFIC PRODUCTION

The thesis consists of four papers that are organized to be published opportunely. The first paper of this thesis has been published in the following journal article:

- **1th journal paper:** Evangelista Jr., F. and Almeida, I. F. "Machine learning RBF-based surrogate models for uncertainty quantification of age and time-dependent fracture mechanics". Engineering Fracture Mechanics, vol. 258, pp.: 108037, 2021. Doi: https://doi.org/10.1016/j.engfracmech.2021.108037.

The following paper is being prepared to be submitted:

- **2th journal paper**: Evangelista Jr., F. and Almeida, I. F. "A novel compactly support radial basis function as surrogate model in structural reliability analysis using very few training samples". (Under review by the supervisor)

This thesis also presents other papers that are going to be published in the future:

- **3th journal paper**: Evangelista Jr., F. and Almeida, I. F. "Surrogate modeling in uncertainty quantification of age and time-dependent statistical moments of a 3D structural component". (Under review by the supervisor)

- **4th journal paper**: "Surrogate models based on RBF with regularization and genetic algorithm in uncertainty quantification process". (Under review by the supervisor)

1.5. THESIS LAYOUT

This thesis is a series of four papers that are going to be published opportunely. The chapter one presents a brief introduction regarding what is covered in the papers and contains the respective objectives, motivation, contributions of this thesis and the scientific production.

The second chapter is the first article and addresses the uncertainty quantification process of a concrete structural component for age and time dependent fracture mechanics problem. The problem studied consists of determining the uncertainty of the energy release rate from a plate with an initial crack. This problem is analyzed using surrogate models and the response surface method through the construction of cumulative distribution functions. The use of the response surface method is associated only with the Chapter 02, and in the rest of the chapters the response surface methodology is not considered.

The third chapter deals with a process of uncertainty quantification of the statistical moments of the structural component presented in the second chapter for only the surrogate model based on RBF of the Wendland-C2 type. Thus, this chapter aims to continue the second chapter, adding new situations dependent on time and age. The moments studied are: mean, kurtosis and skewness. In the case of the second statistical moment, we chose to study the standard deviation instead of the variance.

The fourth chapter contains adaptive surrogate models to analyze the reliability index and the failure probability of pre-selected problems from the literature. The surrogate models are based on RBFs, and a new compact support RBF is also introduced. In this chapter, it is important to be mentioned that the structural reliability parameters are obtained from the uncertainty quantification process, through the probability density functions.

The fifth chapter consists of an uncertainty quantification with surrogate models of the RBF type with regularization and genetic algorithm. The ridge regression process is considered to improve the fit of the surrogate model, with both the regularization parameter and the RBF parameters being determined using a genetic algorithm. This chapter is the only one focused on the study of regularization and genetic algorithm to determine the parameters of the RBFs.

The sixth chapter summarizes all the important conclusions obtained in this thesis and makes some recommendations for future work about this thesis. In addition, an appendix is presented at the end of this thesis with additional information that are going to be considered opportunely in the future publications.

2. MACHINE LEARNING RBF-BASED SURROGATE MODELS FOR UNCERTAINTY QUANTIFICATION OF AGE AND TIME-DEPENDENT FRACTURE MECHANICS

2.1. INTRODUCTION

In many structural problems, creep and aging play important roles in the life cycle and failure behavior, especially in concrete structural components [32, 33, 34, 35, 36, 37, 38, 39, 40]. The creep phenomena are commonly studied through the classical theory of viscoelasticity, and aging models incorporate an aging dependence on the creep compliance derived from empirical [32, 41, 42, 43, 44] to more physics-based models such as the solidification theory [45, 46, 47]. Although the creep phenomena are well studied in the literature, few works report or quantify the inherent uncertainty due to random inputs, materials properties associated with the experimental results (creep compliance and aging function) or prediction models [48]. One of the main reasons is that traditional methods for uncertainty quantification (UQ) are based on the Monte Carlo (MC) simulations which become prohibitive due to the high computational cost in evaluating the convolution integral of the time-dependent problem for a large number of random variables which demands thousands or millions of realizations. Despite those difficulties, several works have employed UQ in various engineering problems [49, 50, 51, 52, 53, 25]. Specifically, on creep applications, Keitel and co-authors [49, 50] adapted variance-based global sensitivity analysis to quantify the influence of different sources of randomness to conclude that the uncertainty of creep prediction of concrete under varying stresses originates from the selection of creep models and parameter uncertainty rather than the choice of time-integration methods. More recent works also studied the creep uncertainties based on several methods such as Taylor’s expansion [29], pooling methods [54], Bayesian update [55], and polynomial chaos expansions [56, 24]. These last two works employed surrogated techniques to replace the direct methods based on time integration to cope with the computational costs of uncertainty quantification that demands thousand to million evaluations of the output response. Surrogate models (also called metamodels) and response surface models are approximation models that emulate the behavior of the physical and/or mathematical system with computationally cheaper functions to evaluate the output response given the input variables/features. In uncertainty quantification problems, those techniques have been demonstrated to be more efficient than direct probabilistic methods [56, 24]. However, surrogate models, especially the ones based on multivariate polynomial regression, have some restrictions for a high number of variables, in which a higher number of coefficients is required, and, therefore, a prohibitive training set size.

The literature on creep and aging of concrete structures also lacks the consideration of realistic failure situations where age and time-dependent behavior has been related to fracture mechanics crack front quantities, which are directly affected by the creep and aging models [57, 58, 59]. One of those quantities is the time-dependent energy release rate that is crucial to understand the service life and potential failures on structural components [60]. None of the existing papers quantifies the uncertainty of the energy release rate considered as an age and time-dependent quantity. As previously stated, surrogate models were used for the uncertainty quantification of the solely creep phenomenon [56, 24]. However, the aging phenomenon and the presence of initial cracks, and therefore the random variables related to them were not addressed. Furthermore, those surrogates used polynomial basis functions that generally require a high number of training samples to adequately fit the coefficients of the basis functions.

The aim of this paper is two-fold. Firstly, it originally develops an efficient machine learning surrogate model based on radial basis functions (RBFs) to accurately quantify the uncertainty of age and time-dependent fracture mechanics problems. The surrogate models are trained to replace a time integral that predicts the output response of the age and time-dependent energy release rate for different times and loading ages. The case study addressed is a concrete plate section with a semi-elliptic surface crack. The main contribution is to develop an innovative computational strategy to replace the time integral of the energy release rate using relatively small training sets and less intensive training methods. Secondly, it quantifies the output response uncertainty based on ten random input variables of geometry, loading, and material parameters. Another contribution here is that these parameters and their uncertainty were estimated from the fitting of an aging creep compliance function and the experimental data for concrete material. The uncertainty quantification is therefore based on the determination of probability density functions and cumulative distribution functions for the output response. The performance of the RBF surrogates is compared with traditional polynomial models, commonly used on response surface methods, and also reference solutions using the direct Monte Carlo method that exhaustively simulates the analytical integral to obtain the output response. The results show that the RBF-based surrogate models reached good accuracy with a small training set rather than the polynomial models which need a higher number of training samples. Moreover, hypothesis tests show that only the RBF surrogates were able to provide cumulative distribution functions statistically similar to the reference ones even with a 79-sample training set for the multivariate problem with ten random variables.

The paper is organized as follows: Section 2.2 presents the formulation of the surrogate models. In Section 2.3, the fracture mechanics problem is defined. Section 2.4 presents the methodology used for the analysis whereas Section 2.5 shows the results and discussion. Finally, Section 2.6 presents the conclusions and final considerations.

2.2. SURROGATE MODELS

Surrogate models can be constructed fitting (training) a generalized linear model with a training set of samples. This strategy is ideal to significantly decrease the computational time necessary to obtain the system response of complex and/or numerous time-consuming evaluations of the mathematical model. Figure 2.1 presents the flowchart using a machine learning data-driven approach for the training and application of surrogate models, either using a polynomial basis or radial basis functions (RBF). The blue boxes indicate the path of the training phase where the surrogate (orange boxes) learns from the training set. The green boxes present the path of the application of an already trained surrogate to predict new output responses for problems of UQ using random sampling in a Monte Carlo (MC) Simulation method used here as a computational algorithm that uses repeated random sampling to obtain numerical results. Random sampling can be accomplished with several techniques [61, 62]. Note that the MC simulations are performed using the surrogate and not the physical/mathematical model once that one is trained.

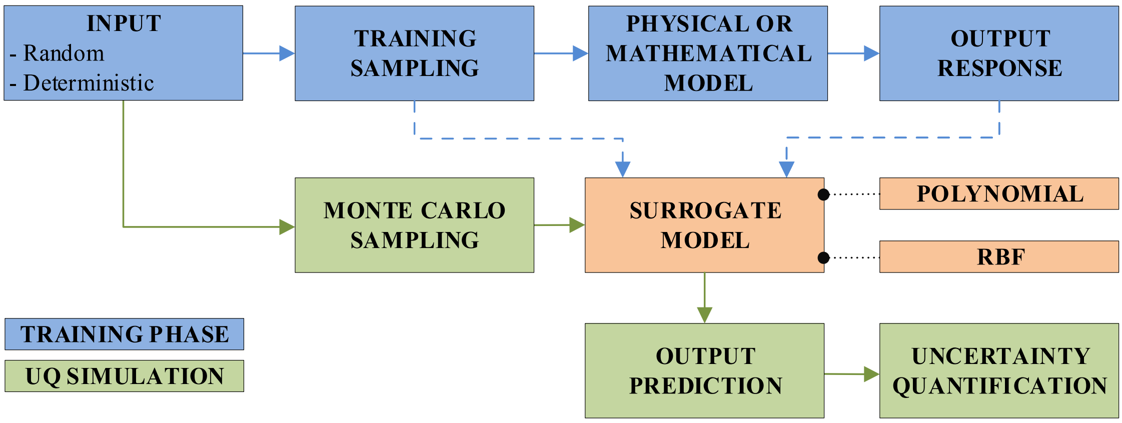


Figure 2.1 – Surrogate model training and simulation strategy flowchart.

Surrogate models can construct a prediction $\hat{y}(.)$ of a given function $y(.)$ through a general linear model:

$$\hat{y}(\xi) = \sum_{j=1}^{m} \omega_j h_j(\xi) \tag{2.1}$$

in which the m regressors $h_j(\boldsymbol{\xi})$ are functions of the inputs $\boldsymbol{\xi} \in \mathbb{R}^k$, with $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)$ and k is the number of input variables, therefore the dimension of the problem. The m unknowns are the coefficients ω_j determined through a fitting process based on a training set $\mathfrak{I} = \{(\boldsymbol{\xi}_i, y_i) | i = 1, \ldots p\}$ of data samples in which the outputs $y_i \in \mathbb{R}$ in a global domain $(\boldsymbol{\Omega})$. The quality of the fitting/training process can be accessed through a set of unseen data by the model or a cross-validation procedure that splits the data set into training and validation (unseen data during training) subsets. Different surrogate models can be constructed using different regressor functions $h_j(\xi)$ as detailed in the following subsections.

2.2.1. Response surface method

Although the literature presents several regression models able to handle the complexities of a data set, a basic approach, the so-called response surface method, is commonly developed using polynomials as the regressor functions of the general model of equation (2.1). The second-order polynomial is one of the most used in literature and is applied as a surrogate model to solve several response surface problems [61, 62]. The generalized form of a k-variate second-order polynomial is:

$$\hat{y}(\xi) = b_0 + \sum_{j=1}^{k} b_j \xi_j + \sum_{j=1}^{k} b_{jj} \xi_j^2 + \sum \sum_{i<j=2}^{k_{bl}} b_{ij} \xi_i \xi_j \tag{2.2}$$

in which the set of regressors, $h(\boldsymbol{\xi})$ in equation (2.1), has linear (ξ_i), quadratic $\left(\xi_i^2\right)$, and bilinear terms $(\xi_i \xi_j)$ with their respective coefficients b_i, b_{ii} e b_{ij}, and the independent term b_0. The total number of linear and quadratic coefficients is given, each, by the actual value of k which is the number of input variables. The number of bilinear coefficients (crossed-terms) is given by $k_{bl} = k!/2!\,(k-2)!$. A traditional least-squares procedure can be used to obtain the coefficients based on a training set $\mathfrak{I}$.

2.2.2. Proposed radial basis function surrogate models

Another type of surrogate model can be constructed using the following rewriting of Equation (2.1):

$$\hat{y}(\xi) = \sum_{i=1}^{m} \omega_i \Psi_i(\xi, \rho) \tag{2.3}$$

in which $\Psi_i(\boldsymbol{\xi}, \rho)$ are the radial basis functions (RBFs) with a given width (or support) ρ, and ω_i their respective coefficients or weights. RBF networks are known to be universal approximators for regression problems [63], and they require less intensive training and relatively smaller training set size when not used within the framework of neural networks. The RBFs considered in this work are not related to a neural network and therefore is no architecture or correction/weighting scheme (such as backpropagation). Thus, the RBF is used here as a regression model in which the weights are determined by least squares.

Equation (2.3) can be also expressed as:

$$\mathbf{y} = \mathbf{\Psi}\boldsymbol{\omega} + \mathbf{e} \tag{2.4}$$

in which $\mathbf{y}$ and $\mathbf{e}$ are the p-dimensional vectors of the training set outputs and the errors, respectively. The vector $\boldsymbol{\omega}$ is the m-dimensional vector of unknown coefficients. The $p \times m$ design matrix $\mathbf{\Psi}$ contains the response of the m regressors to the knowns p training inputs. This linear system can be expressed as:

$$\begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{Bmatrix} = \begin{bmatrix} \Psi_1\left(\left\|\xi_1 - \xi_1^c\right\|\right) & \Psi_2\left(\left\|\xi_1 - \xi_2^c\right\|\right) & \cdots & \Psi_m\left(\left\|\xi_1 - \xi_m^c\right\|\right) \\ \Psi_1\left(\left\|\xi_2 - \xi_1^c\right\|\right) & \Psi_2\left(\left\|\xi_2 - \xi_2^c\right\|\right) & \cdots & \Psi_m\left(\left\|\xi_2 - \xi_m^c\right\|\right) \\ \cdots & \cdots & \cdots & \cdots \\ \Psi_1\left(\left\|\xi_p - \xi_1^c\right\|\right) & \Psi_2\left(\left\|\xi_p - \xi_2^c\right\|\right) & \cdots & \Psi_m\left(\left\|\xi_p - \xi_m^c\right\|\right) \end{bmatrix} \begin{Bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{Bmatrix} + \begin{Bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{Bmatrix} \tag{2.5}$$

in which the regressor Ψ_i is distinguished by its center ξ_i^c which is an input variable combination from the training set $\Im$. Here, we establish the centers among the training data points to guarantee efficiency and to avoid a two-step training method commonly used when RBFs are employed within the ANN framework. In the equation, Ψ_i processes the Euclidean norm (Euclidean distance) from each training input ξ to its center ξ_i^c through the general form:

$$\Psi_i = \psi\left(\frac{\left\|\xi - \xi_i^c\right\|}{\rho}\right) \tag{2.6}$$

in which ψ is a non-linear function that monotonically decreases upon increasing values of ξ. This type of function is the so-called RBF, and the literature presents several types such as Gaussian, multiquadrics, inverse quadratic, thin-plate spline to cite a few [64, 65, 66]. However, this paper focuses attention on the two RBF used herein: Gaussian and Wendland-C2. The reason for these choices is that each one is representative of globally and locally (or compactly) supported RBF types.

The Gaussian function is the classical global RBF whose support can span all over the domain [22, 65]. There are some variations in its general equation, with the one used here given by:

$$\psi = e^{-\left(\frac{\left\|\xi - \xi^c\right\|}{\rho}\right)^2} \tag{2.7}$$

The Wendland-C2 RBF presented in [23] is classified as a compactly support RBF due to its null values outside the support region ($\|\boldsymbol{\xi} - \boldsymbol{\xi}^c\| > \rho$). Its equation is:

$$\psi = \begin{cases} \left(1 - \dfrac{\left\|\xi - \xi^c\right\|}{\rho}\right)^4 \left(1 + 4\dfrac{\left\|\xi - \xi^c\right\|}{\rho}\right), & 0 \le \dfrac{\left\|\xi - \xi^c\right\|}{\rho} \le 1 \\ 0 & \dfrac{\left\|\xi - \xi^c\right\|}{\rho} > 1 \end{cases} \tag{2.8}$$

In order to find the coefficients $\boldsymbol{\omega}$, one can minimize the cost function $E = \mathbf{e}^T\mathbf{e}$ which is the sum of the squared errors of the training set $\mathfrak{I}$. It leads to the solution provided by a least-square procedure:

$$\boldsymbol{\omega} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\mathbf{y} \tag{2.9}$$

The described strategies for the RBF surrogate models of this Section 2.2.2, as well as the polynomial response surface method of Section 2.2.1, were implemented in MATLAB™ as an in-house code called *RBFuQ*. The code has pre-processing, processing, and post-processing routines to perform the machine learning data-driven approach for the random sampling, training, and uncertainty quantification application of surrogate models, either using a polynomial basis or radial basis functions (RBF), as shown in Figure 2.1.

2.3. FRACTURE MECHANICS PROBLEM DEFINITION

Consider the structural component of Figure 2.2 that consists of a section of an initially cracked plate with width W_t and thickness e_0. The plate section is under a tensile stress σ_t and a bending moment M_b. Regarding the initial crack, it is a semi-elliptical crack surface with depth a and length $2c$. The angle ϕ_a gives the position at the crack front edge.

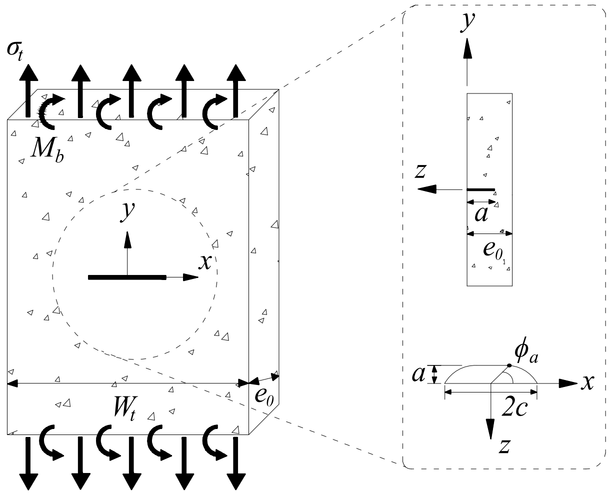


Figure 2.2 – Geometry and loading of a concrete plate section with a semi-elliptical surface crack.

The mode I stress intensity factor (K_I) for the plate section in Figure 2.2 was developed in [67]. The authors used finite element numerical solutions to construct the general solution:

$$K_I = \left[\sigma_t + H_p\sigma_b\right]\left[\sqrt{\frac{\pi a}{Q(a,c)}}\right]F\left(\frac{a}{e_0},\frac{a}{c},\frac{c}{W_t},\phi_a\right) \tag{2.10}$$

in which the tensile stress due to the bending moment is $\sigma_b = 6M_b/(W_t e_0^2)$ and the remain geometric functions $Q()$, $F()$, and H_p are defined in the Appendix A.

In the linear elastic regime, the energy release rate for a stationary crack is $G_I = K_I^2/\bar{E}_m$ with $\bar{E}_m = E$ for plane stress and $\bar{E}_m = E/(1 - \upsilon^2)$ for plane strain conditions; E is the modulus of elasticity, and υ is the Poisson ratio. Following the viscoelasticity theory and the elastic–viscoelastic correspondence principle, the age and time-dependent energy release rate can be expressed as the integral:

$$G_I(t) = D_0^2\int_0^t J(t,t_c)\dot{\sigma}_w^2(t_c)dt_c \tag{2.11}$$

in which t is the current time, t_c is the loading time or age at loading, and $J(t, t_c)$ is the aging creep compliance defined as the unit strain at time t caused by unit stress applied at age t_c. The term $\dot{\sigma}_w^2$ is the time derivative of the squared loading function. The loading function for the concrete plate section is given by:

$$\sigma_w(t) = \sigma_t(t) + H_p\sigma_b(t) \tag{2.12}$$

and the time-independent global geometric factor $D_0(.)$:

$$D_0(a,c,e_0,W_t,\phi_a) = \sqrt{\frac{\pi a}{Q(a,c)}} F\left(\frac{a}{e_0},\frac{a}{c},\frac{c}{W_t},\phi_a\right) \tag{2.13}$$

The solution of the time-dependent energy release rate integral may sometimes be performed analytically (e.g. by Laplace transforms) for the simplest cases of $J(t,\ t_c)$ and $\dot{\sigma}_w^2$, but for the majority of cases, numerical time integration strategies have to be employed [68]. Several models are currently used to describe the aging creep compliance $J(t,\ t_c)$, from empirical [32, 41, 42, 43, 44] to more physics-based models such as the solidification theory [45, 46, 47]. A basic and useful equation for the aging creep compliance is given by the double power-law (DPL) which has been used in the literature over the years [32, 69]. In this paper, the DPL is formulated as:

$$J(t,t_c) = J_0(t_c)\left[1+\phi_0 t_c^{-d}\theta^{p_0}\right] \tag{2.14}$$

in which $J_0(t_c)$ is the initial compliance at loading age t_c, ϕ_0, d, and p_0 are fitting parameters. The elapsed time θ is defined as:

$$\theta = t - t_c \tag{2.15}$$

and the initial compliance is:

$$J_0(t_c) = A_1 e^{-\eta_1 t_c} + A_2 e^{-\eta_2 t_c} \tag{2.16}$$

in which, A_1, A_2, η_1, and η_2 are fitting parameters.

The compliance parameters ϕ_0, d, p_0, A_1, A_2, η_1, and η_2 can be determined by fitting the curves $\varepsilon(\theta)/\sigma_0$ from a series of creep tests performed in different loading ages (t_c). Each of those creep tests is performed under a time constant loading stress σ_0 applied over a period of time and the time-dependent strain $\varepsilon\,(\theta)$ is monitored. Note that the age and time-dependent $J(t,t_c)$ is an increasing function of time t. Therefore, the energy release rate in Equation (2.11) also increases with t once the time derivative of the squared loading function is positive, or constant on loading age. Furthermore, $G_I(t)$ is dependent on several independent variables of geometry, loading, and material parameters whose uncertainty and randomness are propagated through the equations for different times and loading ages. Therefore, to quantify the uncertainty of the time-dependent energy release rate, sampling plans of several thousand or several million points, have to be performed, each one solving the integral Equation (2.11). Therefore, this task can be extremely costly or even prohibitive. In this way, the proposed mathematical surrogate models, described in Sections 2.2.1 and 2.2.2, can replace the physics-mathematical Equation (2.11) based solely on its data results from smaller sampling plans.

2.4. METHODOLOGY

The methodology presented here aims to train the surrogate models to predict the output response under a UQ framework considering the statistical uncertainty of the input variables of geometry, loading, and material characterization. Surrogate models using polynomial (P2) and RBF (GA and WE-C2) functions were trained to predict three different cases of loading age (t_c) and time (t): (i) t_c = 1 day and t = 11 days; (ii) t_c = 7 days and t = 17 days; and (iii) t_c = 7 days and t = 3650 days (10 years). The following subsections present the details of the employed methods.

2.4.1. Material characterization for the age and time-dependent creep compliance

The data for the material characterization used in this paper were obtained from the experimental results performed in [69]. The author conducted detailed experiments on several concrete mixtures. Different series of creep tests were performed on samples loaded at the different loading ages of 1, 3, and 8 days. Two samples for each loading age were tested with very low current time. At a loading time of 1 day, Atrushi [69] considered the current time (t) to be around 7 days; for t_c = 3 days the literature adopted t approximately equal to 10 days and for t_c = 8 days a $t \cong 13$ days was used. More details about the materials and tests can be found directly on [69].

The parameters p_0, d, ϕ_0, A_1, A_2, η_1, and η_2 were obtained through a fitting process that minimized a cost function of the differences between Equation (2.14) and the experimental measurements obtained at the creep tests for each loading age. Figure 2.3 shows the fitted $J(t, t_c)$, represented by the continuous surface, for the tested loading age's data represented by the dots. Therefore, the age-dependent creep function can be used to predict creep behavior at different loading ages.

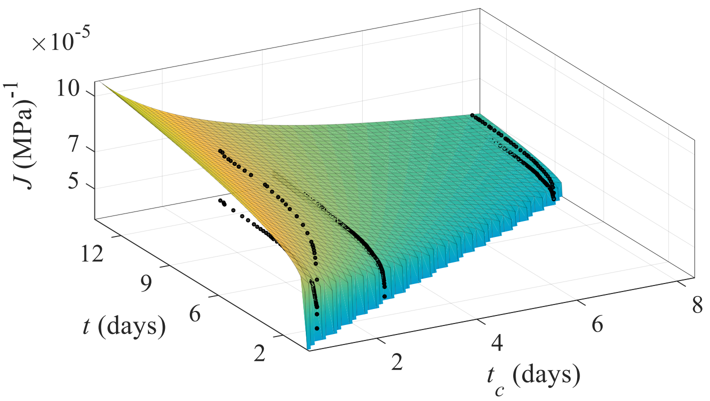


Figure 2.3 – Test data and surface fitting of creep function $J(t, t_c)$ for different loading ages t_c.

The regressed curve for the creep function presented a high fit quality indicated by the very low values of the sum of squares error (SSE = 1.31 E-08) and the root mean square error (RMSE = 5.87

E-06). The coefficient of determination (r^2) also indicated a good accuracy in which r^2 = 0.82. The formulation that describes these errors are shown in appendix A.

2.4.2. Deterministic and random variables

Considering the geometric, loading, and material input variables to determine the value of $G_I(t)$, it was reasonably to consider the geometric dimensions of the plate W_t and e_0 as deterministic variables. The two fitted power parameters η_1 and η_2 for the instantaneous compliance $J_0(t_c)$ were also considered deterministic with their values determined directly from the fitting process. For simplicity and parsimony of graphs and plots, the energy release rate was only evaluated at the position of its maximum value over the crack front (see Figure 2.2) which is reached at the angle $\phi_a = 90°0'$. These variables were defined according to literature recommendations, being based mainly on suggestions from Joint Committee on Structural Safety (JCSS) [70]. Table 2.1 lists the units and values for the deterministic variables.

Table 2.1 – Deterministic variables

Var.	Unit	Value
W_t	mm	500.00
e_0	mm	100.00
ϕ_a	degree	90°00'
η_1	-	0.4166
η_2	-	0.0136

All the other variables were considered random variables as shown in Table 2.2 with their respective mean, standard deviation, and probability density function $f(x)$. Regarding the geometric and loading variables, their mean values were chosen to render a physically sound structural problem, while their standard deviations and $f(x)$ followed the recommendations of the JCSS for loading and geometric variables [70]. The loading σ_t and σ_b were both considered a Heaviside function (step function) over time.

Table 2.2 – Random variables

Parameter	var.	unit	mean	std. dev.	$f(x)$
Geometry	a/c	-	0.40	0.08	GEV I (max)
	a/e_0	-	0.20	0.04	GEV I (max)
	c/W_t	-	0.10	0.02	Lognormal
Loading	σ_t	MPa	5.00	1.00	GEV I (max)
	σ_b	MPa	1.20	0.24	Lognormal
Material	A_1	MPa^{-1}	28.53 E-6	36.85 E-6	Lognormal
	A_2	MPa^{-1}	40.52 E-6	17.81 E-6	Lognormal
	ϕ_0	-	24.48 E-2	1.97 E-2	Lognormal
	d	-	13.06 E-2	6.25 E-2	Lognormal
	p_o	-	42.93 E-2	5.23 E-2	Lognormal

The mean, standard deviation, and probability density functions, $f(x)$, for the material parameters were estimated through a bootstrap method applied to those parameters [71, 72] when fitting the experimental data. This classical statistical technique allows the quantification of bias, variance, confidence intervals, and probability density functions of the estimates through resampling, with replacement, of the original samples [71]. The estimates in our case were the fitted values A_1, A_2, ϕ_0, d, and p_0 which were refitted 200 times through the bootstrap resamples. The mean, standard deviation, and probability density function were estimated based on the results of those 200 refits. The lognormal distribution was tested and identified as the best distribution that fitted each parameter distribution.

2.4.3. Training sampling

Based on the values in Table 2.1 and Table 2.2, a sampling strategy was employed to gather a training set $\mathfrak{I} = \{(\boldsymbol{\xi}_i, y_i) | \; i = 1, \ldots p\}$ to be used to train the surrogate models as depicted by the blue boxes of Figure 2.1. Latin hypercube sampling (LHS) was used to generate p training samples where each interval was equally divided into n_d intervals creating $(n_d)^k$ possible combinations of the k input variables. Then, a random algorithm selects p combinations among the $(n_d)^k$ with no replacement to compose the training vectors $\boldsymbol{\xi}_i$. This LHS method used here was implemented in MATLAB inside the *RBFuQ*. Following the training phase, each combination $\boldsymbol{\xi}_i$ was used to evaluate the energy release rate in Equation (2.11) to determine the y_i values to complete the training set $\mathfrak{I}$. The LHS method was chosen due to its simplicity and efficiency, and more details on the method can be found in [73, 74]. Nevertheless, any other sampling method can be used to generate the training sampling.

The determination of the number of training samples is crucial to the training efficiency and quality of the results. A bigger sample size generally provides better training and generalization of predictions but demands higher computational cost since it demands p evaluations of the expensive physical model. However, a smaller p-size training sample generally decreases the quality of the predictions due to insufficient generalization of the trained surrogate. This is especially true for a high number of variables as in the case of this paper which considers ten random variables (see Table 2.2). This paper performed a series of experiments with different p-size training sets to reach a minimal p-size of 79 samples that the RBF surrogates presented good performance, and also providing a sufficient sample number to fit a second-order polynomial. Another training set size of 180 was also employed to compare the performances among the surrogates and training sets.

2.4.4. Training, parameter optimization, and cross-validation

The surrogate models with second-order polynomial basis (P2), RBF Gaussian (GA), and RBF Wendland C-2 (WE-C2) were then trained with the provided training set. During training and cross-validation, the width parameter ρ of each RBF was determined to optimize the prediction residual error sum of squares (PRESS) of the leave-one-out cross-validation (LOOCV). This procedure was used to measure the predictive performance of the surrogates and to assure the models were not overfitted to the data during the training process. The LOOCV is a type of *S*-fold cross-validation with the number of folds equals to the number of available data ($S = p$). The method hold-out one data point (unseen to the training) to be predicted by the model trained with the other $p - 1$ data points. This procedure is then employed for the p data points and the performance quantity, known as PRESS, is calculated for each model:

$$PRESS = \sum_{i=1}^{p} \left(\hat{y}_{i,-i} - y_i \right)^2 \tag{2.17}$$

in which $\hat{y}_{i,-i}$ is the predicted response of the hold-out data point and y_i is the respective known output response. Although the LOOCV is the most time-consuming *S*-fold cross-validation, it can be the most rigorous one and recommended for small training samples [62, 63, 75].

2.4.5. Performance evaluation of uncertainty quantification prediction of PDFs and CDFs

Once the surrogate was trained and validated, the UQ was performed through the MC simulations also using the LHS sampling method. MC simulations were chosen due to easy implementation, popularity, and convenience that serves the purpose of this paper which is not centered on the efficiency of random simulation techniques. However, any other random number generator and/or sampling strategy can be used to generate the uncertainty quantification samples. This further proves that the surrogate models are so computationally inexpensive that any random simulators can be used. A set of 5E+04 samples was generated for the surrogate models to predict and provide sufficient response outputs to determine the probability density function (PDF) and cumulative distribution function (CDF) as illustrated by the green boxes in Figure 2.1. This was an extra performance procedure specifically to evaluate the quality of the models for the UQ prediction. Further performance of each surrogate model was evaluated through the comparison of their predicted values with a reference solution using the same 5E+04 samples but analytically evaluated using the integral in Equation (2.11). These results are referred to herein as the Monte Carlo reference solution (MC^{REF}).

The problem defined in Equation (2.11) consists of a convolution integral with a high computational cost, especially in high accumulated times, in which the number of steps to evaluate this integral becomes prohibitive, therefore, it is necessary to use surrogate models because they have a lower computational demand. The performance is then compared using the normalized mean square error of the PDFs, ε, to analyze the goodness of fit for the reference and surrogate-predicted PDFs as:

$$\varepsilon = \frac{\left\| f_{MC}(y) - f(\hat{y}) \right\|^2}{\left\| f_{MC}(y) - \mu_{f_{MC}} \right\|^2} \tag{2.18}$$

in which $f_{MC}(y)$ is the PDF values generated by MC^{REF}, $f(\hat{y})$ is the PDF values generated by the surrogate, $\mu_{f_{MC}}$ is the mean value of MC^{REF}, and $\|.\|$ is the Euclidian norm.

The CDFs predicted by the surrogates were also statistically tested with the ones of the reference solution using the two-sample Kolmogorov-Smirnov (KS) statistical test. The KS test variable is the maximum vertical difference between the compared CDFs:

$$KS = \max\left[F(\hat{y}) - F_{MC}(y) \right] \tag{2.19}$$

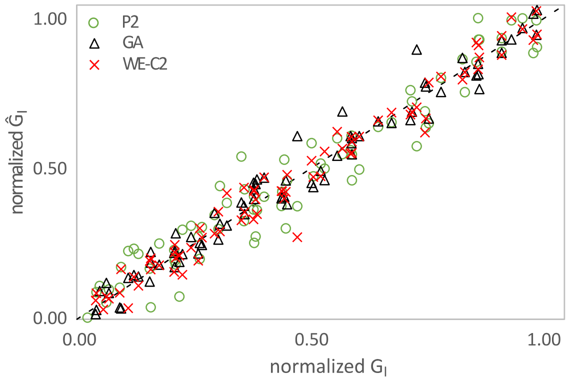
in which $F(\hat{y})$ is the CDF generated by the surrogate model predictions and $F_{MC}(y)$ is the CDF generated with the MC^{REF} values. The H_0-hypothesis consists of $F(\hat{y}) = F_{MC}(y)$ and is rejected in case the two CDFs are not drawn for the same population, and therefore not statistically similar for a given significance level.

2.5. RESULTS AND DISCUSSION

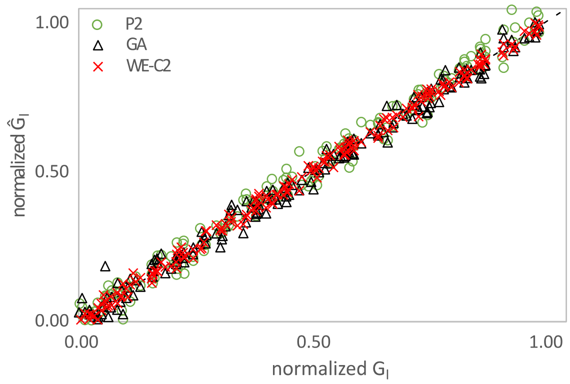
This section presents the cross-validation results for the LOOCV procedures and respective PRESS measurements. It also presents the results and discussion of the uncertainty quantification of the age and time-dependent energy release rate $G_I(t, t_c)$ in three cases: case 1 with t_c = 1 day and t = 11 days; case 2 with t_c = 7 days and t = 17 days; and case 3 with t_c = 7 days and t = 3650 days (10 years).

2.5.1. Cross-validation results

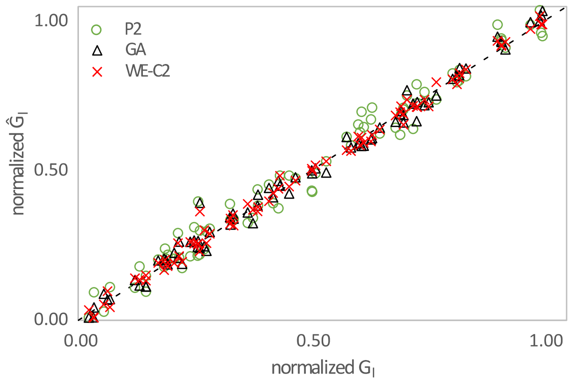
The LOOCV results are shown in Figure 2.4 for the normalized values (between 0 and 1) of the energy release rate G_I. Note that each prediction of the unseen data is performed using the surrogate trained with the remaining $p - 1$ data points. In general, the plots show the better performance and higher generalization capabilities of the RBFs models GA and WE-C2 rather than the spreader predictions of the polynomial models P2 for all cases and data set sizes. Furthermore, the plots indicated slightly higher errors in the predictions for the earlier loading age of t_c = 1 day (case 1) compared to t_c = 7 days (case 2). Case 01 has a lower fit when compared to case 02 due to the very premature loading time, which makes difficult to predict the model. However, the largest errors are found for case 3 (t_c = 7 days and t = 3650 days). The largest magnitude of error occurs for case three because it has a longer interval of elapsed time, being this behavior expected.



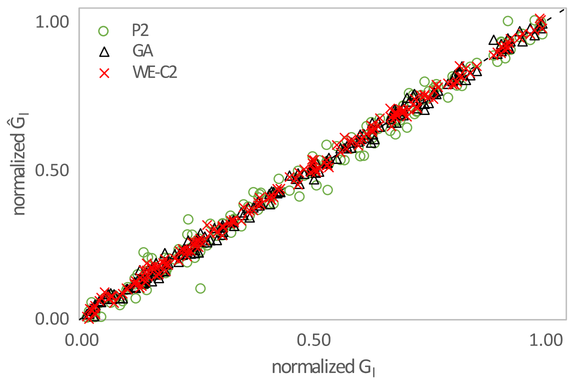
a) case 1, p = 79



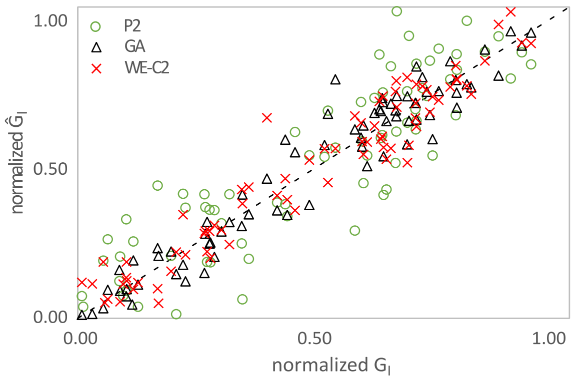
b) case 1, p = 180



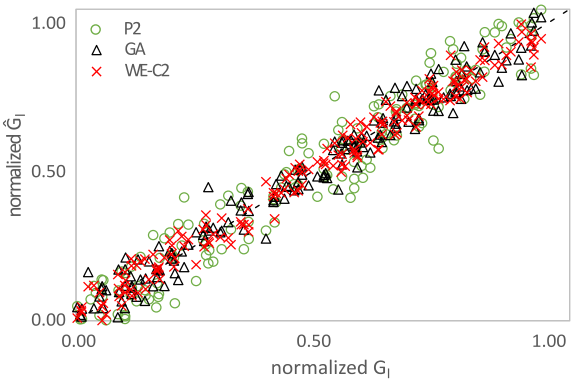
c) case 2, p = 79



d) case 2, p = 180



e) case 3, p = 79



f) case 3, p = 180

Figure 2.4 – Leave-one-out cross-validation (LOOCV) results for the three cases and two data set sizes.

The calculated PRESS values for the cross-validation are presented in Table 2.3 for the three cases and both data set sizes. The calculated PRESS for each case quantitatively confirms the observations in the previous figure. The RBF surrogates always presented better predictive performance than the polynomial surrogate for all cases and *p*-values. Specifically, the compact support surrogate model WE-C2 had the highest generalization capability among the surrogates, with the only exception for case 2 and p = 79 in which GA presented slightly less PRESS than WE-C2. As qualitatively observed in Figure 2.4, the polynomial surrogate presented high predictive errors, especially for the smaller training set. Specifically, the highest PRESS is observed for case 3 where the integral is evaluated up to the time of 3650 days. The PRESS values also show the improvement in the model generalization when larger training set sizes are used (p = 180). The performance of the models for the UQ parameters is further assessed in the next section.

Table 2.3 – LOOCV results for PRESS

p	Surrogate	Case 1	Case 2	Case 3
79	P2	0.63	0.55	2.24
	RBF (GA)	0.24	0.11	0.91
	RBF (WE-C2)	0.12	0.12	0.70
180	P2	0.17	0.14	0.92
	RBF (GA)	0.05	0.07	0.36
	RBF (WE-C2)	0.04	0.04	0.28

2.5.2. Prediction of probability density functions (PDFs) and normalized mean square errors

The reference solutions MC^{REF} for the PDFs, $f(G_I)$, using Equation (2.11), are presented in Figure 2.5 for the three cases considered here. These curves are used as a reference for the validation of the surrogate models because they are based on the analytical evaluation of the physics-based model for the time-dependent energy release rate of Equation (2.11). The curves indicated a slightly higher dispersion of the G_I distribution for the earlier loading age of t_c = 1 day (case 1) compared to t_c = 7 days (case 2). The largest dispersion is found for case 3 (t_c = 7 days and t = 3650 days) as indicated by its long-tail distribution. Indeed, the coefficient of variation for this G_I distribution was 121%. Long-tail PDFs like the one for case 3 pose a real difficulty to surrogate models since the G_I values are very dispersed over a long range of values.

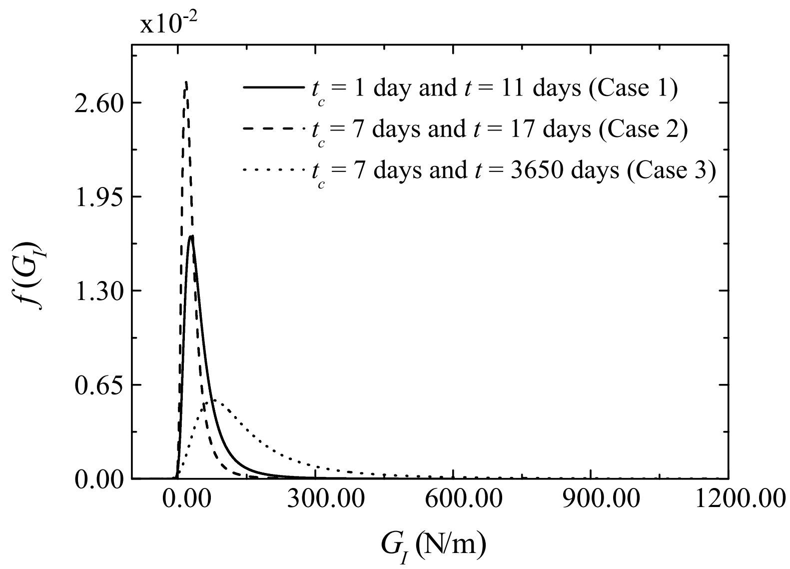


Figure 2.5 – Probability density functions (PDFs) from the solutions of MC^{REF} for the three cases.

The quantitative comparison between the PDFs curves generated from the surrogate predictions and the ones from the reference solutions is assessed by the normalized mean square error ε-values in Figure 2.6. The smaller ε-values for the RBFs surrogates (black and dark-gray colors) show that the PDFs from the RBF predictions are closer to the MC^{REF} for both training sets (79 and 180) rather than the polynomial surrogate P2 (light-gray color). This is confirmed by the smaller values of the normalized errors for both RBFs GA and WE-C2. The expansion of the training set from 79 to 180 samples significantly decreased the ε-values between the surrogates and reference solution, but still with smaller ε-values for the RBF models. The surrogate based on the RBF WE-C2 provided a better approximation in all cases and training sets. Furthermore, the WE-C2 always presented ε-values that were approximately half of the P2 values. Figure 2.6 also shows that the use of the bigger training set with 180 samples decreases the normalized mean square error by approximately five times for all surrogates. Also, note that case 3 presented the highest ε-values due to the long-tail distribution for this case as shown in Figure 2.5. Those long-tail distributions are very difficult to reproduce by a predictive model, but it was accurately predicted by the RBF surrogates.

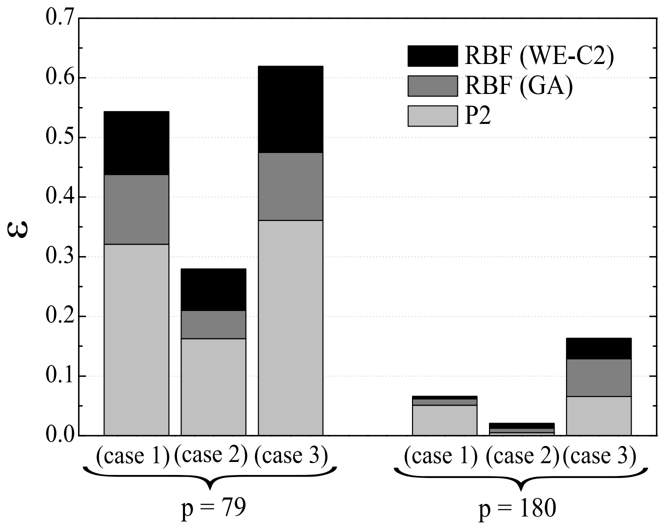
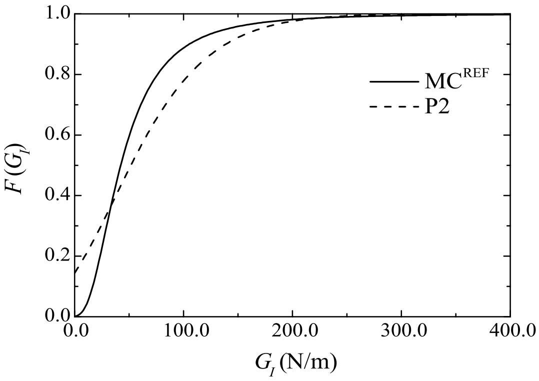


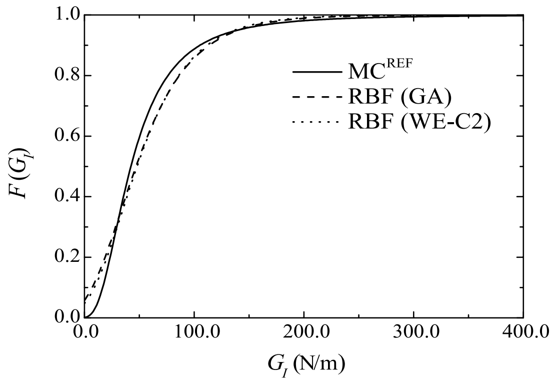
Figure 2.6 – Normalized mean square error (ε) between predicted and reference PDFs.

2.5.3. Prediction of cumulative distribution functions (CDFs) and statistical tests

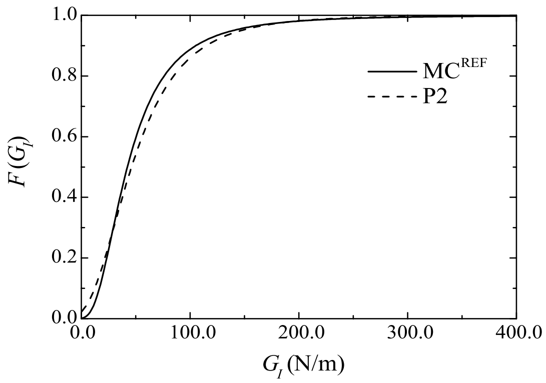
The CDFs $F(G_I)$ were constructed based on predictions of each surrogate model and from the MC reference solution for the three cases. Figure 2.7 through Figure 2.9 graphically shows, for both training sets of 79 and 180 samples, the goodness-of-fit between the generated CDFs from the surrogate predictions and reference solution. Qualitatively, Figure 2.7a, Figure 2.8a, and Figure 2.9a show that the CDFs predicted by the polynomial surrogate (P2), with the 79-training set, could not provide satisfactory results for the three considered cases. However, the CDFs generated from the predictions of the RBF surrogates GA and WE-C2 presented a qualitatively satisfactory goodness-of-fit with the reference solutions considering the 79-training set for all three cases.



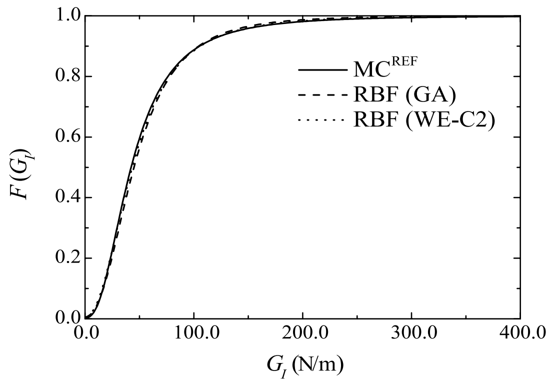
a) P2, p=79



b) RBFs, p=79



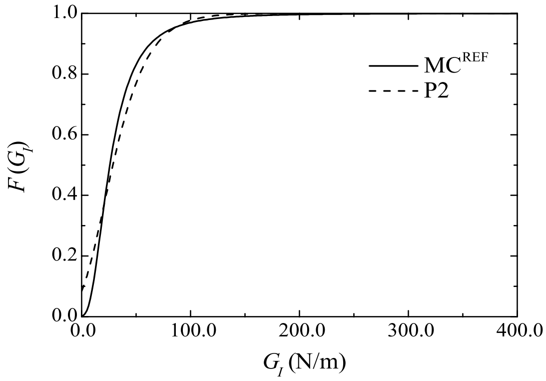
c) P2, p=180



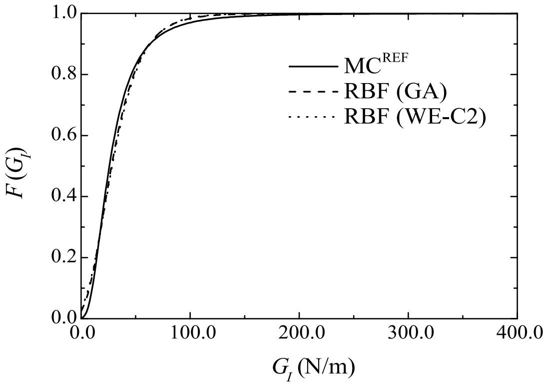
d) RBFs, p=180

Figure 2.7 – Cumulative distribution function (CDF) comparison for case 1.

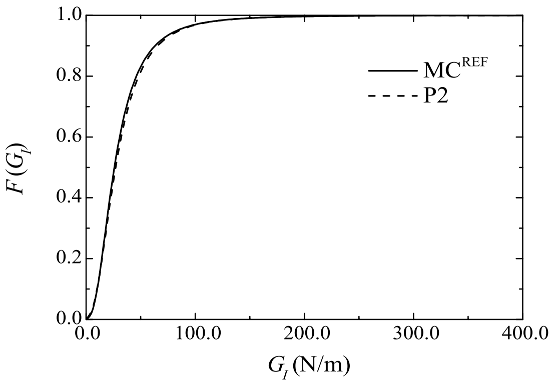
Surrogate model predictions trained with the 180-training set significantly improved the $F(G_I)$ towards the reference solution curves as shown in Figure 2.7c,d, Figure 2.8c,d, and Figure 2.9c,d. Although the approximation for the P2 models also improved with the 180- training set, the RBFs still presented better proximity with their respective reference CDFs. The plots show no visual difference between the RBFs GA and the WE-C2 for all three cases.



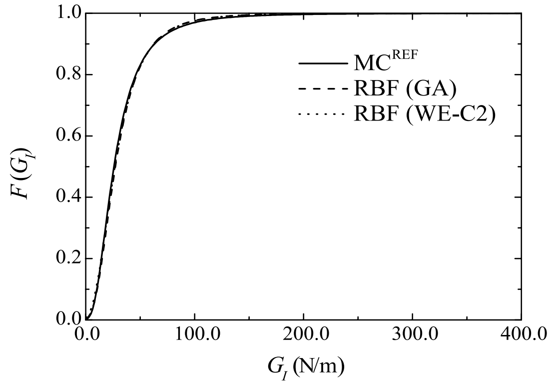
a) P2, p=79



b) RBFs, p=79



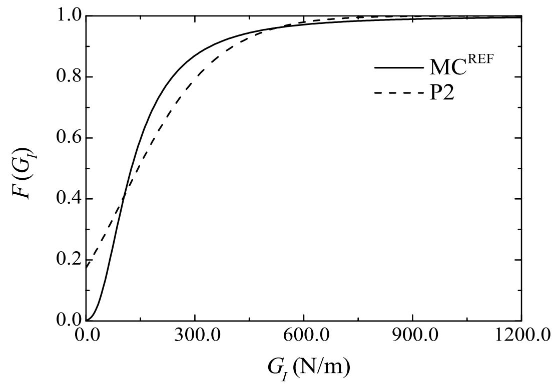
c) P2, p=180



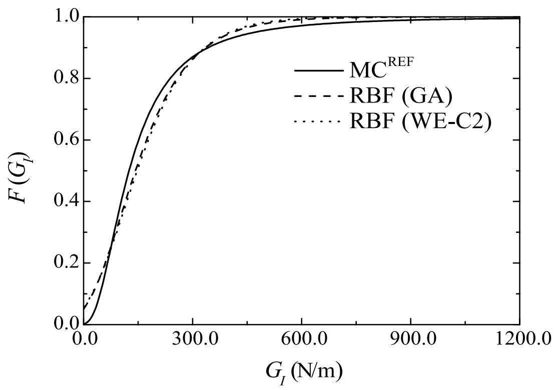
d) RBFs, p=180

Figure 2.8 – Cumulative distribution function (CDF) comparison for case 2.

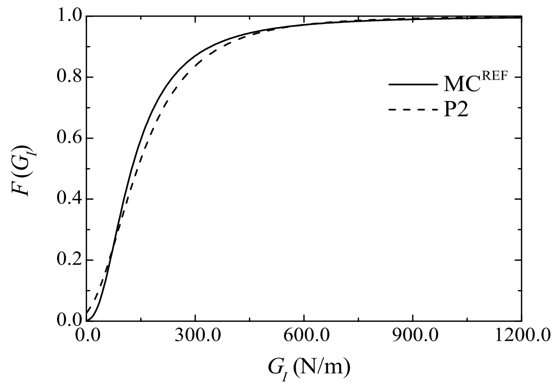
Regarding the shape of the CDFs, note that the ones for case 1 and case 3 exhibit a longer tail approaching the unit of the cumulative value. This further emphasizes extreme values for G_I in those cases, especially for case 3 in which the random nature of geometry, loading, and material properties leads to the prediction of very large uncertainty for farther prediction times from the initial loading time.



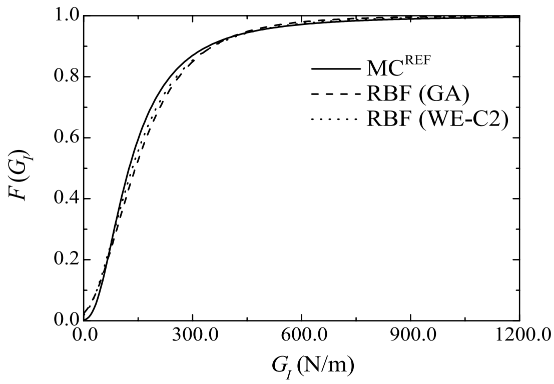
a) P2, p=79



b) RBFs, p=79



c) P2, p=180



d) RBFs, p=180

Figure 2.9 – Cumulative distribution function (CDF) comparison for case 3.

Table 2.4 shows the parameters for the hypothesis test in terms of the KS statistical test (two samples) for each surrogate and the reference solution. The table presents the KS-values and the correspondent resultant hypothesis (H_0 or H_1) for a significance level of α = 5% which is the probability of rejecting the null hypothesis when it is true. The null hypothesis (H_0) is that surrogate predictions and MC reference solutions are drawn from the same population. Therefore, the H_0 result means that the CDF determined from the surrogate predicted values is statistically similar to the CDF determined from the MC reference solution. On the other hand, the alternative hypothesis H_1 indicates the predicted CDF is not statistically similar to the reference solution. As explained earlier, the KS-value is the maximum difference between the $F(G_I)$ values of reference and surrogate solutions as described in Equation (2.19). The table shows that the RBF surrogates presented much smaller KS-values than the P2 model for the 79-training set. The polynomial predictions only presented satisfactory results for the 180-training set as shown in Table 2.4. Furthermore, the higher KS-values for the 79-training set, presented by the P2 polynomials, are reflected in the rejection of H_0. The CDFs based on the predictions of RBF GA and WE-C2 had the acceptance of H_0 for all three cases even for the smaller training set with 79 samples.

Table 2.4 – KS statistical hypothesis test results for a significance level of 5%

p	Surrogate	Case 1		Case 2		Case 3	
		KS	result	KS	result	KS	result
79	P2	0.17	H_1	0.12	H_1	0.19	H_1
	RBF (GA)	0.09	H_0	0.05	H_0	0.07	H_0
	RBF (WE-C2)	0.08	H_0	0.06	H_0	0.09	H_0
180	P2	0.06	H_0	0.03	H_0	0.07	H_0
	RBF (GA)	0.04	H_0	0.03	H_0	0.07	H_0
	RBF (WE-C2)	0.01	H_0	0.02	H_0	0.04	H_0

H_0: surrogate predictions and MC reference solution are drawn from the same population
H_1: surrogate predictions and MC reference solution are drawn from different populations

2.6. CONCLUSIONS

This paper developed a surrogate modeling strategy for the uncertainty quantification of age and time-dependent fracture mechanics problems. Specifically, it was considered a case study of the energy release rate for concrete structures under the presence of initial cracks. Three different types of surrogate models have been employed: the traditional second-order polynomial (P2) and two proposed RBF models (Gaussian and Wendland-C2). The surrogates were trained to replace the time integral to predict the output values of the energy release rate for three different cases of loading ages and elapsed times. The trained surrogate models were used to predict thousands of values from unseen data of Monte Carlo samples to the post-processing of the uncertainty quantification parameters. The uncertainty quantification was then performed by determining the probability density function (PDF) and the cumulative distribution function (CDF) of the output response considering ten random input variables of geometry, loading, and material parameters.

The cross-validation procedures indicated the better performance of both RBF surrogates for predicting unseen data values. The uncertainty quantification was able to demonstrate the effect of the uncertainty of geometry, loading, and material parameters on the probabilistic energy release rate at different loading ages and times. The uncertainty of the input variables produced spreader probability density distributions of the energy release rate for earlier loading times. Also, even longer-tail distribution was observed for farther prediction times showing the high uncertainty of the energy release rate farther in time. The normalized mean square error between the PDFs constructed from the surrogate predictions and the ones from the reference solutions showed that both RBF surrogates performed better than the P2 model for both training set sizes and all three cases of loading ages and times. Specifically, the Wendland-C2 surrogate presented approximately 1/3 of the normalized mean square error presented by the P2 surrogate when both used the smaller training set size with 79 samples. The RBF surrogates always presented better goodness-of-fit between the predicted CDFs and the reference solution, rather than the CDFs determined from the polynomial predictions. The KS statistical test confirmed that only the RBF surrogates predicted statistically similar CDFs to the reference solutions for both training set sizes and all three cases of loading ages and times considered. The P2 model could only present statistically acceptable results when trained with the bigger 180-sample training set. All these performance trends for the uncertainty quantification results were foreseen by the leave-one-out cross-validation assessment, which demonstrates this technique successfully showed the generalization capabilities of each surrogate model for the problem considered in this paper.

The results demonstrated that the proposed RBF surrogates are efficient and accurate computational strategies for large-scale engineering problems involving uncertainty quantification of time-dependent fracture problems that demand thousand to million evaluations of viscoelastic equations involving time-consuming integrals of the energy release rate. The RBF models were able to generalize well the thousands of unseen data from the Monte Carlo sampling to accurately predict the results for those integrals for a multivariate problem with ten random variables with only 79-training samples. Furthermore, the RBF surrogates were able to accurately predict long-tail probability density functions with that relatively small training set size. Although the example considered a concrete plate with a semi-elliptical surface crack, the methodology presented in this paper, as well as the RBF surrogate models, can be directly applied to different materials on other fracture problems which present the energy release rate formulated as a mapping integral.

3. SURROGATE MODELING IN UNCERTAINTY QUANTIFICATION OF AGE AND TIME-DEPENDENT STATISTICAL MOMENTS OF A 3D STRUCTURAL COMPONENT

3.1. INTRODUCTION

Most materials used in structural engineering have complex characteristics that are extremely difficult to measure, e.g. creep and aging in cementitious materials [45, 37, 39]. Several studies address the subject, focused on the construction of mathematical models that aim to represent the behavior of these materials [76, 41, 42, 43, 44, 46, 47, 36, 77]. However, the aging creep analysis in cementitious materials still requires special attention, especially in relation to uncertainty quantification [49, 50].

The uncertainty quantification (UQ) is considered by the literature to obtain the entire spectrum of variation of the output random variable. In this way, it is possible to collect the probability distribution of the values of a variable and consequently determine the probability density function or the cumulative distribution function. The UQ process generally adopted by the literature considers classical simulation models, such as the Monte Carlo method (MC), which is computationally demanding, as it requires millions of evaluations of the physical/mathematical model [78, 48, 29, 79, 52, 53, 80, 81]. The work of Motra [48], for example, verified the cyclical creep in different models and different uncertainty quantification processes. The authors consider the Monte Carlo simulation with Latin Hypercube Sampling (LHS) to determine the uncertainty parameters. The results obtained by the authors in the stochastic sensitivity analysis showed that the variability of cyclic creep strain is influenced more by the modulus of elasticity than other parameters, such as mean stress and cyclic stress amplitude [48]. Another publication that deserves attention is Criel and coauthors [29] who carried out a process of uncertainty quantification of the creep prediction models by using Taylor series. However, in the UQ process, the Taylor series strategy is limited only to estimating the statistical moments and cannot incorporate the probability distribution of the input parameters [29]. In most of the works cited, the use of classical methods such as MC becomes prohibitive, especially for problems that involve a high number of input variables.

A solution to predict the creep and aging behavior is to adopt approximation functions that emulate the physical or mathematical behavior with a lower computational cost to predict the output response [56, 82, 5]. Especially, the work of Evangelista Jr. and Almeida [5] who is a pioneer in the field of uncertainty quantification in age and time-dependent fracture mechanics using Radial

Basis Function (RBF). In this work, the authors verify the energy release rate of a concrete plate with an initial elliptical crack. The uncertainty quantification process focused on the analysis of the cumulative distribution and probability density functions for different ages of loading and time. The results predicted by the models with RBFs showed high efficiency for early and later times. This approach with RBF consists of a type of approximation function that is used in a strategy for training and validating data sets, called surrogate model. However, none of the cited works performed a process of uncertainty quantification focused on the statistical moments from the results predicted by the surrogate models already trained.

The purpose of this article is two-fold. First, it develops an efficient machine learning model with a radial basis function (RBF) to quantify the statistical moments of age and time-dependent fracture mechanics problems. The choice for the quantification of statistical moments is because they are computationally cheaper to apply when compared to the determination of the probability density functions (PDFs) or the cumulative distribution functions (CDFs). The model considers the machine learning approach developed by Evangelista Jr. and Almeida [5] in which RBF surrogate models are trained to substitute a time convolution integral that predicts the energy release rate output response for a concrete plate with an elliptical crack. This integral presents a high computational cost in high values of accumulated times in which the number of steps to evaluate the integral becomes prohibitive, being therefore necessary to use a surrogate model. The novelty and main contribution, therefore, is a process of uncertainty quantification for the statistical moments (mean, standard deviation, skewness, and kurtosis) that are obtained for the results of Evangelista Jr. and Almeida [5] and different initial loading age and elapsed time. Second, perform a reliability analysis for all cases of elapsed times by constructing a performance function and later determining the failure probability and structural reliability index. The performance of the models with RBF is obtained by comparing them with reference solutions using the direct Monte Carlo method. The reference solution was built by exhaustive simulations of the analytical integral of the output response. The results of the surrogate models showed good precision for most of the cases studied, with the exception of the moment of skewness, which requires a larger training set.

The paper is organized as follows: Section 3.2 presents the surrogate model of radial basis function. In Section 3.3, there is the definition of the 3D structural component. Section 3.4 presents the methodology, while Section 3.5 presents the assessment of quality of results. The Section 3.6 presents the analysis of time-dependent energy release rate (ERR) and Section 3.7 shows the conclusions and final considerations.

3.2. SURROGATE MODEL APPROACH

The behavior of the surrogate models based on radial basis function (blue box) is shown in Figure 3.1. This surrogate model based on RBF is composed of two steps: training and uncertainty quantification (UQ) process. The training (purple boxes) consists of the initial step in which a set of input data is used in the surrogate model. The input data are deterministic and random and will be presented in Section 3.4. The training sets are built with Latin hypercube sampling (LHS) and are used to obtain the output response of the mathematical model (purple boxes). The training set obtained in the training phase is then considered by the surrogate model (blue box) which learns from the data obtained from the training phase.

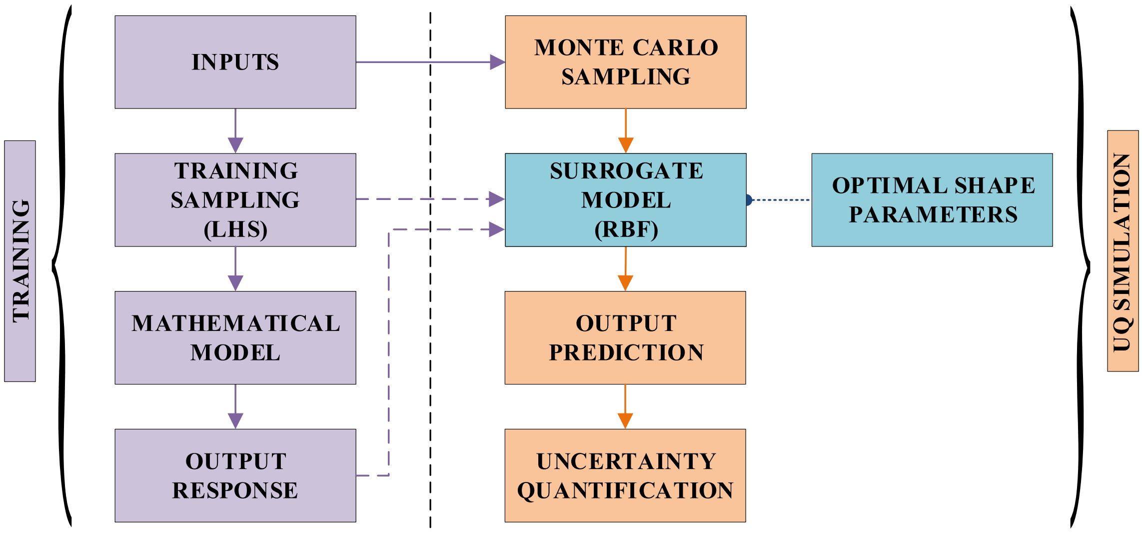


Figure 3.1 – Flowchart with strategy for training and simulating surrogate models

The uncertainty quantification process is shown in the orange boxes. In this process, the surrogate model is used to predict a Monte Carlo simulation dataset. The surrogate model already trained is then able to predict new responses and in this way is done the uncertainty quantification or the structural reliability analysis. The quantitative of the reliability analysis is a subdivision of uncertainty quantification that is also considered in this paper. The structural reliability analysis has the same procedure considered in the UQ process. Just replace the orange box in the figure with the phrase uncertainty quantification by reliability analysis. In subsequent sections, the methodology for both analyzes will be presented. The surrogate radial basis function model builds the prediction $\hat{y}(.)$ of a function $y(.)$ through the following model:

$$\hat{y}(\xi) = \sum_{i=1}^{m} \omega_i \Psi_i(\xi, \rho) \tag{3.1}$$

in which $\Psi_i(\xi, \rho)$ are the radial basis functions (RBFs) with a given support (length) ρ, and ω_i their respective weights or coefficients. The RBFs considered in this work are not related to a neural network and therefore is no architecture or correction/weighting scheme (such as backpropagation). Thus, the RBF is used here as a regression model in which the weights are determined by least squares. The above equation is also expressed in matrix format as:

$$\mathbf{y} = \mathbf{\Psi}\boldsymbol{\omega} + \mathbf{e} \tag{3.2}$$

in which $\mathbf{y}$ is the p-dimensional vector of the known outputs of the training set, $\boldsymbol{\omega} = \{\omega_1, \omega_2, \dots, \omega_m\}$ is the unknown coefficient vector and $\mathbf{e} = \{e_1, e_2, \dots, e_p\}$ is the error vector. The design matrix $\mathbf{\Psi}$ of size $p \times m$ contains the response of the m regressors to the known training inputs p:

$$\mathbf{\Psi} = \begin{bmatrix} \Psi_1\left(\dfrac{\|\xi_1 - \xi_1^c\|}{\rho}\right) & \Psi_2\left(\dfrac{\|\xi_1 - \xi_2^c\|}{\rho}\right) & \cdots & \Psi_m\left(\dfrac{\|\xi_1 - \xi_m^c\|}{\rho}\right) \\ \Psi_1\left(\dfrac{\|\xi_2 - \xi_1^c\|}{\rho}\right) & \Psi_2\left(\dfrac{\|\xi_2 - \xi_2^c\|}{\rho}\right) & \cdots & \Psi_m\left(\dfrac{\|\xi_2 - \xi_m^c\|}{\rho}\right) \\ \cdots & \cdots & \cdots & \cdots \\ \Psi_1\left(\dfrac{\|\xi_p - \xi_1^c\|}{\rho}\right) & \Psi_2\left(\dfrac{\|\xi_p - \xi_2^c\|}{\rho}\right) & \cdots & \Psi_m\left(\dfrac{\|\xi_p - \xi_m^c\|}{\rho}\right) \end{bmatrix} \tag{3.3}$$

in which the regressor $\boldsymbol{\psi}_i$ process the Euclidean norm (Euclidean distance) of each training input $\boldsymbol{\xi}$ to its center $\boldsymbol{\xi}_i^c$ which is a combination of input variables from the training set $\mathfrak{I}$. The centers are adopted as their own training sets to ensure good efficiency ($m = p$). The training inputs were normalized following a standard normal distribution (mean zero and standard deviation equal to 1). In this paper, the focus is on RBFs classified as locally supported RBF with the RBF of type Wendland-C2 being adopted as a representative of this class. The choice of Wendland-C2 is due to its good performance in the uncertainty quantification process shown in the paper of Evangelista Jr. and Almeida [5]. The Wendland-C2 RBF presented in [23] is classified as a compact support RBF because its null values are outside the support region ($\|\boldsymbol{\xi} - \boldsymbol{\xi}^c\| > \rho$). The equation is:

$$\psi = \begin{cases} \left(1 - \dfrac{\|\xi - \xi^c\|}{\rho}\right)^4 \left(1 + 4\dfrac{\|\xi - \xi^c\|}{\rho}\right), & 0 \le \dfrac{\|\xi - \xi^c\|}{\rho} \le 1 \\ 0 & \dfrac{\|\xi - \xi^c\|}{\rho} > 1 \end{cases} \tag{3.4}$$

The range of values for the support parameter (ρ) is $0.0001 \le \rho \le 50$. This parameter was determined through several numerical experiments. The vector of unknown coefficients $\boldsymbol{\omega}$ is then obtained by the least squares process by minimizing the function $E = \mathbf{e}^T\mathbf{e}$. The function E is the sum of squares of the errors of the training set $\Im$ and the vector $\boldsymbol{\omega}$ is defined as:

$$\boldsymbol{\omega} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\mathbf{y} \tag{3.5}$$

The procedures used in the surrogate models with RBF in this section were implemented in MATLAB™ as in-house code denominated *RBFuQ* described in Evangelista Jr. and Almeida [5]. The code presents pre-processing, processing and post-processing routines to perform random sampling, training and application of uncertainty quantification and reliability analysis of surrogate models as shown in Figure 3.1. Furthermore, once the model is trained, it can simulate MC to evaluate the statistical moments that are detailed in Section 3.4.5.

3.3. DEFINITION OF THE 3D STRUCTURAL COMPONENT

A section of a concrete plate with an initial crack has length W_t and thickness e_0 according to Figure 3.2. The section of the structural component is subjected to a tensile stress σ_t and a bending moment M_b. The initial crack in the plate consists of a semi-elliptical surface with depth a and length $2c$. In Figure 3.2 it is possible to observe that the angle ϕ_a provides the position in front of the crack.

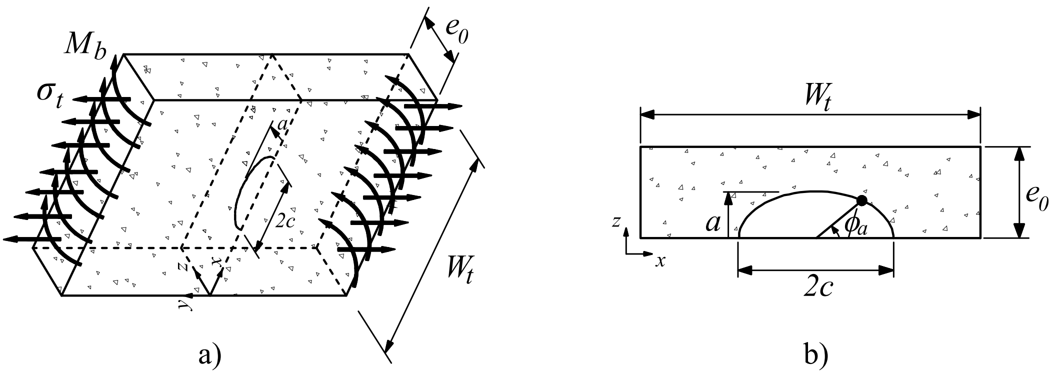


Figure 3.2 – Geometry of a concrete plate section with a semi-elliptic crack and cross-section detail (Adapted from Biglari and coauthors [83]).

The structural component of Figure 3.2 was initially studied by Evangelista Jr. and Almeida [5] with the use of numerical solutions of finite elements obtained by Newman and Raju [67] for fracture mode I, with the fracture intensity factor, or stress intensity factor (K_I) defined as:

$$K_I = \left[\sigma_t + H_p\sigma_b\right]\left[\sqrt{\frac{\pi a}{Q(a,c)}}\right]F\left(\frac{a}{e_0},\frac{a}{c},\frac{c}{W_t},\phi_a\right) \tag{3.6}$$

in the tensile stress due the bending moment is $\sigma_b = 6M_b/(W_t e_0^2)$. The other geometric functions $Q()$ and $F()$ are:

$$Q(a,c) = 1 + 1.464\left(\frac{a}{c}\right)^{1.65} \quad \text{for} \quad \frac{a}{c} \le 1$$

$$Q(a,c) = 1 + 1.464\left(\frac{c}{a}\right)^{1.65} \quad \text{for} \quad \frac{a}{c} > 1$$

$$F\left(\frac{a}{e_0}, \frac{a}{c}, \frac{c}{W_t}, \phi_a\right) = \left[M_1 + M_2\left(\frac{a}{e_0}\right)^2 + M_3\left(\frac{a}{e_0}\right)^4\right] f_{\phi_a} f_w f_g$$

$$M_1 = 1.13 - 0.09\left(\frac{a}{c}\right)$$

$$M_2 = -0.54 + \frac{0.89}{0.2 + (a/c)} \tag{3.7a-i}$$

$$M_3 = 0.5 - \frac{1.0}{0.65 + (a/c)} + 14\left(1.0 - \frac{a}{c}\right)^{24}$$

$$f_{\phi_a} = \left[\left(\frac{a}{c}\right)^2 \cos^2\phi_a + \sin^2\phi_a\right]^{1/4}$$

$$f_g = 1 + \left[0.1 + 0.35\left(\frac{a}{e_0}\right)^2\right](1 - \sin\phi_a)^2$$

$$f_w = \left[\sec\left(\frac{\pi c}{W_t}\right)\sqrt{\frac{a}{e_0}}\right]^{1/2}$$

The geometric function is from Newman and Raju [67] and were obtained by a finite element approximation.

The energy release rate for a stationary crack in a linear elastic regime is defined as $G_I = K_I^2/\bar{E}_m$. The term $\bar{E}_m = E$ for the plane stress with E as the modulus of elasticity. In the plane strain, $\bar{E}_m = E/(1-\upsilon^2)$ with the term υ as the Poisson ratio. The theory of viscoelasticity and the elastic-viscoelastic correspondence principle define the age and time-dependent energy release rate as an integral:

$$G_I(t) = D_0^2 \int_0^t J(t, t_c)\dot{\sigma}_w^2(t_c)\,dt_c \tag{3.8}$$

in which t is the actual time, t_c is the loading age or age loading, and $J(t, t_c)$ is the aging creep compliance defined as the unit strain at time t caused by the unit stress applied at age t_c. The integral of the energy release rate given in Equation (3.8) is a convolution integral that demands a high processing time. For each elapsed time to be verified by the integral, j integration time steps are required. Consequently, the longer the elapsed time, the more processing time is required to solve the integral. This presents a greater complexity when the number of Monte Carlo simulation samples is one million, as the computational time becomes very high, and it may take hours to solve this integral. In this case, the integral will have 1E+06 j^2 integration steps. Therefore, it is important to use the surrogate model to approximate the response of this integral. The term $\dot{\sigma}_w^2$ is the time derivative of the squared loading function. The loading function for the concrete plate section is given by:

$$\sigma_w(t) = \sigma_t(t) + H_p \sigma_b\left(t\right) \tag{3.9}$$

The time-independent global geometric factor $D_0(.)$:

$$D_0(a, c, e_0, W_t, \phi_a) = \sqrt{\frac{\pi a}{Q(a,c)}} F\left(\frac{a}{e_0}, \frac{a}{c}, \frac{c}{W_t}, \phi_a\right) \tag{3.10}$$

The auxiliary functions also obtained by Newman and Raju [67] which are given in Equation (3.9) are:

$$\begin{aligned}
&H_p = H_1 + (H_2 - H_1)\sin^p \phi_a \\
&p = 0.20 + \left(\frac{a}{c}\right) + 0.60\left(\frac{a}{e_0}\right) \\
&H_1 = 1 - 0.34\left(\frac{a}{e_0}\right) - 0.11\left(\frac{a}{c}\right)\left(\frac{a}{e_0}\right) \\
&H_2 = 1 + G_1\left(\frac{a}{e_0}\right) + G_2\left(\frac{a}{e_0}\right)^2 \\
&G_1 = -1.22 - 0.12\left(\frac{a}{c}\right) \\
&G_2 = 0.55 - 1.05\left(\frac{a}{c}\right)^{3/4} + 0.47\left(\frac{a}{c}\right)^{3/2}
\end{aligned} \tag{3.11a-f}$$

The time-dependent energy release rate integral solution is defined according to Evangelista Jr. and Almeida [5]. In the cited paper, the authors consider a basic equation widely used in aging creep models [32, 69, 5], which consists of the double power-law (DPL), being defined as:

$$J(t, t_c) = J_0(t_c)\left[1 + \phi_0 t_c^{-d} \theta^{p_0}\right] \tag{3.12}$$

in which $J_0(t_c)$ is the initial compliance at the loading age t_c, ϕ_0, d, and p_0 are fitting parameters. The elapsed time θ is defined as:

$$\theta = t - t_c \tag{3.13}$$

and the initial compliance is:

$$J_0(t_c) = A_1 e^{-\eta_1 t_c} + A_2 e^{-\eta_2 t_c} \tag{3.14}$$

in which, A_1, A_2, η_1, and η_2 are fitting parameters.

The creep parameters ϕ_0, d, p_0, A_1, A_2, η_1, and η_2 are obtained by fitting the time-dependent strain curves $\varepsilon\,(\theta)$ when a constant load σ_0 is applied to the creep models. These parameters were obtained from the literature and more details on the process of adjusting these parameters are found in Evangelista Jr. and Almeida [5]. In the present paper, the fit of the parameters of the creep model is not our focus.

3.4. METHODOLOGY

The methodology considered in this paper aims to train surrogate models to predict the output response in uncertainty quantification and structural reliability processes in which is considered the randomness of the input variables. The surrogate models are based on radial basis functions and were trained to predict five different cases of loading age (t_c) and time (t) which are: (*i*) t_c = 1 day and t = 11 days; (*ii*) t_c = 1 day and t = 365 days (1 year); (*iii*) t_c = 7 days and t = 17 days; (*iv*) t_c = 7 days and t = 365 days (1 year) and (*v*) t_c = 7 days and t = 3650 days (10 years). The following subsections present the details of the used procedures.

3.4.1. Age and time-dependent creep function

The characterization of the material as well as the parameters fitted for the age and time-dependent creep models come from Evangelista Jr. and Almeida [5]. The authors considered experimental results of creep tests from Atrushi [69] for a concrete mixture in which several creep tests were performed for different loading ages of 1, 3 and 8 days. For each loading age, two samples were considered to be tested. The final creep function regressed by Evangelista Jr. and Almeida [5] is illustrated in Figure 3.3.

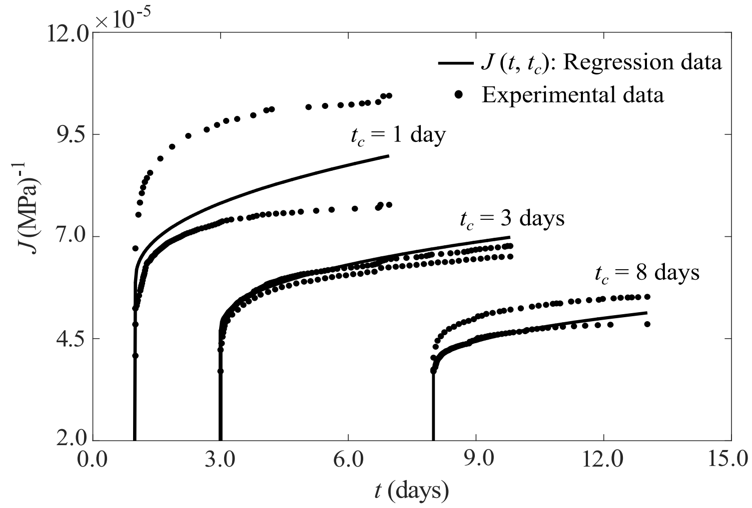


Figure 3.3 – Creep function $J(t, t_c)$ for different loading ages t_c

In Figure 3.3, the regressed creep function is represented by a solid line and the experimental data of Atrushi [69] are represented by points for different ages t_c and times t. The respective graph was constructed only for initial times t, however, the regressed function can be used for longer times, for example, times of 1, 5 or 10 years. It is also noted that the loading age t_c = 1 day presents a significant variability in the results from literature (experimental data), and this behavior is expected and characteristic of low ages of t_c. The parameters of the $J(t, t_c)$ were obtained by a process of adjustment between Equation (3.12) and the experimental results of the creep tests in the literature for the indicated loading times.

3.4.2. Deterministic and random variables

The geometric variables (W_t and e_0) were defined as deterministic variables. Furthermore, the power parameters (η_1 and η_2) of the initial compliance were also considered as deterministic variables. The energy release rate was analyzed only for the position of its maximum value in front of the crack with angle $\phi_a = 90°0'$. These variables were defined by Evangelista Jr. and Almeida [5] according to the recommendations of the Joint Committee on Structural Safety (JCSS) [70]. Table 3.1 shows the values considered for the deterministic variables.

Table 3.1 – Deterministic variables from Evangelista Jr. and Almeida [5]

Var.	**Unit.**	**Value**
W_t	mm	500.00
e_0	mm	100.00
ϕ_a	degree	90°00’
η_1	-	0.4166
η_2	-	0.0136

The random variables are shown in Table 3.2 with the respective means, standard deviations and probability density function f (x). The definition of the deterministic variables as well as the random variables of geometry and loading was performed by Evangelista Jr. and Almeida [5] by following the recommendations of the JCSS [70].

Table 3.2 – Random variables from Evangelista Jr. and Almeida [5]

Par.	Geometry			Loading		Material				
	a/c	*a/e₀*	*c/Wₜ*	σ_t (MPa)	σ_b (MPa)	A_1 (MPa^{-1})	A_2 (MPa^{-1})	ϕ_0	d	p_o
μ	0.40	0.20	0.10	5.00	1.20	28.53E-6	40.52E-6	24.48E-2	13.06E-2	42.93E-2
σ	0.08	0.04	0.02	1.00	0.24	36.85E-6	17.81E-6	1.97E-2	6.25E-2	5.23E-2
$f(x)$	GEV I (max)	GEV I (max)	LN	GEV I (max)	LN	LN	LN	LN	LN	LN

The material parameters were determined using the bootstrap method, making it possible to quantify the respective means, standard deviations and probability density functions. These parameters were also determined by Evangelista Jr. and Almeida [5] with readjustments of 200 times through bootstrap resampling. More details on the bootstrap method can be found in [71, 72]. It is also noted in Table 3.2 that the distribution that best represented the behavior of most random variables was the lognormal distribution (LN).

3.4.3. Training sampling

The training set $\mathfrak{I} = \{(\boldsymbol{\xi}_i, y_i) |\ i = 1, \dots p\}$ was generated to train surrogate models based on RBFs. The construction of the training sets was established according to the procedure represented in the blue boxes of Figure 3.1. The sampling method used to build the training sets is the Latin hypercube sampling (LHS) which was chosen due to its simplicity in implementation and also due to its high efficiency [73, 74]. The LHS method is considered to determine p training samples, where each interval is divided equally into n_d intervals creating $(n_d)^k$ possible combinations of the k input variables. Soon after, a random algorithm selects p combinations among the $(n_d)^k$ without substitution to compose the training vectors $\boldsymbol{\xi}_i$ [5]. After obtaining the training set in the training phase, each combination of $\boldsymbol{\xi}_i$ is used to analyze the energy release rate that is verified by the Equation (3.8). Each training sample is then used to determine the value y_i and thus complete the training set $\mathfrak{I}$.

The definition of training sets is based on the paper of Evangelista Jr. and Almeida [5] in which several experiments with different training sets of size p were rigorously analyzed in order to obtain a minimum size p of 79 samples for the surrogate models with RBFs so that the models could achieve a good performance. In addition, the authors also considered another training set with the size of 180 samples in order to compare it to the minimum set of 79 samples. The definition of sample sizes was obtained through a convergence test. The initial size used for the convergence test consisted of the minimum number required for response surface models. Thus, from this minimum number, the convergence test began, in which the best results occurred for p =

79 and p = 180. In this article, no emphasis will be given to the details regarding the training process, since this paper continues the article of Evangelista Jr. and Almeida [5].

3.4.4. Parameter training and optimization

The surrogate models with RBFs Wendland-C2 (WE-C2) were trained with the training set describe in the previous item. In the training and also in the cross-validation process, the length parameter ρ of the RBFs was determined in order to optimize the prediction residual error sum of squares (PRESS) of the leave-one-out cross-validation (LOOCV). This procedure was performed using a genetic algorithm implemented in the internal code of *RBFuQ* in order to minimize PRESS. The LOOCV procedure is a cross-validation with S-folds in which the number of folds is equal to the number of available data (S = p). Both PRESS and LOOCV procedures are outside the scope of this paper, but were considered for training samples. The PRESS and LOOCV procedures were verified in the paper of Evangelista Jr. and Almeida [5] and for more details see [75, 62, 63].

3.4.5. Performance evaluation of the uncertainty quantification prediction of statistical moments, PDFs and CDFs.

After training and validation, the UQ was then performed through MC simulations in which the LHS sampling method was also considered. The MC simulations, as described above, were chosen due to their ease of implementation and popularity, as this article does not focus on the efficiency of random simulation techniques. In the UQ process, a set of 5 E+04 samples were considered, which is similar to the one adopted by Evangelista Jr. and Almeida [5]. The definition of the number of samples was determined after a convergence analysis that found that 5 E+04 samples is enough to obtain the statistical moments and also carry out the process of uncertainty quantification and structural reliability. This sample set from the MC simulation is then used to predict and provide enough output responses to determine the probability density functions (PDFs) and cumulative distribution functions (CDFs) of the surrogate models, as illustrated by the green boxes (see Figure 3.1). The set of 5 E+04 samples is also used in a reference solution that consists of analytically evaluating Equation (3.8), and the results obtained are called Monte Carlo reference solutions (MC^{REF}).

The performance of each surrogate model was compared using the normalized root mean squared error (ε_a). This normalized error consists of the square root of the normalized mean square error of the PDFs considered by Evangelista Jr. and Almeida [5]. The normalized error ε_a for the verification of the goodness of fit of the reference PDFs and the predictions of the surrogate models is defined as:

$$\varepsilon_a = \frac{\| f_{MC}(y) - f(\hat{y}) \|}{\| f_{MC}(y) - \mu_{f_{MC}} \|} \tag{3.15}$$

in which $f_{MC}(y)$ is the values of PDFs generated by MC^{REF}, $f(\hat{y})$ are the values of PDFs generated by the surrogate model, $\mu_{f_{MC}}$ is the mean value of MC^{REF}, and $\|.\|$ is the Euclidean norm.

The CDFs predicted by the surrogate models will be presented as a complementary analysis. The results of the statistical test are shown in Appendix B. The statistical test compares the CDFs constructed of the surrogate models with the CDF constructed of the reference solution. This CDF fit process was performed for all the CDFs of all the cases, however, will not be presented in this paper, as it is not part of the objective of this work. Thus, if there is interest in verifying the statistical test (Kolmogorov-Smirnov test), just check [5], where the adjustment of the CDFs was verified for almost all of the cases presented in this paper.

The statistical moments obtained from the predictions of the surrogate models are then compared to those of the reference model. In this paper, the statistical moments to be verified are: mean, variance (standard deviation), skewness and kurtosis. The first statistical moment consists of the expected value or mean and for a vector of random variables $\boldsymbol{\xi}$ is given as:

$$\mu_{\xi} = \int_{-\infty}^{+\infty} \xi \, f(\xi) d\xi \tag{3.16}$$

in which $f()$ is the probability density function. The second statistical moment [4] is the variance that measures the dispersion of the variable around the mean. The variance is defined as:

$$\sigma_{\xi}^2 = \int_{-\infty}^{+\infty} \left(\xi - \mu_{\xi}\right)^2 f\left(\xi\right) d\xi \tag{3.17}$$

In the expression of Equation (3.17), the term σ corresponds to the standard deviation, which is defined as the square root of the variance:

$$\sigma_{\xi} = \sqrt{\int_{-\infty}^{+\infty} \left(\xi - \mu_{\xi}\right)^2 f\left(\xi\right) d\xi} \tag{3.18}$$

The standard deviation is a parameter that has been constantly used instead of the variance to measure the dispersion of the output response set of the surrogate models [61, 62]. Thus, in this paper, only the standard deviation is used as a representative for the second statistical moment. The study of statistical moments is generally considered by the literature only for the first two presented. However, as in this paper we have the construction of a surrogate model, it is interesting to analyze the third and the fourth moments. The third statistical moment or skewness of a random variable is:

$$\gamma_{\xi} = \int_{-\infty}^{+\infty} \frac{\left(\xi - \mu_{\xi}\right)^{3}}{\sigma_{\xi}^{3}} f(\xi) d\xi \tag{3.19}$$

The moment of skewness describes the degree of asymmetry of a distribution around the mean that can be symmetric ($\gamma_{\xi} = 0$), positive skew ($\gamma_{\xi} > 0$) and negative skew ($\gamma_{\xi} < 0$). The knowledge of the type of asymmetry of a variable can provide important information to be considered in the uncertainty quantification process. In negative asymmetry, there is a high concentration of data greater than the mean of the variable. This behavior is inverse for positive skew [4, 1]. The fourth statistical moment to be verified is the kurtosis defined as:

$$\chi_{\xi} = \int_{-\infty}^{+\infty} \frac{\left(\xi - \mu_{\xi}\right)^{4}}{\sigma_{\xi}^{4}} f(\xi) d\xi \tag{3.20}$$

The kurtosis measures the flatness of a probability density distribution. This statistical moment is also compared to a normal distribution when its magnitude is zero. Positive kurtosis values indicate that the distribution has a sharp peak, and negative values indicate that the distribution is flat when compared to a normal distribution [1].

3.4.6. Structural reliability analysis

The samples obtained in the MC simulation were also used in the structural reliability analysis. The reliability analysis consists of an additional assessment considered in this paper for the results of the reference models and for the values predicted by the surrogate model with p = 180 in all the five cases through a performance function. Structural reliability is then obtained for a performance function $G(\boldsymbol{\xi})$ that fails when the values $G(\boldsymbol{\xi}) < 0$ [84, 85, 2]. The probability of failure is defined as:

$$P_f \left\{ G(\xi) < 0 \right\} = \int_{G(\xi)<0} \int \cdots \int f(\xi_1, \xi_2, \ldots, \xi_n) d\xi_1 d\xi_2 \ldots d\xi_n \tag{3.21}$$

in which f (.) is the multivariate probability density function (PDF) constructed from the input variables $\boldsymbol{\xi}$. The failure probability constructed from the prediction results of the surrogate models is then compared to the P_f value constructed for the reference model (MC^{REF}). The coefficient of variation [86] of the estimated probability of failure of the reference model is defined as:

$$V_{Pf} = \sqrt{\frac{(1-P_f)}{p(P_f)}} \tag{3.22}$$

Another analysis consists of the reliability index β [10, 87] defined as:

$$\beta = -\Phi^{-1}(P_f) \tag{3.23}$$

in which Φ^{-1} is the inverse of the standard normal cumulative distribution function. The reliability index obtained from the surrogate models is compared with the reliability index of the reference solution. In item 3.6.3, more details about the performance function will be defined and the results obtained will also be presented.

3.5. ASSESSMENT OF QUALITY OF RESULTS

This section presents the results of the statistical moments of the surrogate models and also of the reference solutions for the uncertainty quantification process. The statistical moments to be considered are: mean, standard deviation, skewness and kurtosis. It is important to emphasize that the variance will not be verified, but the standard deviation, as described above. In the process of quantifying the statistical moments of time-dependent energy release rate (ERR) output response, five cases were considered: (*i*) t_c = 1 day and t = 11 days; (*ii*) t_c = 1 day and t = 365 days (1 year); (*iii*) t_c = 7 days and t = 17 days; (*iv*) t_c = 7 days and t = 365 days (1 year) and (*v*) t_c = 7 days and t = 3650 days (10 years). The following subsections present the results obtained.

3.5.1. Normalized root mean square error (ε_a)

The quantitative comparison between the curves of the PDFs generated from the predictions of the surrogate models and the reference solutions is given by the values of the normalized root mean square error (ε_a) according to Equation (3.15). The smallest values of ε_a are from the surrogate models of RBF (WE-C2) built with a training set of 180 samples (black color). These results show that the PDFs of the predictions are close to the MC^{REF} instead of the models built with 79 training samples (light gray color). This expansion of the training set from 79 to 180 considerably reduced the values of normalized errors in all the cases. Moreover, the case of t_c = 1 day / t = 11 days presented the smallest error magnitude and the case of t_c = 7 days / t = 3650 days the largest magnitude of error normalized to p = 180.

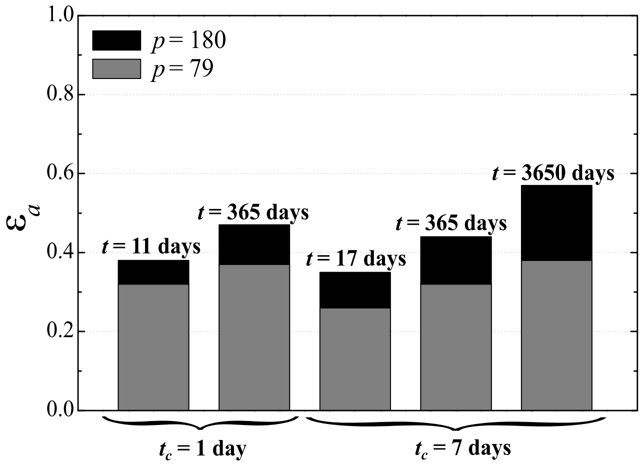
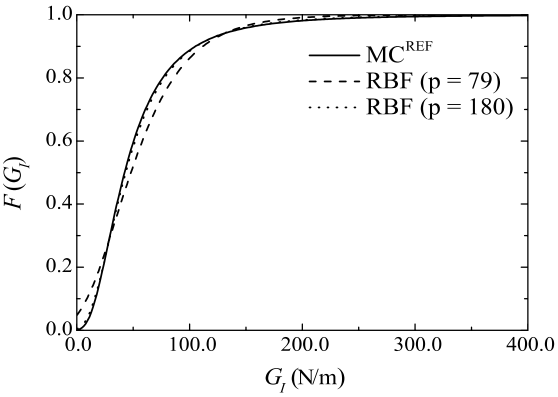


Figure 3.4 – Root mean square normalized errors, ε_a

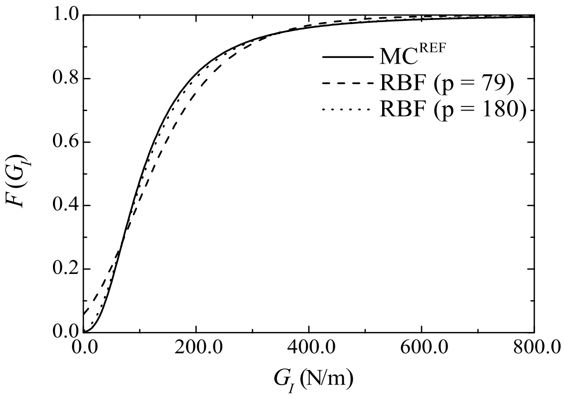
3.5.2. Cumulative distribution functions (CDFs) predictions

The CDFs of $F(G_I)$ were obtained from the prediction results of the surrogate models and the MC^{REF} reference solution for the five studied cases. These CDFs are graphically represented in Figure 3.5 for the training sets p = 79 and p = 180.

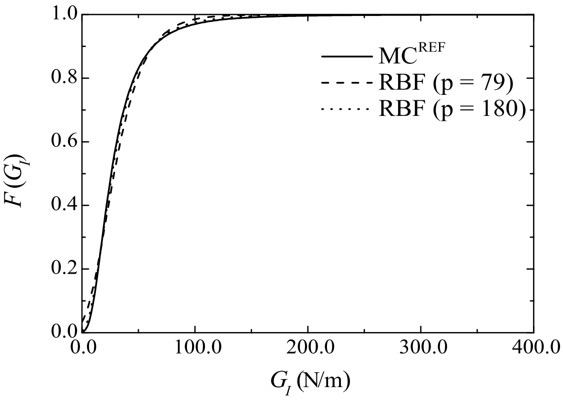
In Figure 3.5 it is noted that the surrogate models with the training set of 180 samples present a better approximation in relation to the curves of the reference solution. However, for the training set with 79 samples, only the cases t_c = 7 days / t = 17 days and t_c = 7 days / t = 365 days show a significant approximation of the CDFs in relation to the reference model. Although the statistical test to verify the fit of the CDFs is not presented in this paper, the results of this statistical test are shown in Appendix B. In addition, it is observed that for very early times (t = 11 days) and very long times (t = 3650 days) there is a difficulty in predicting the behavior of the CDFs, and this behavior is expected and initially observed in the paper of Evangelista Jr. and Almeida [5].



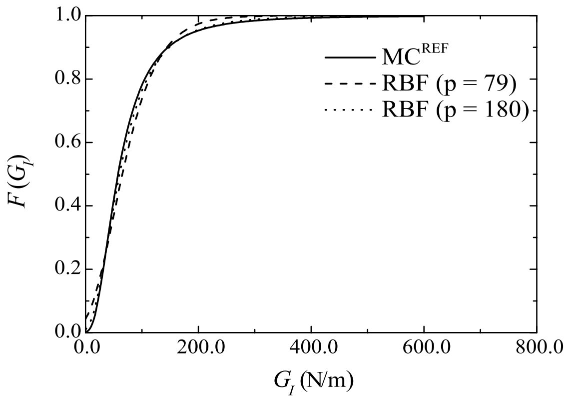
a) t_c = 1 day and t = 11 days



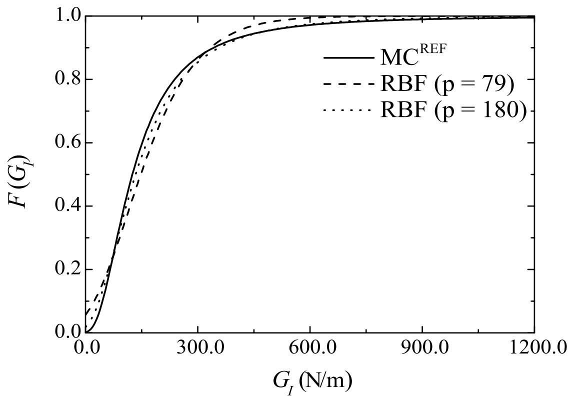
b) t_c = 1 day and t = 365 days



c) t_c = 7 days and t = 17 days



d) t_c = 7 days and t = 365 days



e) t_c = 7 days and t = 3650 days

Figure 3.5 – Comparison of the cumulative distribution function (CDF) for all cases

3.5.3. Table of moments

The results of the means, standard deviations, skewness and kurtosis of the MC^{REF} reference solutions and of the surrogate models with RBF (WE-C2) are presented in Table 3.3. The statistical moments of the case of t_c = 1 day / t = 11 days; t_c = 7 days / t = 17 days and t_c = 7 days / t = 3650 days (10 years) were extracted from PDFs and CDFs predicted by the RBF (WE-C2) models from Evangelista Jr. and Almeida [5].

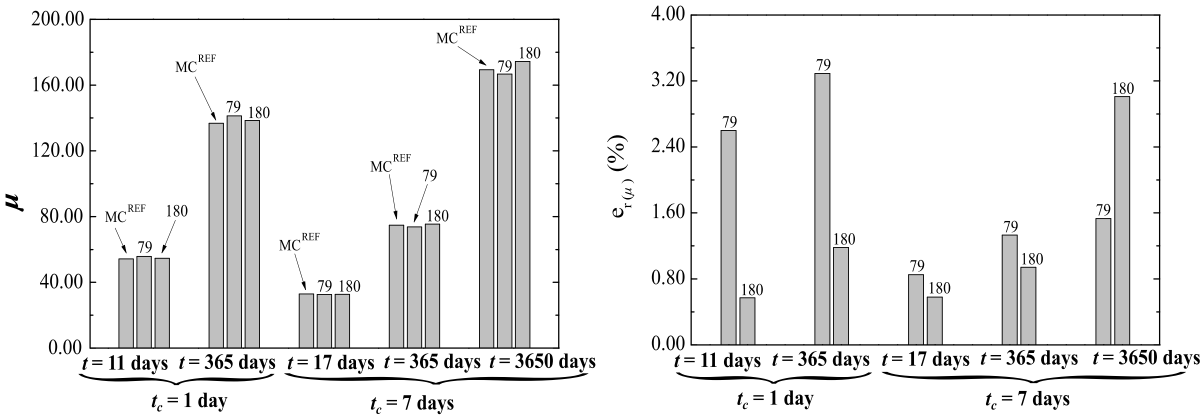
Table 3.3 – Statistical moments of the five cases

Case	p	Surrogate	μ	$e_r(\mu)$ (%)	σ	$e_r(\sigma)$ (%)	γ	$e_r(\gamma)$ (%)	χ	$e_r(\chi)$ (%)
t_c = 1 day / t = 11 days	79	RBF (WE-C2)	55.75	2.60	43.12	5.03	1.11	34.93	4.70	27.04
	180	RBF (WE-C2)	54.65	0.57	40.89	0.41	1.59	6.15	6.44	0.12
	-	MC^{REF}	54.34	-	41.06	-	1.70	-	6.45	-
t_c = 1 day / t = 365 days	79	RBF (WE-C2)	141.29	3.29	114.24	6.34	1.09	36.81	4.51	30.30
	180	RBF (WE-C2)	138.41	1.18	108.54	1.03	1.58	8.71	6.22	3.76
	-	MC^{REF}	136.80	-	107.43	-	1.73	-	6.46	-
t_c = 7 days / t = 17 days	79	RBF (WE-C2)	32.61	0.85	23.36	4.21	0.96	42.12	4.10	35.50
	180	RBF (WE-C2)	32.70	0.58	23.29	4.50	1.32	20.52	5.11	19.57
	-	MC^{REF}	32.89	-	24.38	-	1.67	-	6.35	-
t_c = 7 days / t = 365 days	79	RBF (WE-C2)	73.74	1.33	53.59	6.75	0.88	47.93	3.89	38.35
	180	RBF (WE-C2)	75.44	0.94	57.12	0.62	1.42	15.80	5.49	12.96
	-	MC^{REF}	74.73	-	57.47	-	1.68	-	6.31	-
t_c = 7 days / t = 3650 days	79	RBF (WE-C2)	166.71	1.53	126.61	8.91	0.84	51.69	3.79	40.28
	180	RBF (WE-C2)	174.38	3.01	145.21	4.47	1.55	10.75	6.04	4.89
	-	MC^{REF}	169.29	-	139.00	-	1.73	-	6.35	-

Table 3.3 also shows that the adjustment of skewness and kurtosis moments show a better accuracy for the training set with 180 samples. The results with 79 samples show predicted values of skewness and kurtosis a little distant from the reference solution. This behavior demonstrates the difficulty of predicting viscoelastic models and age and time-dependent fracture mechanics.

3.5.4. Figures of statistical moments

Another way of verifying the results of the statistical moments is shown in the plots of means, standard deviations, skewness, and kurtosis. Figure 3.6 shows the expected values for the cases with different t_c and t, in which, for each case, the value of the MC^{REF} reference solution and the training sets (p = 79 and p = 180) and the respective relative error are presented.

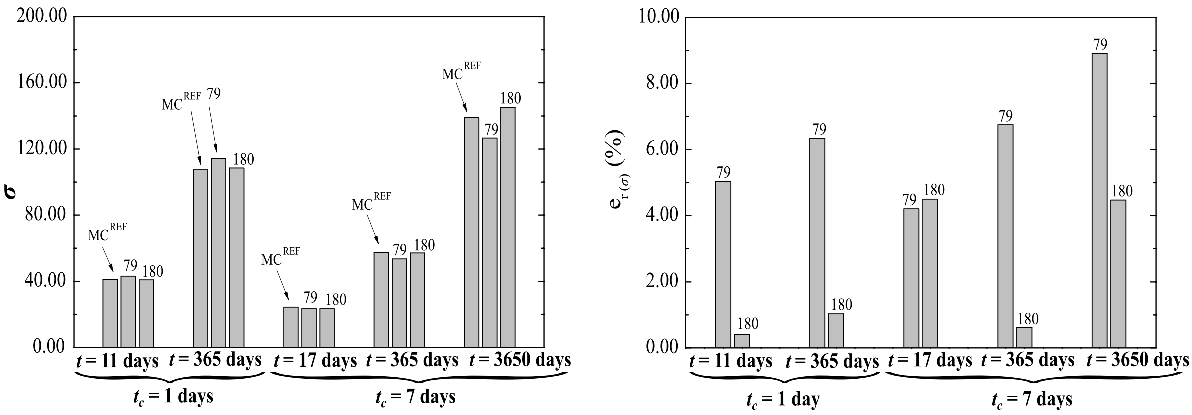


a) Mean value b) Relative error

Figure 3.6 – Expected values or mean of the predictions and the reference model

The results predicted from the mean by the surrogate models presented a good approximation in relation to the mean obtained from the output response of the reference model. The case with t_c =

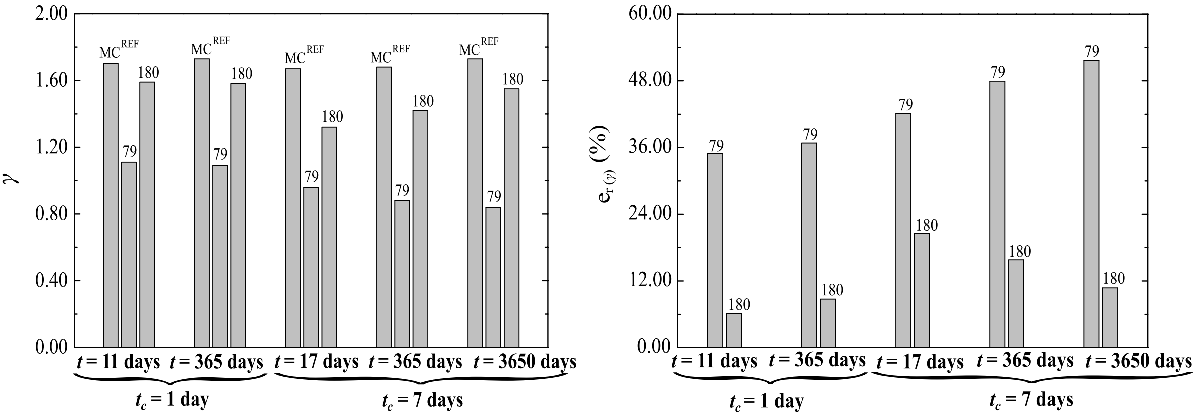
7 days and with t =17 days showed magnitudes practically similar to the solution obtained analytically for the respective case for both training sets. In addition, an analysis was performed for the relative error of the means in which for most of the mean predicted results, the relative error for the initial training set of 79 samples has a high magnitude which indicates the need for a larger training set. This relative error is clear mainly for the lowest loading age, which is t_c = 1 day, decreasing for the age of t_c = 7 days. However, for the training set of 180 samples, it is noted that the relative error is less than half of the error with the set of 79 training samples for the cases with t_c = 1 day. The case with $t_c = 7$ days and with t =3650 days is the only exception, as the best approximation will occur for the smallest training set of 79 samples. This shows how difficult it is to predict high times, such as 10 years or 50 years. The results of standard deviations are shown in Figure 3.7.



a) Standard deviation value b) Relative error

Figure 3.7 – Standard deviation values of the predictions and the reference model

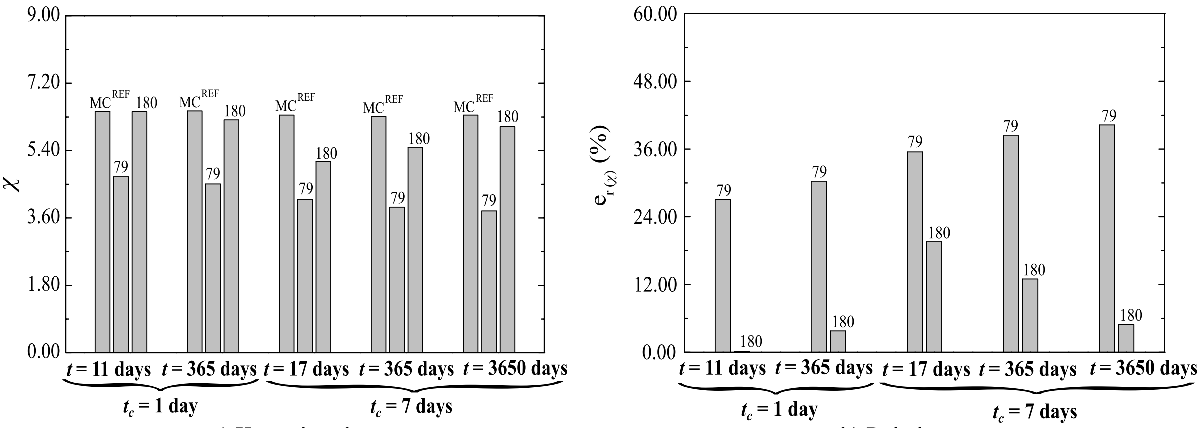
The standard deviation predictions show a high accuracy for most cases with p = 180, with the exception of the case with t = 3650 days which showed a significant difference in the values obtained by the surrogate models in relation to the MC^{REF} solution. In the case of t_c = 7 days / t = 3650 days, high values of relative error are observed both in the results of the surrogate models with p = 79 and in the results of p = 180. This behavior, as mentioned earlier, is due to difficulty in predicting long elapsed times. The case with t_c = 7 days / t = 17 days are also showed a peculiar behavior, as it can be seen in the figure that the best approximation of the standard deviation occurs for the smallest training set p = 79. For the training sets p = 79, it is also important to emphasize that even the relative error magnitudes being greater than 4%, the standard deviation results do not present a bad fit, as verified in Evangelista Jr. and Almeida [5], the uncertainty quantification process indicated a good accuracy of the surrogate models with RBF (WE-C2). In addition, it is also noted that the high magnitudes of standard deviation indicate a high dispersion of the predicted results, which explains the difficulty of the prediction, mainly for times of 1 and 10 years. The results of the third statistical moment or skewness are shown in Figure 3.8.



a) Skewness value b) Relative error

Figure 3.8 – Skewness values of the predictions and the reference model

The skewness results show a positive value in which the output response vector presents a high concentration of data smaller than the expected value of the respective distribution. This behavior indicates that the longest tail of the distribution is to the right of the probability density function. The results predicted by the surrogate models with the minimum number of 79 samples were not able to adequately predict the third statistical moment. For the training set of 180 samples, it is noted that the relative error is less than half of the relative error of the smallest training set. Thus, it is observed that the skewness consists of the statistical moment that presents the greatest difficulty of prediction for the surrogate models. One solution to get a better fit is to increase the size of the training set. The last statistical moment or kurtosis is shown in Figure 3.9.



a) Kurtosis value b) Relative error

Figure 3.9 – Kurtosis values of the predictions and the reference model

The kurtosis values show that, for all the cases, the results of the MC^{REF} reference models and the surrogate models show a tendency to taper at the peak of the PDFs. Furthermore, it is also noted that the smallest training set (79 samples) is not able to predict the kurtosis of the cases studied. The training set with 180 samples presents very similar magnitudes for the cases of t_c = 1 day and for the case with t_c = 7 days / t = 3650 days. The relative errors for the set with 180 training samples show magnitudes much lower than those obtained for the sets with 79 training samples, especially for the case t_c = 1 day / t = 11 days in which the error is practically zero.

3.6. ANALYSIS OF TIME-DEPENDENT ENERGY RELEASE RATE (ERR)

The section presents the PDFs built from the reference solutions of the five case studies and also shows the statistical moments as a function of age and loading time. In addition, a reliability analysis is shown for the results of the reference models through the construction of a performance function shown in item 3.6.3.

3.6.1. Prediction of probability density functions (PDFs)

The PDFs built for the MC^{REF} reference solutions are shown in Figure 3.10 for the five cases considered here. These curves are used as a reference for the validation of the surrogate models. The results of the reference models are obtained for the time-dependent energy release rate by analytical evaluations of Equation (3.8).

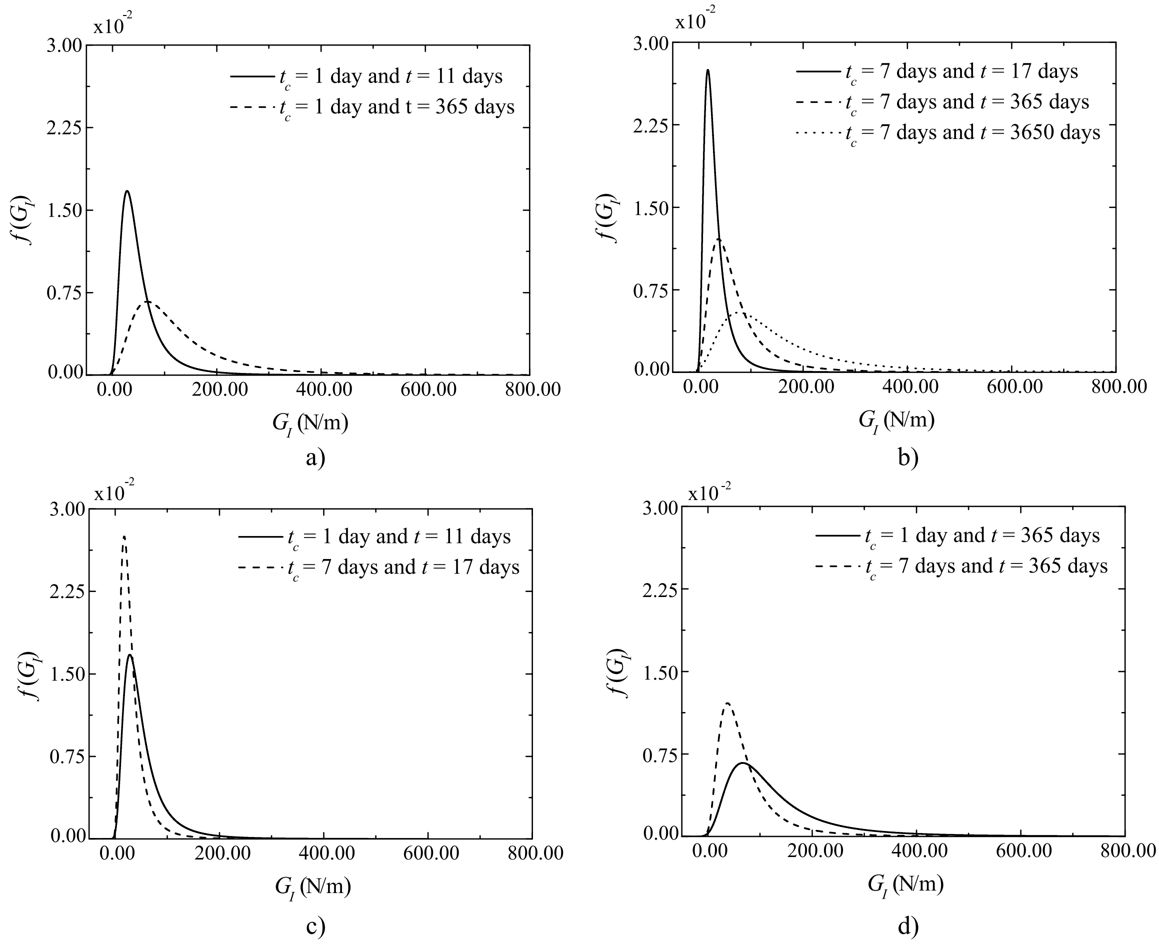
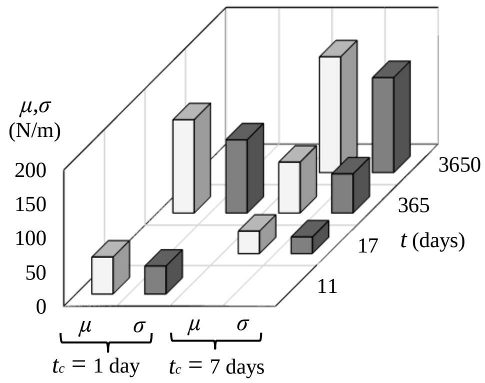


Figure 3.10 – Probability density functions (PDFs) of the MC^{REF} solutions for the five cases

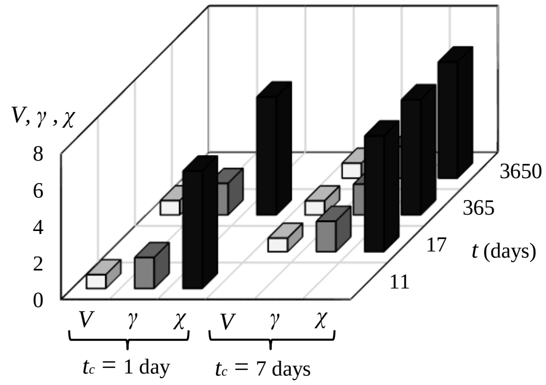
The curves show a greater dispersion for the cases where $t_c = 1$ day / t = 365 days and for $t_c = 7$ days / t = 3650 days as indicated by their long tail distributions. These long tail PDFs show a real difficulty for the surrogate models, as the energy release rate values are very dispersed over a long range of values. In Figure 3.10a and Figure 3.10b it can be seen that as the elapsed time increases there is a greater dispersion of the energy release rate (E.R.R.). In Figure 3.10c it is observed that for the same elapsed time there is a greater concentration of the E.R.R around the mean as t_c increases.

3.6.2. Figures 3D

The values of the statistical moments (mean, standard deviation, skewness, kurtosis) constructed with the samples of the MC^{REF} reference solution for the five cases are shown in Figure 3.11 as a function of loading times and ages.



a) mean and standard deviation



b) coefficient of variation, skewness and kurtosis

Figure 3.11 – Statistical moments of G_I as a function of age and loading time for the five cases

The results of the means show that for premature ages there is a tendency for the mean value of the energy release rate to increase as time also increases. For the cases with time t = 365 days, it is noted that the mean results tend to present higher magnitudes for smaller ages of loading, that is, the mean for t_c = 1 day / t = 365 days is greater than the mean of t_c = 7 days / t = 365 days. The standard deviations also show the same behavior of the mean values and for initial ages t_c = 1 and 7 days, there is a significant increase in the values of standard deviations for longer times. Furthermore, for times of t = 365 days (1 year), it is noted that the deviation tends to present higher values for lower loading ages. The skewness results show a different behavior from the one showed for the means and standard deviations. It is noted that the magnitude of the skewness practically does not change for longer times. The behavior of the statistical moment of kurtosis is similar to the skewness, that is, the kurtosis moment for premature loading ages does not change significantly in its magnitude for longer elapsed times.

3.6.3. Reliability analysis

In the failure analysis of the structural component, a performance function was constructed for the output response of the reference solution. The performance function that verifies the energy release rate is defined as:

$$G = G_I^R - G_I^S \tag{3.24}$$

in which $G_I{}^R$ is a G_I resistance function and $G_I{}^S$ is the demand function. The $G_I{}^S$ function is defined as the answer to Equation (3.8) and modeled in this paper for PDFs of the Section 3.6.1. The $G_I{}^S$ function is therefore the modeling of soliciting forces. The $G_I{}^R$ function is estimated and based on the mean of the PDF of $\mu_{G_I^S}$ with a safety factor as $\mu_{G_I^R} = \gamma_0 \mu_{G_I^S}$ and a coefficient of variation of 20% and value γ_0 = 6. The $G_I{}^S$ function corresponds to the MC^{REF} results. This performance function configuration follows the recommendations prescribed by the JCSS [70] which were also considered to determine the random variables of the concrete plate. The configuration considered for this performance function was due to the fact we did not known the behavior of the resistance function. In this way, an alternative solution to construct the resistance function was to consider a value related to the mean of the demand function. Figure 3.12 shows the respective failure probabilities, P_f, and the reliability index, β, for the MC^{REF} assessments.

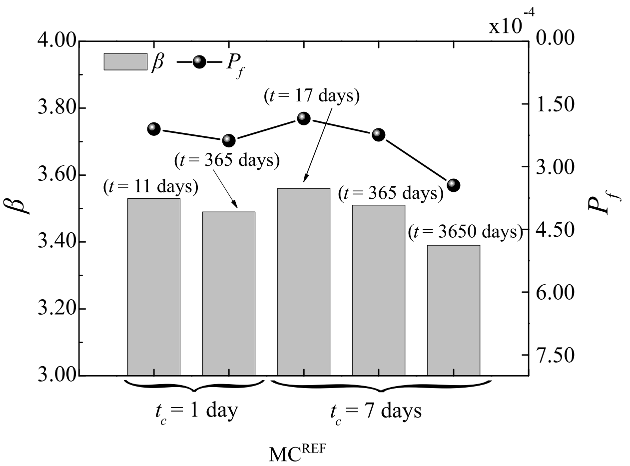


Figure 3.12 – Failure and reliability analysis for all the cases

The P_f values present a low probability of failure around 10^{-4} in which 5 E+04 evaluations of the integral of Equation (3.8) were necessary to determine the time-dependent energy release rate for the reference solution. The coefficient of variation of P_f for $t_c = 1$ day / $t = 11$ days is 0.31, for $t_c = 1$ day / $t = 365$ days it is 0.29. For the case $t_c = 7$ days / $t = 17$ days, the coefficient of variation is equal to 0.33 and for the case $t_c = 7$ days / $t = 365$ days, the value of V_{P_f} is equal to 0.30. The smallest coefficient of variation (V_{P_f}) is for $t_c = 7$ days / $t = 3650$ days with value equal to 0.24. The performance functions that were determined for all the cases also show that for both the MC^{REF} results, there is an increase in the influence of the energy release rate at initial ages of loading as times increases. This behavior shows, therefore, that the probability of failure tends to be higher and, consequently, the reliability index tends to be lower. However, as mentioned above, the failure is small and consequently difficult to occur due to the high values of reliability index.

3.7. CONCLUSIONS

The paper developed a procedure to quantify the uncertainty using surrogate models in age and time-dependent fracture mechanics problems. The problem under study consisted of a concrete plate with an initial crack to verify the energy release rate. Two different types of training sets were used: with 79 samples and with 180 samples. These training sets were used to train surrogate models based on radial basis functions of the Wendland-C2 in order to predict output values of the energy release rate for five cases of loading ages and acting times. The surrogate models were able to predict thousands of unknown Monte Carlo simulation values that were unseen by the already trained surrogate model. Then, the post-processing was done through the uncertainty quantification given by the moments of the output response obtained by the surrogate models. These results were compared to the reference solutions constructed by solving the convolution integral of the energy release rate.

The verification of the goodness of fit of the PDF was performed by the analysis of the root mean square normalized error (ε_a). The normalized error constructed from the PDFs of the surrogate model predictions and the reference model solutions showed that both training sets for all cases of loading age and elapsed time performed well with error magnitudes less than 0.60. Another verification consisted in the study of the behavior of the CDFs that presented a better fit only for the training set of 180 samples, and it was also observed that at premature loading ages and very long initial times or acting times there is a difficulty in predicting the behavior of the energy release rate when the comparison was made between the CDFs constructed from the predictions of the surrogate models and the CDFs from the reference solutions.

The statistical moments verified were: mean, standard deviation (variance), skewness and kurtosis. The results obtained for the expected value or mean of the output response indicated a good fit of the surrogate models in relation to the reference solution, mainly for the largest training sets (180 samples) that present relative error magnitudes smaller than the training set of 79 samples. The second statistical moment or variance was replaced by the verification of standard deviations. The standard deviations obtained by the results of the surrogate models also present a better accuracy for the largest training set, with a significant decrease in the relative errors being observed with the increase in the size of the training set. For the third statistical moment or skewness, it was observed that in all the cases both the reference solution and the values predicted by the surrogate models showed a positive skewness behavior that indicates a greater concentration of data in the region to the left of the expected value and consequently the presence of a long tail to the right of the distribution. On the other hand, this statistical moment consisted of the uncertainty quantification process that presented the greatest difficulty of prediction by the surrogate models with a better fit only for the largest training sets. The fourth statistical moment or kurtosis indicated the presence of a tapering at the peak of the PDFs constructed by both the surrogate models and the reference solutions. The kurtosis results also show only a good fit for the largest training set.

In the analysis of the PDFs built from the analytical evaluations of the energy release rate of the reference solution for the five studied cases it was also observed a tendency which consists of a greater dispersion of the energy release rate as the elapsed time increases. Furthermore, the mean and standard deviation of these PDFs also show an increase in their magnitude as time increases. This increase in means and standard deviations is proportional and does not directly affect the coefficient of variation. However, the statistical moments of kurtosis and skewness do not show significant variation in their magnitudes for different elapsed times. Finally, in a structural reliability analysis, it was found that at initial loading ages there is a tendency to increase the probability of failure as the elapsed times are greater. This behavior was observed for the results of the reference solution. However, the magnitude of the failure which is around 10^{-4} indicates a high structural reliability index for all cases. Thus, it can be concluded that the computational procedure with surrogate models based on RBFs is efficient and accurate for structural engineering problems in uncertainty quantification processes involving time-dependent fracture mechanics.

4. A NOVEL COMPACTLY SUPPORT RADIAL BASIS FUNCTION AS SURROGATE MODEL IN STRUCTURAL RELIABILITY ANALYSIS USING VERY FEW TRAINING SAMPLES

4.1. INTRODUCTION

Structural reliability is an important vehicle to ensure the safety of the structures and the assessment of this reliability is usually given through the probability of failure of the output response of the structural problem and also through the quantification of the inherent uncertainty of the material properties, geometry, loading and age of the structural component. This structural reliability generally presents techniques based on statistical moments, being therefore quantities derived from the variation spectrum of the variable. The quantitative of structural reliability analysis is treated here as a subdivision of uncertainty quantification. The uncertainty quantification process allows us to obtain the entire spectrum of variation of the output random variable. Consequently, it is possible to determine the probability density functions and cumulative distribution functions. Moreover, structural reliability analysis is an area constantly studied over the years [88, 11, 30, 53] in which the study is basically made in two ways considering classical methods or approximation methods. However, only the second has gained more force to solve stochastic models due to its lower computational cost and accuracy of these models [30, 5]. The basic idea of these models is to approximate a limit state of a performance function with high nonlinearity and a prohibitive computational cost.

The classical methods such as Monte Carlo simulation [89, 90, 53]; First-order Reliability Method, FORM [91, 4, 25]; and the Second-order Reliability Method, SORM [92, 93, 2] have been considered over the last 40 years in the analysis of the performance functions of structural components. The Monte Carlo simulation method is a well-known and easy-to-use method. However, the Monte Carlo method has several disadvantages, such as a prohibitive cost for a large number of samples. Furthermore, several other methods derived from Monte Carlo simulation have been used, such as directional simulation [94, 95] and importance sampling [96, 97]. On the other hand, these presented methods still have some difficulty predicting the behavior of multivariate cases. As a solution to solve the structural reliability there is the use of approximation methods called surrogate models that are employed by several researchers [88, 11, 98, 99, 100, 101].

The surrogate modeling techniques are considered constantly to obtain an approximate solution to many problems of structural components, such as truss bars, frame bars, and concrete beams [98, 102, 103, 104, 105, 106]. The first cited paper from Dai and Cao [98] considered the use of Wavelet Neural Networks in Support Vector Machine (SVM) to approximate the response of structural components. The structural components studied by the authors are a dynamic element of six degrees of freedom and a truss string structure with multiple random variables. The results demonstrated the efficiency of the surrogate model in the failure analysis. Another important paper is from Kroetz and co-authors [102] that analyzed the reliability and the uncertainty propagation of complex problems using Artificial Neural Networks (ANN), Polynomial Chaos Expansion (PCE), and Kriging. The results demonstrated that the three surrogate models are competitive to solve stochastic problems. Guo and co-authors [104] studied the reliability of two structural engineering cases using the maximum entropy method with PCE. The results demonstrated that PCE-based multiplication decreased the number of evaluations of the limit state function and the method proposed by the authors was similar to FORM and SORM when the number of evaluations was close [104]. Finally, the last relevant paper to be mentioned is from Hong and co-authors [105] in which Radial Basis Functions (RBFs) and an adaptive sequential sampling method were used in a proposed surrogate model for reliability analysis with five numerical examples. The results demonstrated high precision and applicability in structural reliability analysis [105]. The RBF technique is also an important strategy to be considered as a surrogate model and has been successfully used to solve stochastic engineering problems [107, 108, 99].

The radial basis function is a technique used initially in the approximation of functions and derivatives [109, 110, 111]. This technique is also used in structural reliability and uncertainty quantification analysis as surrogate models in which adaptive models are adopted in several papers [107, 108, 99, 30, 5, 112]. Wang and co-authors [108] considered surrogate modeling with RBFs in reliability analysis of tunnels in which was observed an effective surrogate model for predicting tunnel engineering problems [108]. Li and co-authors [30] studied benchmark problems, such as a circular tube beam and an oscillator through surrogate modeling techniques with Monte Carlo simulation to obtain the probability of failure and the reliability index. The reliability results demonstrated that the RBF-based method was able to approximate the examples. Another important paper is from Evangelista Jr. and Almeida [5] on which the radial basis function technique was used to quantify the uncertainty of the energy release rate of a concrete plate under the presence of an initial crack. The results predicted from the RBFs were compared with the traditional second-order polynomial and the RBFs were able to predict long-tail probability density functions with a small training set size. In this way, RBF surrogate model technique is an alternative to approximate complex surfaces, solve multivariate cases and avoid the Curse of dimensionality.

The aim of this paper is two-fold. Firstly, it originally develops an efficient surrogate model based on a new radial basis function (RBF) to accurately study the reliability of selected numerical examples from the literature. The first and the second numerical example are mathematical problems from the literature and third and the fourth examples are structural engineering problems. The main contribution is to propose the adaptation of the Flat-top equation as an RBF of three parameters and also to develop a computationally innovative strategy for the new RBF with a small training set. Secondly, quantify the reliability through a comparison of the obtained results of the new RBF with the traditional RBFs and with other comprehensive methods from literature (Kriging, Neural Networks, PCE) in some classes of problems: nonlinear functions and dynamic oscillator. This comparison of RBFs results is in terms of relative errors of the reliability index and probability of failure in the reliability analysis. The performance of the new RBF and also the traditional RBFs showed a good approximation with a relatively small training point set. Another study was done on the output response of the reference solution to quantify the uncertainty of the input variables. The results showed that in all the numerical examples the high uncertainty of the input parameters demands a prohibitive computational cost of the reference solution.

This paper is organized as follows: Section 4.2 presents the background on surrogate models with the definitions of RBFs. Section 4.3 shows the methodology while Section 4.4 presents the numerical examples with the results and discussion. Finally, Section 4.5 presents the conclusions and final considerations.

4.2. BACKGROUND ON SURROGATE MODELS

A surrogate model or metamodel aims to approximate the answer quickly (low computational cost) and with the required precision. This strategy can be constructed by fitting (training) a generalized linear model with a set of training samples. It is important to emphasize that in this paper; the application of the surrogate model occurs in structural reliability problems. Figure 4.1 shows the flowchart of training and application of surrogate models with radial basis function (RBF).

The blue boxes indicate the training phase path in which the surrogate models learn from the training set. The green boxes show the application path of the already trained surrogate models to predict new output responses in reliability analysis (RA) using random sampling in the Monte Carlo (MC) simulation method. It is important to emphasize that random sampling can be performed with several techniques, but in this paper, the MC simulations are used only in the surrogate model and not for the physical/mathematical model once it is trained.

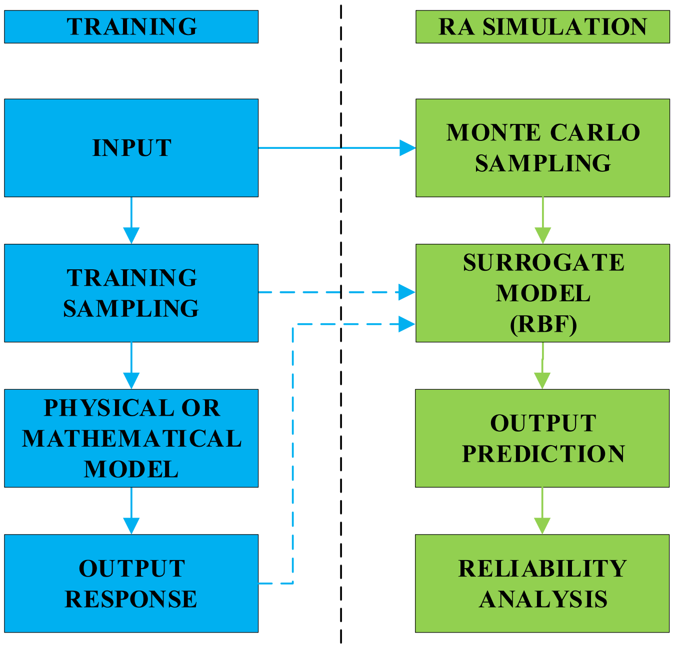


Figure 4.1 – Surrogate model training and simulation strategy flowchart.

The surrogate models can construct $\hat{y}(.)$ of a given function $y(.)$ through a general linear model:

$$\hat{y}(\xi) = \sum_{j=1}^{m} \omega_j h_j(\xi) \tag{4.1}$$

in which the m regressors $h_j(\boldsymbol{\xi})$ are fixed functions of the inputs $\boldsymbol{\xi} \in \mathbb{R}^k$, with $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_k)$ and k is the number of input variables, therefore the dimension of the problem. The m unknowns are the coefficients ω_j determined through a fitting process based on a training set $\mathfrak{I} = \{(\boldsymbol{\xi}_i, y_i) | i = 1, \dots p\}$ of data samples in which the outputs $y_i \in \mathbb{R}$ in a global domain $(\boldsymbol{\Omega})$. The quality of the fitting/training procedure can be accessed through a set of unseen data by the model or a cross-validation process that splits the data set into training and validation (unseen data during training) subsets. Different surrogate models can be constructed using different regressor functions $h_j(\xi)$, but in this paper, the regressors is detailed in the following subsections.

4.2.1. Traditional radial basis function surrogate models

The RBF networks are known to be universal approximators for regression problems [63] and follow the rewriting of Equation (4.1):

$$\hat{y}(\xi) = \sum_{i=1}^{m} \omega_i \Psi_i(\xi, \rho) \tag{4.2}$$

in which $\Psi_i(\boldsymbol{\xi}, \rho)$ is the radial basis function (RBF) with a given width (or support) ρ, and ω_i their respective coefficients or weights. The RBFs considered in this work are not related to a neural network. The RBF is used here as a regression model in which the weights are determined by least squares. Equation (4.2) can be also expressed as:

$$\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\omega} + \mathbf{e} \tag{4.3}$$

in which $\mathbf{y}$ is the p-dimensional vector of the known output sets, $\boldsymbol{\omega} = \{\omega_1, \omega_2, \dots, \omega_m\}$ is the unknown coefficient vector, and $\mathbf{e} = \{e_1, e_2, \dots, e_p\}$ is the vector error. The $p \times m$ design matrix $\boldsymbol{\Psi}$ contains the response of the m regressors obtained for the already known p input data:

$$\boldsymbol{\Psi} = \begin{bmatrix} \Psi_1\left(\frac{\|\xi_1 - \xi_1^c\|}{\rho}\right) & \Psi_2\left(\frac{\|\xi_1 - \xi_2^c\|}{\rho}\right) & \cdots & \Psi_m\left(\frac{\|\xi_1 - \xi_m^c\|}{\rho}\right) \\ \Psi_1\left(\frac{\|\xi_2 - \xi_1^c\|}{\rho}\right) & \Psi_2\left(\frac{\|\xi_2 - \xi_2^c\|}{\rho}\right) & \cdots & \Psi_m\left(\frac{\|\xi_2 - \xi_m^c\|}{\rho}\right) \\ \cdots & \cdots & \cdots & \cdots \\ \Psi_1\left(\frac{\|\xi_p - \xi_1^c\|}{\rho}\right) & \Psi_2\left(\frac{\|\xi_p - \xi_2^c\|}{\rho}\right) & \cdots & \Psi_m\left(\frac{\|\xi_p - \xi_m^c\|}{\rho}\right) \end{bmatrix} \tag{4.4}$$

in which the regressor Ψ_i is differentiated by its center ξ_i^c which corresponds to a combination from the training set $\mathfrak{I}$. The training inputs (ξ) were normalized following a standard normal distribution (mean is zero and standard deviation equal to 1). The centers were established between the training samples to ensure efficiency ($m = p$). The Equation (4.4) shows that Ψ_i process the Euclidean norm (Euclidean distance) of each input data ξ to its respective center ξ_i^c. This type of function is termed RBF and is classified into two different classes that are the globally supported RBFs and the locally (or compactly) supported RBFs [64, 65, 66]. This paper focuses attention on the two RBFs used herein: Gaussian and Wendland-C2. The reason for these choices is that each one is representative of the globally and locally supported RBF types.

The Gaussian function is the classical global RBF whose support can span all over the domain [22, 65]. There are some variations in its general equation, with the one used here given by:

$$\psi = e^{-\left(\frac{\|\xi - \xi^c\|}{\rho}\right)^2} \tag{4.5}$$

The Wendland-C2 RBF presented in Wendland [23] is classified as a compactly support RBF (CSRBF) due to its null values outside the support region ($\|\boldsymbol{\xi} - \boldsymbol{\xi}^c\| > \rho$). The equation is:

$$\psi = \begin{cases} \left(1 - \dfrac{\|\xi - \xi^c\|}{\rho}\right)^4 \left(1 + 4\dfrac{\|\xi - \xi^c\|}{\rho}\right), & 0 \le \dfrac{\|\xi - \xi^c\|}{\rho} \le 1 \\ 0 & \dfrac{\|\xi - \xi^c\|}{\rho} > 1 \end{cases} \tag{4.6}$$

The behavior of a global and local support RBF is shown in Figure 4.2 for different support values. Note that both RBFs present the same behavior, being differentiated by the actuation of the parameter ρ. For the RBF (GA), values of $\rho > 0.125$ tend to have a high influence of the RBFs in the training set, however, may presents a poor conditioning in the fit to obtain the least square coefficients because this type of RBF presents a full $\boldsymbol{\psi}$ matrix. On the other hand, the RBF (WE-C2) shown a small influence for the same parameters, but this type of RBF presents a better matrix conditioning, and consequently, may show a better accuracy than the RBF (GA).

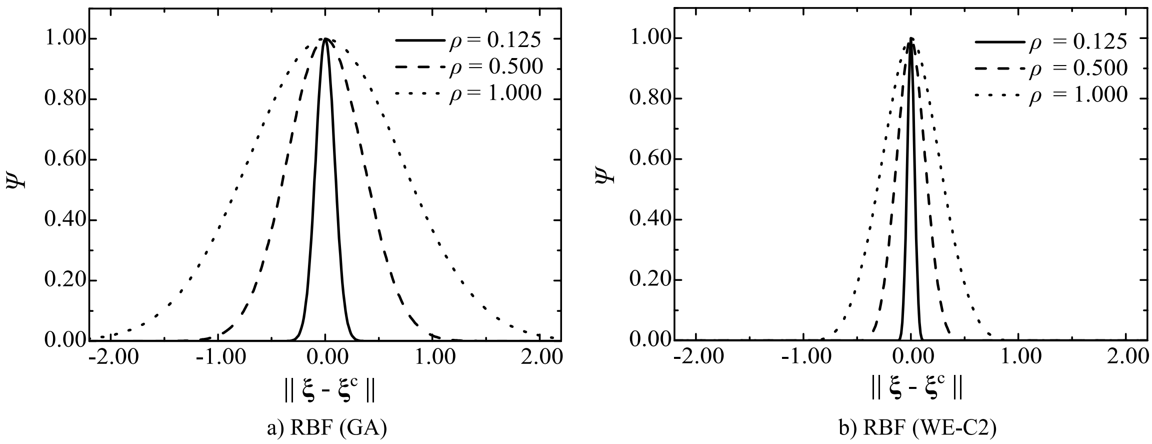


Figure 4.2 – Behavior of RBFs: a) RBF (GA); b) RBF (WE-C2).

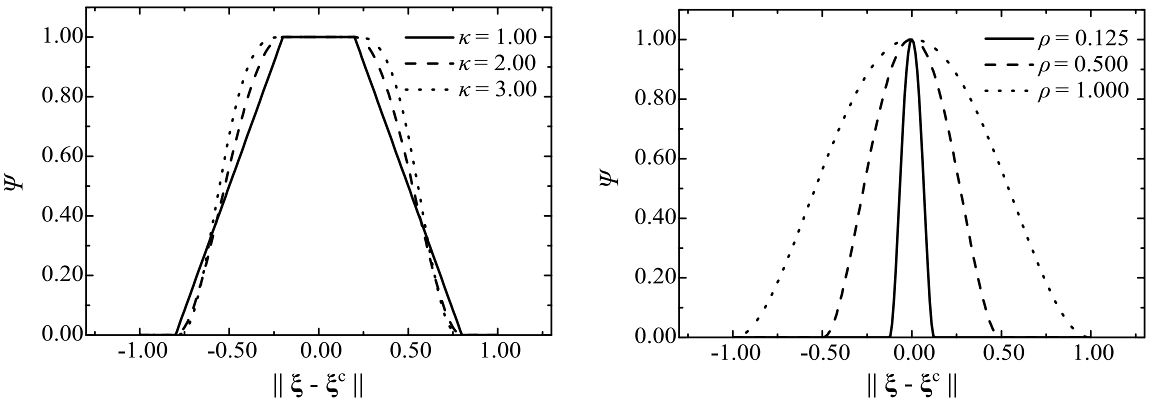
4.2.2. Proposed radial basis function surrogate model

The proposed RBF is based on the Flat-Top Unit Partition model [113, 114, 115] used in the Generalized Finite Element Method (GFEM) to study super convergence points [116]. The method is characterized by a compact support behavior, and can be defined as a new CSRBF by adapting the formulation of Zhang and co-authors [114]. The new RBF is defined as Almeida Evangelista Jr. of type Flat-top or RBF (AEJ-FT) according to:

$$\psi = \begin{cases} 1, & \text{if } \dfrac{\|\xi-\xi^c\|}{\rho} \le \beta_0 \\ \left(1-\left(\dfrac{\dfrac{\|\xi-\xi^c\|}{\rho}-\beta_0}{1-2\beta_0}\right)^{\kappa}\right)^{\kappa}, & \text{if } \beta_0 < \dfrac{\|\xi-\xi^c\|}{\rho} < (1-\beta_0) \\ 0, & \text{if } (1-\beta_0) \le \dfrac{\|\xi-\xi^c\|}{\rho} \le 1 \text{ or } \dfrac{\|\xi-\xi^c\|}{\rho} > 1 \end{cases} \quad (4.7)$$

in which, κ is a shape parameter indicating the smoothness of the curve that interconnects the flat regions of the approximation function and β_0 is the Flat top size parameter with the range $0 \le \beta_0 < 0.50$. The new RBF also presents a considerable difference from the traditional RBFs, as it has a greater number of shape parameters to ensure a better approximation of the surrogate models. The shape parameters ρ (Gaussian, Wendland, and AEJ-FT) and the additional κ and β_0 (only for AEJ-FT) can be determined using optimization procedures and genetic algorithm.

The behavior of the RBF (AEJ-FT) is showed in Figure 4.3 for different magnitudes of κ and ρ with a fixed value for the parameter β_0. In Figure 4.3a, different values of κ are presented, with the parameter $\beta_0 = 0.20$; and in Figure 4.3b, several magnitudes of ρ are shown with $\beta_0 = 0.01$ ($\beta_0 \longrightarrow 0$) and $\kappa = 2.00$.



a) RBF (AEJ-FT), $\beta_0 = 0.20$, $\rho = 1.00$ b) RBF (AEJ-FT), $\beta_0 = 0.01$, $\kappa = 2.00$

Figure 4.3 – Behavior of RBFs (AEJ): a) RBF (AEJ-FT), $\beta_0 = 0.20$, $\rho = 1.00$; b) RBF (AEJ-FT), $\beta_0 = 0.01$, $\kappa = 2.00$

In Figure 4.3a, it is observed that values of $\kappa > 1.00$ show a greater slope in the RBFs curve; and for high values of κ, the parameter κ tends to have poor conditioning in the RBFs matrices. In Figure 4.3b, the parameter ρ which is characteristic of a compact support RBF has a behavior similar to an RBF (WE-C2). It is also important to mentioned that the high values of ρ shows a high influence of the RBFs in the training set.

4.3. METHODOLOGY

The methodology of this paper aims to train the surrogate models to predict the output response under a reliability analysis framework considering the statistical uncertainty of the input variables. The strategies for the RBF surrogate models described in this item were implemented in MATLAB™ as an in-house code called *RBFuQ*. The code has pre-processing, processing, and post-processing routines to perform the data-driven approach for the random sampling, training, and reliability analysis application of the surrogate models as shown in Figure 4.1. Surrogate models using RBF (GA and WE-C2) and the new RBF (AEJ-FT) were trained to predict four numerical examples. The first and the second numerical examples consist of two mathematical problems described later. The third example consists of a structural performance function that determines the maximum deflection of a cantilever beam and the last numerical example consists of the dynamic response of an oscillator. All the numerical examples were selected because they are examples reported by the literature and they are capable of being reproduced. These examples are also solved by the literature with different techniques adopted as surrogate models. These results from the different techniques were compared with the results of the RBF-based surrogate models. The following subsections present the details of the employed methods.

4.3.1. Training sampling

A sampling procedure was employed to determine a set of training samples $\mathfrak{J} = \{(\boldsymbol{\xi}_i, y_i) | i = 1, \ldots p\}$ to be considered for training the surrogate models as can be seen in the blue boxes of Figure 4.1. Latin hypercube sampling (LHS) was considered to generate p training samples in which each interval was equally divided into n_d intervals creating $(n_d)^k$ possible combinations of the k input variables. Then, a random algorithm selects p combinations among the $(n_d)^k$ with no replacement to compose the training vectors $\boldsymbol{\xi}_i$. Finally, in the training phase, each combination $\boldsymbol{\xi}_i$ was used to evaluate the output response y_i of the four numerical examples and then complete the training set $\mathfrak{J}$. The LHS method considered here was implemented in MATLAB™ and was chosen due to its simplicity and efficiency. However, any other sampling method can be used to generate the training sampling and more details on the LHS method can be found in the papers of Shields and co-authors [73, 74].

The determination of the number of training samples is an important stage to build the surrogate models in which a bigger sample size generally provides a better generalization of predictions but demands higher computational cost and a smaller p-size training sample generally decreases the quality of the predictions due to insufficient generalization of the trained surrogate models [5]. In this way, this paper performed a series of experiments with different p-size training sets in the four numerical examples to ensure the efficiency and quality of the results. The first and the fourth numerical examples considered a minimal p-size of 19 samples to establish that the RBFs presented good performance in the reliability analysis. The second numerical example used 12 samples to build the surrogate models and the third numerical example considered a minimal p-size of 15 training samples. These minimum size of training samples were defined having as a starting point the minimum number used by the examples in the literature, and later a convergence test was carried out in search of a decrease in the sample size. This paper will not focus on these convergence tests.

4.3.2. Training, parameter optimization, and cross-validation

The surrogate models built with RBF Gaussian (GA), RBF Wendland-C2 (WE-C2), and the new RBF Almeida-Evangelista Jr. (AEJ-FT) were trained with a specific training set previously described. To find the coefficients $\boldsymbol{\omega}$, one can minimize the cost function $E = \mathbf{e}^T\mathbf{e}$ which is the sum of the squared errors of the training set $\Im$. It leads to the solution provided by a least-square procedure:

$$\boldsymbol{\omega} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\mathbf{y} \tag{4.8}$$

in which, $\mathbf{y}$ is the vector of data samples with the observed outputs. In the training and cross-validation process, the parameters (ρ, κ, β_0) of each RBF were determined to optimize the PRESS of the LOOCV. This procedure was used to measure the prediction performance of surrogate models and to guarantee that the models were not over-fitted during the training process. The LOOCV is a cross-validation of S-folds with the number of folds equal to the number of available data ($S = p$). The method removes a sample from the set of initial samples to be predicted by the trained model with the other $p - 1$ training data. This procedure is then used for the p data and the quality of the performance, known as PRESS, is then calculated for each model as:

$$PRESS = \sum_{i=1}^{p}(\hat{y}_{i,-i} - y_i)^2 \tag{4.9}$$

in which, $\hat{y}_{i,-i}$ is the predicted response of the hold-out sample and y_i is the respective known output response. Although the LOOCV is the most time-consuming S-fold cross-validation, it can be a more rigorous one and is recommended for small training samples [75, 62, 63]. The PRESS and the LOOCV are considered here to obtain the shape parameters ρ, κ, and β_0 with a genetic algorithm. The range of values of the RBFs parameters used for scanning the genetic algorithm were defined as being $0.0001 \leq \rho \leq 50$, $0 \leq \kappa \leq 6$ and $0 \leq \beta_0 \leq 0.499$ according to the suggestions in the literature. This genetic algorithm was implemented in MATLAB™ and found the optimal shape parameters through the minimization of PRESS of the LOOCV. The LOOCV results and PRESS values do not consist of the focus of this paper but are presented in the Appendix C.

4.3.3. Optimization of RBF parameters through genetic algorithm

The Genetic Algorithm consists of a probabilistic method of optimization whose principle is to determine the best solution (optimal solution) within a space of results [117, 118]. The behavior of the genetic algorithm is shown in Figure 4.4. In this figure, an initial population is evaluated into an objective function. The optimization ends when the tolerable error of 1E-04 (stop criterion) with respect to the previous solution is reached. The objective function considered in this paper consists of the PRESS minimization process presented in the previous section. The present work considers the use of genetic algorithm functions from MATLAB™ to minimize the objective function.

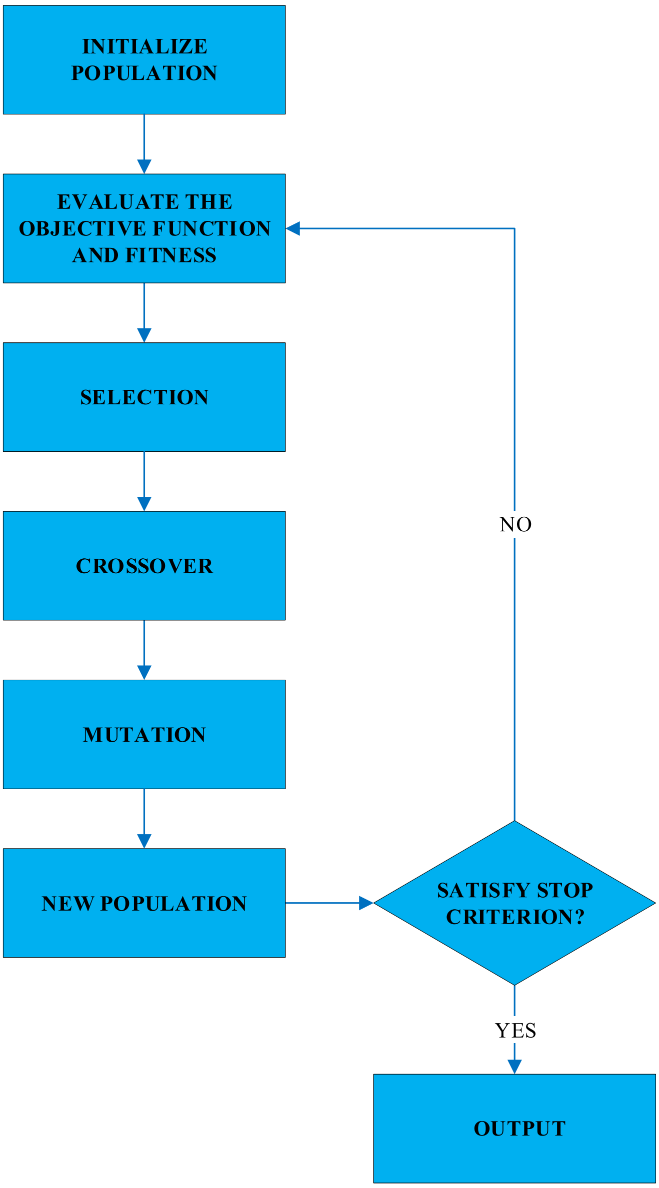


Figure 4.4 – Flowchart of the behavior of the genetic algorithm.

Future generations are then obtained through the steps of selection, crossover, and mutation. The selection consists of selecting certain chromosomes (usually binary numbers that encode the variables) that will contribute to the reproduction of the population. This reproduction is achieved by the crossover operator and the mutation consists of small changes in the chromosomes to ensure a greater population variety [118]. More details on the behavior of the genetic algorithm are presented in the literature [119, 117, 118, 120].

4.3.4. Performance evaluation of uncertainty quantification and reliability analysis

Once the surrogate models were trained and validated, the reliability analysis was performed through the MC simulations also using the LHS sampling method. MC simulations are chosen due to the easy implementation, popularity, and convenience that serve the purpose of this paper and are not centered on the efficiency of random simulation techniques. It is also important to emphasize that any other random number generator and/or sampling can be used in the reliability analysis. The definition of the size of the sample sets used in the MC simulation were obtained through a convergence analysis of the probability of failure. In this way, first, second and third numerical examples considered 1 E+06 samples in the MC simulation. However, the only exception here is for the last example which considered only 1 E+05 samples. This number of samples is different because of the convergence of the results demonstrated in the reliability analysis that there is no need for a larger sample set.

The samples obtained from the MC simulation were used first to predict analytically the solution of each numerical example. These results are referred to here as the Monte Carlo reference solution (MC^{REF}) and these results were then plotted to quantify the uncertainty of the limit state function of each numerical example. The samples obtained from the MC simulation were also used to validate the surrogate models as can be seen in the green boxes of Figure 4.1. These predicted results were then used in the reliability analysis. In this paper, the reliability analysis consists of the main study to be carried out with the surrogate models. The structural reliability analysis was done through the probability of failure (P_f) of a performance function $G(\boldsymbol{\xi})$ that occurs for values $G(\boldsymbol{\xi}) < 0$ [84, 85, 2]. The probability of failure is defined as:

$$P_f \left\{ G(\xi) < 0 \right\} = \int_{G(\xi)<0} \int \cdots \int f(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 d\xi_2 \dots d\xi_n \tag{4.10}$$

in which $f(.)$ is the multivariate probability density function (PDF) built from the input's variables $\boldsymbol{\xi}$. The probability of failure built from the results of the predicted surrogate models is then compared to the P_f value built into the reference model (MC^{REF}). The coefficient of variation [86] of the estimated probability of failure is defined as:

$$V_{Pf} = \sqrt{\frac{(1-P_f)}{p(P_f)}} \tag{4.11}$$

Another analysis consists of the reliability index β [10, 87]. The reliability index is defined as:

$$\beta = -\Phi^{-1}(P_f) \tag{4.12}$$

in which Φ^{-1} is the inverse of the standard normal cumulative distribution function. The reliability index obtained from the surrogate models is compared with the reliability index from the reference solution. These verifications were done in all the numerical examples.

4.4. NUMERICAL EXAMPLES

This section presents the numerical examples selected from the literature to be predicted with RBF-based surrogate models of type Gaussian (GA), Wendland-C2 (WE-C2), and the new Almeida-Evangelista Jr (AEJ-FT). The Example 01 and Example 02 consist of two limit state analytical functions and the Example 03 and Example 04 consist of physical models applied to structural engineering. The third function (Example 03) consists of a performance function analysis for the maximum deflection of a cantilever beam. Finally, the last example (Example 04) aims to analyze the dynamic response of an oscillator. The shape parameters of the RBFs for all four numerical examples are shown in Table 4.1.

Table 4.1 – Radial Basis Function parameter values

Example	Surrogate	Parameters
01	RBF (GA)	ρ = 0.10
	RBF (WE-C2)	ρ = 11.0
	RBF (AEJ-FT)	ρ = 7.00
		κ = 1.00
		β_0 = 0.01
02	RBF (GA)	ρ = 0.10
	RBF (WE-C2)	ρ = 12.0
	RBF (AEJ-FT)	ρ = 8.70
		κ = 2.00
		β_0 = 0.01
03	RBF (GA)	ρ = 0.10
	RBF (WE-C2)	ρ = 11.00
	RBF (AEJ-FT)	ρ = 4.00
		κ = 3.00
		β_0 = 0.01
04	RBF (GA)	ρ = 0.10
	RBF (WE-C2)	ρ = 11.00
	RBF (AEJ-FT)	ρ = 7.00
		κ = 1.00
		β_0 = 0.01

4.4.1. Example 01: Nonlinear Polynomial Function

The example is a nonlinear performance function studied by the literature [121, 99] with two random variables and is defined as:

$$G(x) = -18 + x_1^3 + x_1^2 x_2 + x_2^3 \tag{4.13}$$

The random variables are independent and normally distributed with mean (μ) and standard deviation (σ) defined in Table 4.2. This example was selected because it is easy to reproduce and is constantly explored by the literature.

Table 4.2 – Random variables of Example 01

Var	μ	σ	$f(x)$
x_1	10.00	5.00	Normal
x_2	9.90	5.00	Normal

In Table 4.2 the respective μ of variables x_1 and x_2 have practically similar magnitudes. The number of samples in the MC simulation was obtained by a convergence of the surrogate models' techniques and the analytical solution. This analytical solution consists of evaluating Equation (4.13) using the MC samples, defined here as MC^{REF}. Figure 4.5 depicts the analytical solution of MC^{REF} in which is defined $G(x) < 0$ as the failure region, $G(x) > 0$ as the confidence or safe region, and $G(x) = 0$ as the limit state.

Note that the domain $G(x) < 0$ is relatively smaller than the domain $G(x) > 0$, thus requiring a large number of simulations to construct an accurate solution for the reliability analysis. The variables x_1 and x_2 have similar behavior, and for both, the failure zone is constructed for combinations between ($-\mu - \sigma$) and $\mu + \sigma$. It is also important to emphasize that for any combination of random variables greater than 2μ, the non-randomness of samples will not cause the failure of the performance function. The limit state in Figure 4.5 is constructed for a range of values of $G(x)$ between -0.7 and 0.7 which consist of magnitudes of $G(x)$ tending to zero ($G(x) \cong 0$). The failure region is constructed with $G(x) < -0.7$ and the safe region shows values of $G(x) > 0.7$.

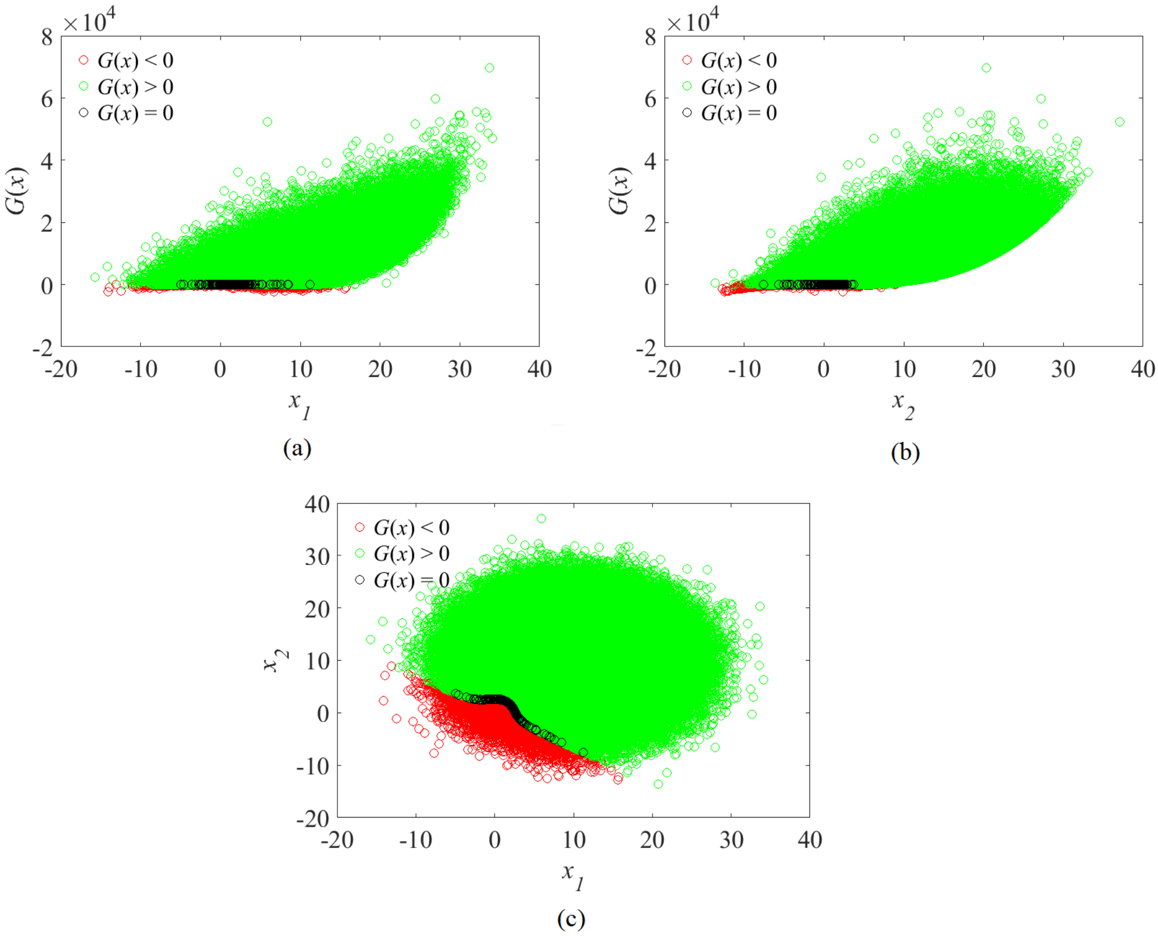


Figure 4.5 – MC^{REF} simulation for Example 01: (a) x_1 *versus* $G(x)$, (b) x_2 *versus* $G(x)$, (c) x_1 *versus* x_2.

The probability of failure estimate of the analytical reference solution (MC^{REF}) and the surrogate model's approximations are presented in Table 4.3. Column p represents the number of calls for the analytical performance function. The p-value for the surrogate models consists of the sampling number used in the training process of the surrogate models (see item 4.3.1). The definition of sample sizes for training the surrogate model was performed considering the initial sample size as being equal to the size adopted by the methods from the literature that evaluated the Example 01. Then, a convergence test was carried out with the objective of get a smaller training sample size. The method RBFN – RSM2 consists of an RBF neural network response surface method of Tan and co-authors [121] using FORM and Monte Carlo simulation to obtain the reliability index. The method A-RBF consists of an adaptive RBF technique of Wang and Fang [99] using FORM to estimate the reliability index. In both methods, the authors only determined the reliability index β. Here, it is used Equation (4.12) to have a P_f value for comparison purposes. The coefficient of variation of the probability of failure (V_{P_f}) for the MC^{REF} solution is 0.01.

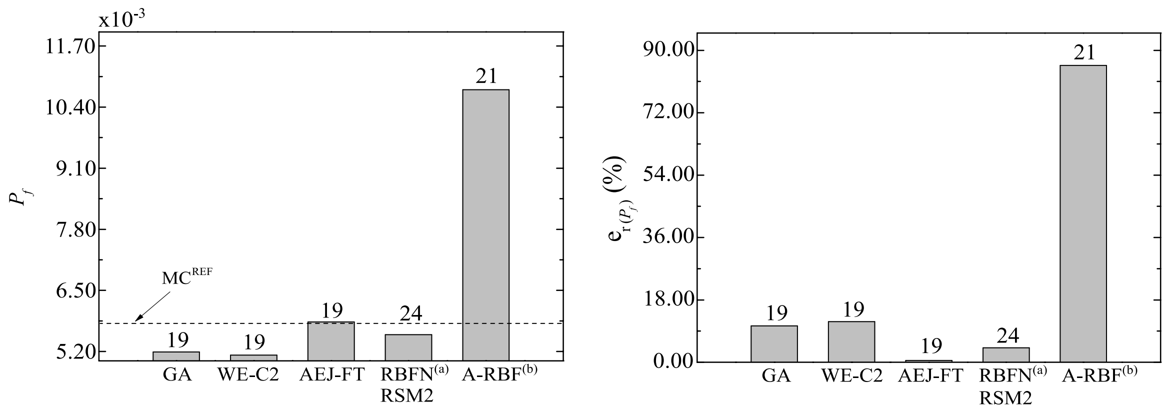
Table 4.3 – Probability of failure estimative of Example 01

Method	P_f	e_r (%)	V_{P_f} (%)	p
RBF (GA)	5.190 E-03	10.517	1.384	19
RBF (WE-C2)	5.120 E-03	11.724	1.394	19
RBF (AEJ-FT)	5.830 E-03	0.517	1.306	19
RBFN-RSM2 [(a)]	5.559 E-03	4.155	1.337	24
A-RBF [(b)]	1.077 E-02	85.690	-	21
MC^{REF}	5.800 E-03	-	1.309	1 E+06

[(a)] Reference [121]; [(b)] Reference [99].

Table 4.3 shows that the RBF (AEJ-FT) approximation is more accurate than the RBFN-RSM2 and A-RBF methods. It is also noted that the model A-RBF which has a smaller number of training samples to construct the surrogate model when compared to RBFN-RSM2, has a larger relative approximation error. The RBF (GA) and RBF (WE-C2) show relative error inferior to the A-RBF technique, but values of relative error are superior to 10%.

An additional observation to be mentioned here is the verification of the processing cost. The referenced methods in the literature do not make a specific study of the computational time for Example 01. A viable solution to verify the processing time will be given as a function of the number of radial bases to be built by the RBF surrogate model during the training phase. Each base function is built with one sample of the training set and therefore it is possible to determine a relative cost for the surrogate models as a function of the number of training samples. The both methods in the literature (A-RBF and RBFN-RSM2) are methods based on radial basis functions and becomes interesting to analyze that the number of bases to build the model has a certain computational cost. Consequently, there is a reduction in the computational demand for the RBF (GA), RBF (WE-C2) and RBF (AEJ-FT), since they present a reduction of approximately 10% in the number of evaluations of the training set when compared to the A-RBF method; and for the RBFN-RSM2 method a reduction of around 21%. So, it appears that the RBFs have a lower computational cost when compared to the A-RBF and RBFN-RSM2 methods. Another analysis of Table 4.3 is presented in Figure 4.6.



a) Probability of Failure
b) Relative Error

Figure 4.6 – Failure analysis of Example 01: a) Probability of Failure; b) Relative Error

Figure 4.6a show that the predicted P_f values of the RBF (GA) and the RBF (WE-C2) are distant from the MC^{REF}. However, RBF (AEJ-FT) and RBFN-RSM2 are the ones that best represent the failure analysis. In Figure 4.6b, the relative error of RBF (AEJ-FT) is much lower than the rest of the surrogate modeling techniques. The reliability index of surrogate models is presented in Table 4.4.

Table 4.4 – Reliability index estimative of Example 01

Method	$\boldsymbol{\beta}$	**e$_r$ (%)**	$\boldsymbol{p}$
RBF (GA)	2.563	1.545	19
RBF (WE-C2)	2.568	1.743	19
RBF (AEJ-FT)	2.522	0.079	19
RBFN-RSM2 [(a)]	2.539	0.594	24
A-RBF [(b)]	2.298	8.954	21
MC^{REF}	2.524	-	1 E+06

[(a)] Reference [121]; [(b)] Reference [99].

In Table 4.4, the reliability index of Wang and Fang [99] was obtained by the FORM method and thus presents a greater approximation error to the reference solution from MC^{REF}. It is also observed that the RBF (AEJ-FT) is the surrogate model with the lowest relative error. The RBFN-RSM2 of Tan and co-authors [121] considered MC simulation and the FORM method to obtain the reliability index. The result of RBFN-RSM2 with the FORM method is 2.289, in other words, shows a relative error superior to 9 %. Thus, it can be concluded that MC simulation shows a better approximation of the exact solution of reliability analysis for RBFN-RSM2. The traditional RBF (GA) and RBF (WE-C2) show a relative error inferior to 2%.

4.4.2. Example 02: Exponential Function

In this example, an exponential state limit function [122, 84, 102] is expressed as:

$$G(x) = \exp\left(0.2x_1 + 1.4\right) - x_2 \tag{4.14}$$

in which x_1 and x_2 are random variables with mean and standard deviations showed in Table 4.5. Although this example is simpler, it is easy to reproduce and serves as a primary analysis of the new RBF. This same example is also widely used in the literature to test surrogate models.

Table 4.5 – Random variables of Example 02

Var	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	$\boldsymbol{f(x)}$
x_1	0.00	1.00	Normal
x_2	0.00	1.00	Normal

The probability of failure and reliability index were constructed for the surrogate models and the MC^{REF} reference solution with 1 E+06 samples. The analytical solution consists of evaluating Equation (4.14) using the MC samples, defined here as MC^{REF}. The behavior of the MC^{REF} limit state is presented in Figure 4.7. The limit state in Figure 4.7 is constructed for a range of values of $G(x)$ between -0.03 and 0.03 which consist of magnitudes of $G(x)$ tending to zero ($G(x) \cong 0$). The failure region is constructed with $G(x) < -0.03$ and the safe region shows values of $G(x) > 0.03$.

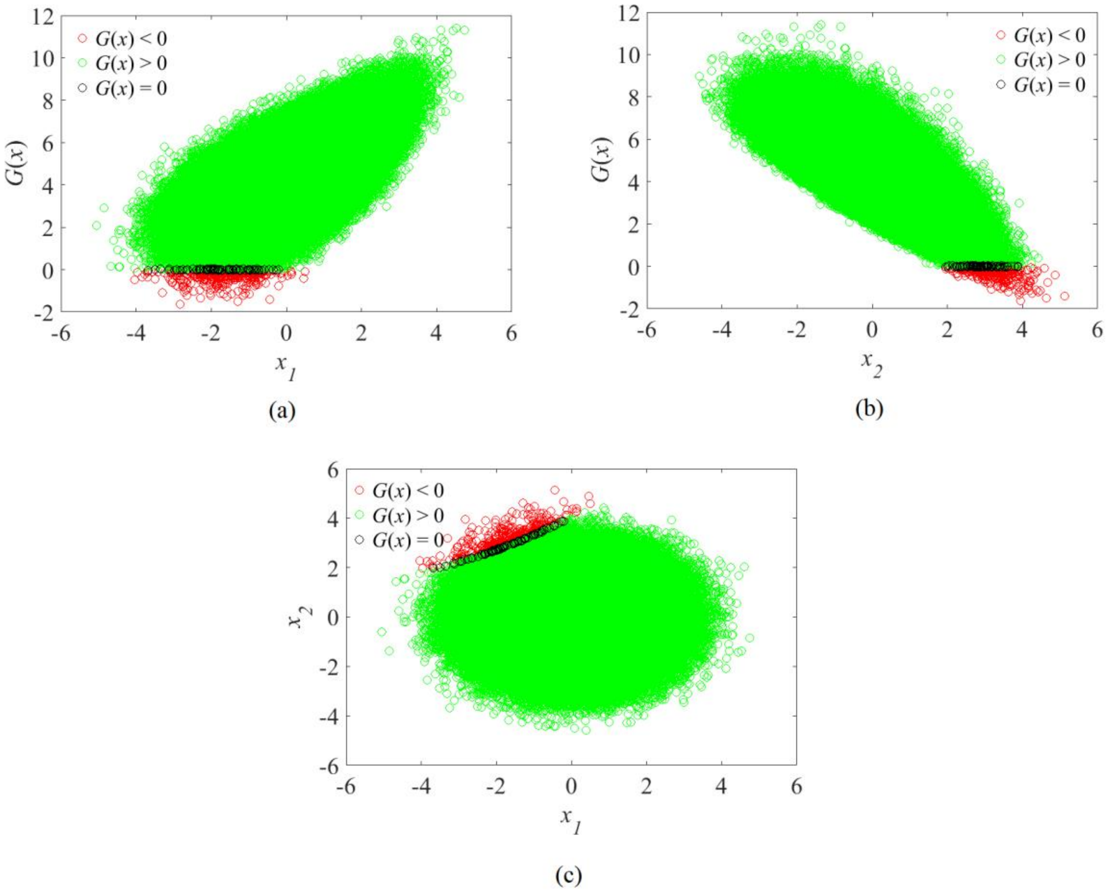


Figure 4.7 – MC^{REF} simulation for Example 02: (a) x_1 *versus* $G(x)$, (b) x_2 *versus* $G(x)$, (c) x_1 *versus* x_2.

Figure 4.7 show that the random variables x_1 and x_2 present a contrary linear behavior with a high dispersion of the safe region, which is responsible for the high-reliability index of the function. The failure region is expressed by the variable x_1 in the range of [-4, 1] and for the variable x_2 in the range of [2, 6] approximately. However, for values of $x_1 > 2$ and $x_2 < 0$, the function will always be in the safe region. The results of P_f values are presented in Table 4.6. The method ANN consists of an artificial neural network technique used by Kroetz and co-authors [102] to obtain the reliability index. The methods KRIG and PCE correspond to Kriging and the 4th degree Polynomial Chaos Expansion, respectively. In both methods, the authors only determined the reliability index β. Here, it is used Equation (4.12) to have a P_f value for comparison purposes. The coefficient of variation of the probability of failure (V_{P_f}) for the MC^{REF} solution is 0.05. The size of the training sets considered for the RBFs was obtained through a convergence test. The convergence started with a sample size similar to the literature and it was found at the end of the convergence that the minimum sample size for RBFs was $p = 12$. In this paper, emphasis will not be given to the convergence process, as it is something that has already been extensively explored in benchmark problems by the literature.

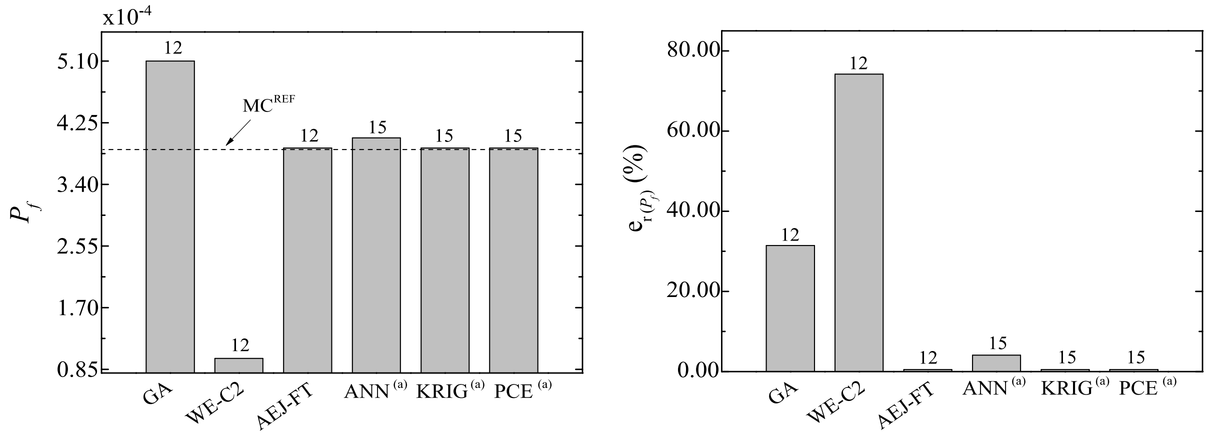
Table 4.6 – Probability of failure estimative of Example 02

Method	P_f	**e**$_r$ **(%)**	V_{P_f} **(%)**	p
RBF (GA)	5.10 E-04	31.44	4.43	12
RBF (WE-C2)	1.00 E-04	74.23	10.00	12
RBF (AEJ-FT)	3.90 E-04	0.52	5.06	12
ANN [(a)]	4.04 E-04	4.12	4.97	15
KRIG [(a)]	3.90 E-04	0.52	5.06	15
PCE [(a)]	3.90 E-04	0.52	5.06	15
MC^{REF}	3.88 E-04	-	5.08	1 E+06

[(a)] Reference [102].

Table 4.6 presents RBF (AEJ-FT) as a competitive surrogate model to the ANN, KRIG, and PCE surrogate models with a relative error inferior to 1%. However, the RBF (GA) presents a high relative error and this error is superior to 30%. The RBF (WE-C2) also shows that the relative error is more than twice the RBF (GA), and thus, does not show a good failure approximation when compared to MC^{REF}. As for the computational time required to build the KRIG, PCE and ANN; the literature did not present a study for Example 02. Therefore, an analysis of the computational cost similar to that adopted in Example 01 will be considered. In this case, it can be considered that there is a lower computational cost for the RBFs, as there is a decrease of around 20% in the number of samples to be evaluated (in the training step) by the RBFs when compared to the surrogate models from the literature (KRIG, ANN and PCE). This analysis becomes even more interesting when one observes that the KRIG method because it uses Euclidean distance quite similar to that of the RBFs. In this way, there is an indirect comparison by the number of samples to build the basis functions. The KRIG method requires a greater number of bases functions to be build and consequently a greater computational effort.

During the validation phase in which the MC samples are used, the economy in the number of bases to be built by the RBFs is greater. The reduction is of 2 E+05 bases to be constructed by the RBFs when compared to the ANN, KRIG and PCE methods. Consequently, there is a relevant computational economy, since the UQ process via MC requires a lot of computational demand. Another analysis of failure is presented in Figure 4.8.



a) Probability of Failure b) Relative Error

Figure 4.8 – Failure analysis of Example 02: a) Probability of Failure, b) Relative Error

Figure 4.8a shows that RBF (GA) and RBF (WE-C2) surrogate models cannot adequately predict failure, requiring a larger number of training samples ($p > 12$) to represent the MC^{REF} behavior. The approximation error of the RBF (AEJ-FT), presented in Figure 4.8b, is much lower than RBF (GA) and RBF (WE-C2), presenting value similar to ANN, KRIG and PCE. The relative error of the ANN showed superior to KRIG, and PCE. The reliability index analysis is presented in Table 4.7.

Table 4.7 – Reliability index estimative of Example 02

Method	$\boldsymbol{\beta}$	**e$_r$ (%)**	$\boldsymbol{p}$
RBF (GA)	3.28	2.38	12
RBF (WE-C2)	3.72	10.71	12
RBF (AEJ-FT)	3.36	0.00	12
ANN [(a)]	3.35	0.30	15
KRIG [(a)]	3.36	0.00	15
PCE [(a)]	3.36	0.00	15
MC^{REF}	3.36	-	1 E+06

[(a)] Reference [102].

Table 4.7 shows that RBF (AEJ-FT) presents a better accuracy when compared to the proposed surrogate models with practically zero approximation error. The number of decimals considered for the reliability index of Example 02 is the same used by the literature in a direct comparison. The RBF (GA) surrogate model showed a relative error of less than 3% with $p = 12$ and the RBF (WE-C2) surrogate models presented a relative error superior to 10%. In this way, it is also observed that in the reliability index, the RBF (WE-C2) does not show a good approximation in relation to the MC^{REF} analytical solution.

4.4.3. Example 03: Cantilever beam

A structural engineering problem consist of a cantilever beam with a rectangular cross section ($b_w \times h_w$) and uniformly distributed load (w). The cantilever beam is showed in Figure 4.9.

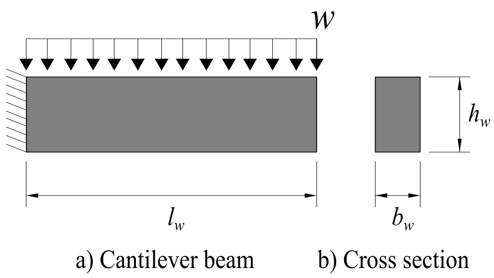


Figure 4.9 – Cantilever beam with uniform distributed loading: a) Cantilever beam, b) Cross section

Example 03 is constantly studied in the literature, especially by researchers [10, 123, 30]. The built and comparison of the surrogate models adopted the performance function from Li and co-authors [30] and is given as:

$$G(x) = \frac{l_w}{325} - \frac{w b_w l_w^4}{8EI} \tag{4.15}$$

in which, l_w is the span length, E is the elasticity modulus, and $I = b_w h_w^3/12$. The state limit function is defined by deflections in the cantilever beam greater than l_w / 325 and the output response is given in millimeters (mm) [30]. The variables l_w and E are deterministic with values presented in Table 4.8.

Table 4.8 – Deterministic variables of Example 03

Var	**Unit**	**Value**
l_w	mm	6.00 E+03
E	Mpa	2.60 E+04

Table 4.9 shows the random variables that are w and h_w and these variables are rewrite as x_1 and x_2 respectively.

Table 4.9 – Random variables of Example 03

Var	**Unit**	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	$\boldsymbol{f(x)}$
x_1	N/m²	1000.0	200.0	Normal
x_2	mm	250.0	37.5	Normal

Substituting the random and deterministic variables of Table 4.8 and Table 4.9, the real performance function to be analyzed is:

$$G(x) = 18.46154 - 74769.23 \frac{x_1}{x_2^3} \tag{4.16}$$

The probability of failure and the reliability index are obtained from 1 E+06 samples through MC simulation. The analytical solution consists of evaluating Equation (4.16) using the MC samples, defined here as MC^{REF}. The MC^{REF} limit state is presented in Figure 4.10. The limit state in Figure

4.10 is constructed for a range of values of $G(x)$ between -0.1 and 0.1 which consist of magnitudes of $G(x)$ tending to zero ($G(x) \cong 0$). The failure region is constructed with $G(x) < -0.1$ and the safe region shows values of $G(x) > 0.1$.

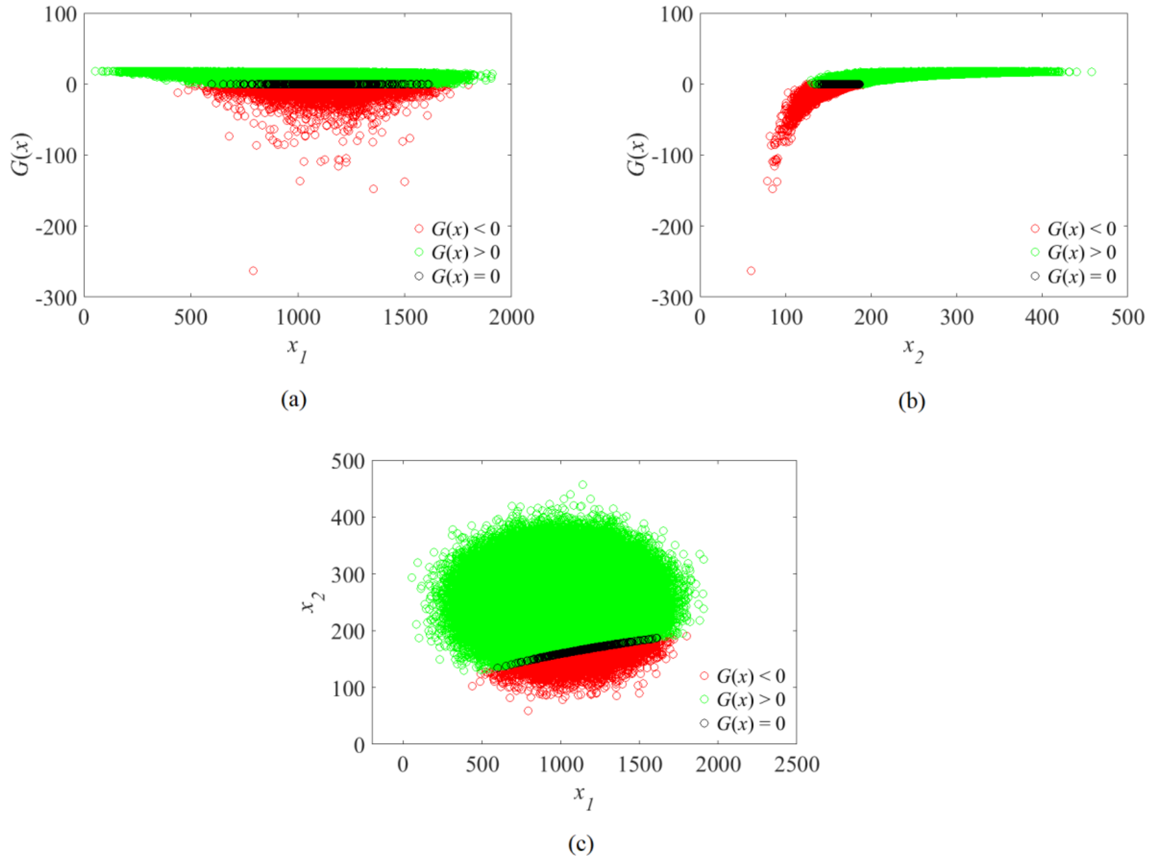


Figure 4.10 – MC^{REF} simulation for Example 03: (a) x_1 *versus* $G(x)$, (b) x_2 *versus* $G(x)$, (c) x_1 *versus* x_2.

The limit state of MC^{REF} is defined by the variable x_1, in other words, by the load that acts in cantilever beam, which presents a greater variability and uncertainty of the combinations of this variable in practically the entire domain of the variable. However, for the variable x_2, beam height (Figure 4.10b), the limit state is checked for magnitudes between 100 and 200 mm. It is important to emphasize that for beam heights greater than μ and any loading combinations, the performance function will not be in the failure region. The failure of the performance function is presented in Table 4.10. The SSRM consists of a sequential surrogate reliability method by Li and co-authors [30] using MC simulation to obtain the probability of failure P_f. The methods FORM and SORM are also used by the authors and they are not incorporated in the SSRM. These two methods were used for comparison with the SSRM. The authors determined for FORM and SORM only the probability of failure, being the reliability index determined by Equation (4.12) as shown in Table 4.11. The DS+NN corresponds to a method using directional sampling with neural networks with more calls than the other methods presented here. The definition of sample sizes for training the RBFs was performed considering the initial sample size as equal to the size adopted by the methods from the literature that evaluated the Example 03. Then, a convergence test was carried out with the objective of get a smaller training sample size (p = 15). The coefficient of variation of the probability of failure (V_{P_f}) for the MC^{REF} solution is 0.01.

Table 4.10 – Probability of failure estimative of Example 03

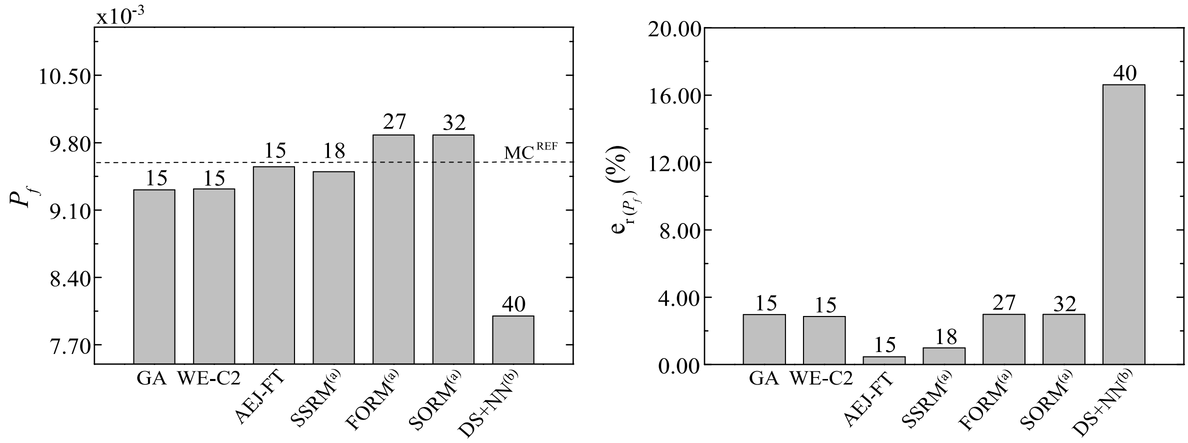
Method	$\boldsymbol{P_f}$	**e_r (%)**	**V_{P_f} (%)**	$\boldsymbol{p}$
RBF (GA)	9.310 E-03	2.960	1.032	15
RBF (WE-C2)	9.320 E-03	2.856	1.031	15
RBF (AEJ-FT)	9.550 E-03	0.459	1.018	15
SSRM [(a)]	9.499 E-03	0.990	1.021	18
FORM [(a)]	9.880 E-03	2.981	-	27
SORM [(a)]	9.880 E-03	2.981	-	32
DS+NN [(b)]	8.000 E-03	16.615	-	40
MC^{REF}	9.594 E-03	-	1.016	1 E+06

[(a)] Reference [30]; [(b)] Reference [123].

The RBF (AEJ-FT) approximation presents a good accuracy for a smaller number of training samples (p =15) than the surrogate models of Li and co-authors [30] and the relative error of the RBF (AEJ-FT) is less than 0.50% in relation to MC^{REF} as can be seen in Table 4.10. It is also observed that the RBF (AEJ-FT) is competitive, presenting a better approximation in relation to SORM and DS+NN. However, the traditional RBF (GA) and RBF (WE-C2) surrogate models contain a similar approximation error when compared with the Li and co-authors [30] results, but the number of training samples (p = 15) is smaller than the used by the surrogate models from the literature.

The computational effort is verified in a similar way to the previous examples, since the literature did not verify the CPU time for Example 03. Therefore, it is observed that RBFs have a lower computational cost because they present a reduction of around 17% of the number of samples to be evaluated (number of bases to be build) during the training phase when compared to the SSRM and a reduction of around 63% when compared to the DS+NN method. Thus, there is a lower computational effort for the RBFs when compared to the surrogate models from the literature.

In the uncertainty quantification process, the RBFs and the SSRM method use 1 E+06 samples to determine the probability of failure. However, the RBFs have a reduction of 1.7 E+05 bases to be built when compared to SSRM. This reduction causes a relevant computational economy for the RBFs, since the UQ process via MC requires a lot of computational demand. For the DS+NN method, it is not possible to compare the cost of processing time, as the method uses directional sampling instead of Monte Carlo simulation. The results from Table 4.10 can also be seen in Figure 4.11.



a) Probability of Failure b) Relative Error

Figure 4.11 – Failure analysis of Example 03: a) Probability of Failure, b) Relative Error.

Figure 4.11a shows that the RBF (AEJ-FT) is competitive with SSRM. However, the traditional FORM and SORM methods are far from the reference MC^{REF}. In Figure 4.11b, the neural network method shows a relative error much higher than the rest of the surrogate model techniques. In Figure 4.11b, it can also be noted that the RBF (GA) and RBF (WE-C2) present relative errors similar to FORM and SORM, and thus, competitive to the classical methods. The reliability index is presented in Table 4.11.

Table 4.11 – Reliability index estimative of Example 03

Method	β	**e_r (%)**	p
RBF (GA)	2.353	0.478	15
RBF (WE-C2)	2.353	0.461	15
RBF (AEJ-FT)	2.344	0.073	15
SSRM [(a)]	2.346	0.159	18
FORM [(a)]	2.331	0.469	27
SORM [(a)]	2.331	0.469	32
DS+NN [(b)]	2.420	3.337	40
MC^{REF}	2.342	-	1 E+06

[(a)] Reference [30]; [(b)] Reference [123].

In Table 4.11, the RBF (AEJ-FT) is competitive to the traditional methods and has a relative error inferior a 0.10%. The RBF (GA) and RBF (WE-C2) show a relative error inferior to 0.50%. It can also be noted that the RBF (WE-C2) shows a better approximation than the traditional methods FORM and SORM. The method DS+NN has a high relative error value; however, it should not be disregarded as it is competitive in relation to the other surrogate models.

4.4.4. Example 04: Dynamic response of an oscillator

Consider a non-damped oscillating system of a single degree of freedom [10, 123, 30]. The performance function is defined for six normal random variables:

$$G(x) = 3r_0 - \left| \frac{2F_1}{m_0 \omega_0^2} \sin\left(\frac{\omega_0 t_1}{2} \right) \right| \tag{4.17}$$

with:

$$\omega_0 = \sqrt{\frac{\bar{c}_1 + \bar{c}_2}{m_0}} \tag{4.18}$$

in which, m_0 is the mass of the oscillator body, r_0 is the displacement at which one of the springs yields, $z(t)$ is the maximum displacement response of the system, F_1 is the amplitude force, $\bar{c}_1$ and $\bar{c}_2$ are spring constants and t_1 is the acting time. Figure 4.12 shows the oscillator system to be considered in the performance function.

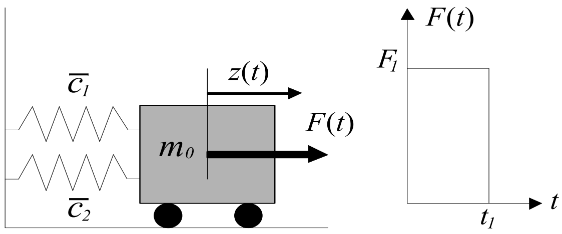


Figure 4.12 – Oscillator definition and applied load [123]

Figure 4.12 shows that the system consists of two springs coupled to a mass body m_0 that depends on the time and amplitude force. Table 4.12 presents the random variables and their respective probability distributions.

Table 4.12 – Random variables of Example 04

Var	μ	σ	$f(x)$
m_0	1.00	0.05	Normal
$\bar{c}_1$	1.00	0.10	Normal
$\bar{c}_2$	0.10	0.01	Normal
r_0	0.50	0.05	Normal
F_1	1.00	0.20	Normal
t_1	1.00	0.20	Normal

The random variables in Table 4.12 follow normal distributions. This example has also shown in the literature several works with different mean and standard deviation values for the random variables, such as, for example, different mean values for the amplitude force shown in Table 4.12 [124, 125, 126]. The probability of failure and the reliability index construction required in this paper is 1 E+05 samples. The analytical solution consists of evaluating Equation (4.17) using the MC samples, defined here as MC^{REF}. The results of the samples used to obtain analytically the reference solution MC^{REF} are presented in Figure 4.13. The limit state in Figure 4.13 is constructed for a range of values of $G(x)$ between -0.01 and 0.01 which consist of magnitudes of $G(x)$ tending to zero ($G(x) \cong 0$). The failure region is constructed with $G(x) < - 0.01$ and the safe region shows values of $G(x) > 0.01$.

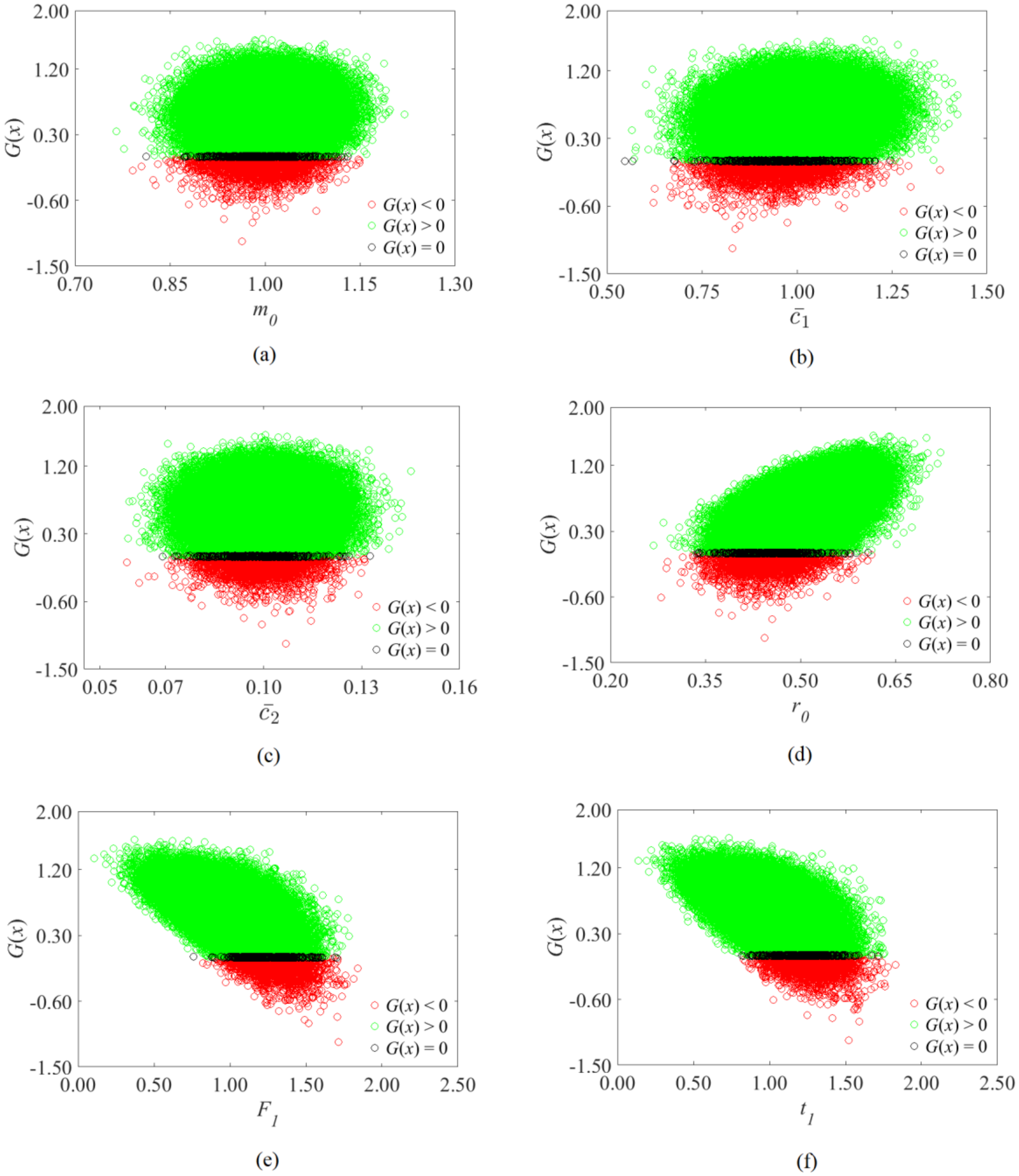


Figure 4.13 – MC^{REF} simulation for Example 04: (a) m_0 *versus* $G(x)$, (b) $\bar{c}_1$ *versus* $G(x)$, (c) $\bar{c}_2$ *versus* $G(x)$, (d) r_0 *versus* $G(x)$, (e) F_1 *versus* $G(x)$, (f) t_1 *versus* $G(x)$

Figure 4.13 shows the output response and the input random variables of the limit state function $G(x)$ for the MC^{REF} reference solution. In Figure 4.13a-c, a high dispersion of the variables m_0, $\bar{c}_1$, and $\bar{c}_2$ is observed and this randomness is a characteristic of the high uncertainty in the failure prediction of the function $G(x)$. In Figure 4.13d, values of $r_0 > 0.65$ will always be in a safe region for the performance function. In Figure 4.13e-f, the variables F_1 and t_1 show similar behavior, but contrary to the variable r_0. It can also be noted that for variables F_1 and t_1, the values inferior to 0.50 are in the non-failure region.

Table 4.13 presents the results of the probability of failure for the surrogate models with RBF (GA), RBF (WE-C2), and RBF (AEJ-FT). The method SSRM shown in Example 03 is used by Li and co-authors [30] to obtain only the probability of failure P_f. The method FORM is also used by the authors for comparison with the SSRM. The DS+NN as shown before is used by Schueremans and Van Gemert [123] to obtain the probability of failure. The reliability index β is obtained using Equation (4.12). The MC^{REF} analytical solution is also shown in Table 4.13. It is also important to emphasize that Li and co-authors [30] considered 7 E +04 samples in the MC^{REF} to obtain the analytical reference solution. The size of the training sets considered for the RBFs was obtained through a convergence test. The convergence started with a sample size similar to that adopted by the DS+NN method (p = 86) and it was found at the end of the convergence that the minimum sample size for the RBFs was p = 19. This sample size is similar to that considered by SSRM method. The convergence process is already been extensively explored by the literature and consequently will not be studied in Example 04. The coefficient of variation of the probability of failure (V_{P_f}) for the MC^{REF} solution is 0.02.

Table 4.13 – Probability of failure estimative of Example 04

Method	P_f	**e$_r$ (%)**	V_{P_f} **(%)**	p
RBF (GA)	2.892 E-02	2.011	1.832	19
RBF (WE-C2)	2.896 E-02	2.152	1.831	19
RBF (AEJ-FT)	2.850 E-02	0.529	1.846	19
SSRM [(a)]	2.880 E-02	1.587	2.195	19
FORM [(a)]	3.108 E-02	9.630	-	35
DS+NN [(b)]	2.800 E-02	1.235	-	86
MC^{REF}	2.835 E-02	-	1.851	1 E+05

[(a)] Reference [30]; [(b)] Reference [123].

The number of training samples considered by the RBF surrogate models is the same as Li and co-authors [30] in the SSRM method as can be seen in Table 4.13. However, the approximation obtained for the RBF (AEJ-FT) is superior to the literature. It is also noted that the relative error of the RBF (AEJ-FT) is much less than 1 %, confirming that the methodology used for the new RBF is accurate. In Table 4.13, the coefficient of variation of the SSRM method is greater, as the number of MC samples considered by the literature is 7 E+04.

The CPU time study was not verified by the literature for the SSRM and DS+NN methods for Example 04. Consequently, the verification will be carried out similarly to Example 03. The method DS+NN is a based on Neural Networks in which there is also a construction of bases for the surrogate model. The surrogate models based on RBF showed a reduction of the computational effort due to decrease of 78% in number of bases to be built by the RBFs when compared to the

DS+NN method. In contrast, the computational time required to evaluate the SSRM is considered here to be similar to that of the RBFs due to the fact that both are based on radial basis functions and have the same number of bases to be build in the training phase. The results from Table 4.13 are also presented in Figure 4.14.

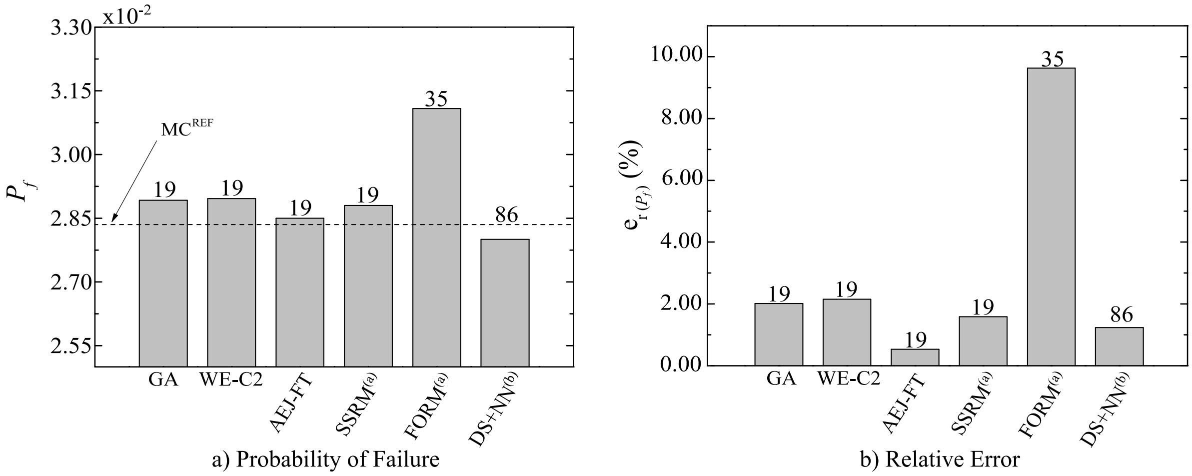


Figure 4.14 – Failure analysis for Example 04: a) Probability of Failure, b) Relative Error

Figure 4.14 shows that the FORM method has a higher error than the DS+NN adaptive surrogate model and the RBFs. It can also be noted that the RBF (AEJ-FT) is competitive with the methods SSRM and DS+NN. The RBF (GA) and RBF (WE-C2) present relative errors equal to 2% approximately. The performance of the reliability index (β) is presented in Table 4.14 for each surrogate model and is compared to other surrogate models from the literature.

Table 4.14 – Reliability index estimative of Example 04

Method	β	**e_r (%)**	***p***
RBF (GA)	1.897	0.472	19
RBF (WE-C2)	1.896	0.525	19
RBF (AEJ-FT)	1.903	0.157	19
SSRM [(a)]	1.899	0.367	19
FORM [(a)]	1.865	2.151	35
DS+NN [(b)]	1.910	0.210	86
MC^{REF}	1.906	-	1 E+05

[(a)] Reference [30]; [(b)] Reference [123].

Table 4.14 presents the reliability indices obtained by the RBF approximations and shows errors below 1%. The FORM method has a difference of around 2% from the reference model, requiring for the method a larger number of training samples to construct the reliability index. The RBF (GA) and RBF (WE-C2) have relative errors higher than SSRM and the RBF (AEJ-FT) shows the best result of the reliability index.

4.5. CONCLUSIONS

The present paper developed a surrogate modeling strategy for the reliability analysis of benchmark problems in which four numerical examples were selected from the literature. The first and the second numerical examples correspond to mathematical problems formulated as performance functions with two random variables. The third and the fourth examples consist of structural engineering problems. The third example corresponds to a study of a performance function of the maximum deflection of a cantilever beam and the last example has a performance function of a dynamic response of an oscillator. Three different types of surrogate models have been employed: two traditional RBF models (Gaussian and Wendland-C2) and a new CSRBF (Almeida-Evangelista Jr.). The surrogate models were trained to predict the output response of the numerical examples. The trained surrogate models were used to predict thousands of values from unseen data of Monte Carlo samples to the post-processing of the reliability analysis through the probability of failure and the reliability index.

The results of the reliability analysis demonstrated that the RBF surrogate models were able to predict the reliability index and the probability of failure of the MC^{REF} reference solution from the numerical examples. The first numerical example presented small relative approximation errors of the RBF surrogate models, principally to the RBF of type Almeida-Evangelista Jr. which has a probability of failure and reliability index values very similar to the reference model and better efficiency than the surrogate models from the literature. The second example showed that the probability of failure has only a competitive RBF surrogate model which was the RBF of type Almeida-Evangelista Jr. which demonstrated a high efficiency. On the other hand, the traditional RBFs (Gaussian and Wendland-C2) only have a small relative approximation error on the prediction of the reliability index. This behavior demonstrated the difficulty to predict the probability of failure of the second numerical example. The third example also presented a better approximation of the new RBF when compared to the other traditional RBFs and these results of reliability analysis also were competitive with the other surrogate models from the literature. Finally, the last example presented small relative approximation errors when compared to the traditional FORM and MC reference solution. The last example also had a better approximation to the new RBF and had a competitive performance to the traditional RBF surrogate models.

Another important study of this paper consisted in the uncertainty quantification of the limit state function. This study focused only on the MC^{REF} samples on which, a large variability of the random variables of each performance function was observed. Each performance function demonstrated that the random variables had a high dispersion and, consequently, a high influence on the failure region, and the combinations that caused the failure was therefore presented in practically the entire domain of input variables. Furthermore, the models based-RBF also demonstrated that these models are competitive surrogate model to be used in the reliability analysis, principally because, in all the numerical examples, the number of training samples used to build the surrogate models is smaller than the other surrogate models from the literature and the new RBF type Almeida-Evangelista Jr. had a better efficiency than the well-known traditional models based-RBF. Although the examples consider benchmark problems from the literature, the methodology presented in this paper, as well as the surrogate models with RBFs, can be applied in other structural engineering problems, principally the new RBF Almeida-Evangelista Jr. that are a tool of a high potential to be considered in engineering problems.

5. SURROGATE MODELS BASED ON RBF WITH REGULARIZATION AND GENETIC ALGORITHM IN UNCERTAINTY QUANTIFICATION PROCESS

5.1. INTRODUCTION

Surrogate model techniques have gained more importance to approximate structural engineering problems over the years [10, 24, 30, 105]. These models consist of techniques with the objective to approximate the output that is usually obtained by a high computational cost or that demand an exhaustive experimental study. Thus, surrogate models can approach the solution of these problems quickly and efficiently. Among the various techniques, the use of Radial Basis Functions (RBFs) stands out, which has shown a high efficiency in the analysis of several structural engineering problems [127, 5, 105]. This method presents a considerable difference in relation to other types of surrogate models, especially compared to multivariate polynomial regression models, because for the polynomial regression models, which require a high number of coefficients and, consequently, a prohibitive training set [5]. In the paper of Evangelista Jr. and Almeida [5], an age and time-dependent uncertainty quantification process was performed in fracture mechanics. The paper considered the uncertainty quantification surrogate models with RBFs to approximate the energy release rate in a concrete plate with an initial crack. The results showed a good efficiency of the models with RBF that were able to represent an unseen sample set of the already trained model. Another important point to be mentioned in relation to the surrogate models is the determination of the respective coefficients for its construction. These coefficients can be obtained in different ways, but a solution that avoids overfitting and allows a better generalization in the presence of noise is the use of regularization process.

A surrogate model with overfitting consists of a model capable of adequately predicting the output response of the training set, but incapable of predicting a set of unseen data from the already trained model [128, 129, 130]. Thus, the regularization process, in addition to solving an old problem of surrogate models, which consists of the ill-conditioning of the RBF matrices, also allows a decrease in the generalization errors and avoids overfitting with the addition of a penalty parameter to the regression model [131, 20, 21, 132, 133, 134, 135]. The regularization technique originated with the initial work of Andre Tikhonov [136] to solve ill-conditioned problems around the 20th century [19]. However, it was only with Hoerl and Kennard [137] that ridge regression was introduced, and this method was used to solve ill-conditioning of linear regression problems [19]. In the case of models with RBFs, the use of regularized regression appear to be mathematically and computationally convenient, as this method allows a reduction in the size of the training set and, consequently, a less sensitive model [19]. In this way, several adaptive methods emerged for the RBF models with regularized regression, such as: regularized forward selection, RFS [20, 18]; lasso regression, LR [138, 139, 140, 141]; regularized orthogonal least squares, ROLS [21, 19]. The ROLS method adds regularization to the regularized orthogonal least squares (OLS) to be able to build small training sets for the RBFs.

One way to avoid overfitting in models with regularization and to reduce the generalization errors is to properly determine the parameters of the surrogate model during the training phase. These parameters are determined by minimizing prediction error indicators. Among the error indicators used in the literature, two deserve to be mentioned which are: prediction residual error sum of squares, PRESS [18, 142, 143, 5], and generalized cross-validation, GCV [131, 144, 145, 146, 147]. The first prediction error considers the use of leave-one-out cross-validation (LOOCV), in which a single sample is predicted considering parameters that are obtained (estimated) from the sampling set excluding the preselected sample. However, when the sample set size is large, LOOCV has a prohibitive computational cost [148]. A viable solution to decrease the computational time was obtained by Golub and coauthors [131] in regularized regression using the GCV. The GCV indicator is a version of the PRESS rotation-invariant [131]. One advantage of GCV over PRESS is the correction of the problem of the sample left out during cross-validation. This samples taken off from the training during the LOOCV can cause high values of PRESS and consequently high values of error in the prediction. The correction given by the GCV consists of multiplying a factor in each sample of the sample set [144]. In addition, the algorithms presented and implemented by the regularized adaptive methods may present limitations, as they are usually deterministic methods that have recruitment of the training set and generally have constant shape parameters that are not always the best shape parameters in order to be adopted in the surrogate models.

The construction of models with RBF has constantly used probabilistic techniques to determine the model parameters, such as the use of genetic algorithm [149, 150, 119, 151, 152]. In general, a genetic algorithm is defined as a method that is based on the evolution of populations and tends to widely explore the decision space with a large probability of obtaining (finding) an improved solution [153]. For RBFs, the main focus of using the genetic algorithm is to determine the training set of the RBFs [150, 154, 155, 127] and to determine the shape parameters [156, 157, 120, 158]. The work of Chen and coauthors [156] combines the ROLS algorithm with the genetic algorithm. The genetic algorithm is considered to determine the shape parameter and the regularization parameter. The algorithm used is verified in nonlinear modeling examples and in time series. The results obtained showed good efficiency with an excellent generalization of the properties shown in the examples used by the authors. The paper of Afiatdoust and Esmaeibeigi [120] used the genetic algorithm to determine the shape parameters of the RBFs in the collocation method. The results showed an excellent fit of the parameters for linear and nonlinear cases of ordinary differential equations with very small approximation error magnitudes.

In view of the arguments presented, it is possible to observe that there is still a great demand for strategies with surrogate models based on RBFs. This question is further accented by the fact that RBFs are usually added to other existing methods or surrogate models. In these models, the RBFs are treated as parameter correction or are used in artificial neural networks. However, in this paper the RBFs are treated as the main functions that approximates the response of a structural problem. Another differential presented in this publication is the use of a genetic algorithm to determine the parameters of the RBFs and mainly the regularization parameter. This differential is not new, but it is little explored in the area of regularized regression, since several methods with RBFs, such as those mentioned above, fix the parameters that will be used in the surrogate models, without, therefore, determining the best combinations of parameters that must be used in each example. In this way, the model can be overfitted and consequently does not allow a good generalization. The uncertainty quantification (UQ) is used here because it allows obtaining the entire spectrum of variation of the output variable (e.g., probability density functions, cumulative distribution functions, statistical moments). The UQ process is carried out for models based on RBF with regularized regression and for the RBF Almeida Evangelista Jr. of type flat-top (AEJ-FT). Models with RBF and regularized regression has very few works in literature on uncertainty quantification processes and the RBF (AEJ-FT) has not yet been studied in UQ process with cumulative distribution functions. The only study of RBF (AEJ-FT) was done in terms of structural reliability in Chapter 4 of this thesis.

The article presents three objectives and three contributions. The first objective of this paper is to develop an RBF-based surrogate model with ridge regression. The RBF shape parameters are obtained with the use of genetic algorithm. In the training phase, an optimization process of the PRESS and GCV functions is done with a genetic algorithm in order to avoid overfitting. After the training stage of the surrogate models, the uncertainty quantification of numerical examples selected from the structural engineering benchmark problems is performed. The first and main contribution is the development of a computationally innovative strategy to a new RBF (RBF AEJ-FT) to approximate structural engineering problems. In this contribution, the new RBF AEJ-FT is no record of publications that verify its behavior in regularized regression. The second objective of this work is to quantify the uncertainty of the output response of numerical examples. The second contribution is the use of cumulative distribution functions (CDFs) for the uncertainty quantification process. The CDF is considered here because it gives a better qualitative analysis of the results and it is possible to observe better the tails errors. The quantitative analysis of the results is given by methods based on distances from the CDFs (e.g., Kolmogorov-Smirnov test). The third and final objective is to increase the generalization power of surrogate models in order to avoid overfitting, especially in situations where the training set is very small. The last contribution consists of the use of three statistical errors. The errors measured are: mean square error (MSE), root mean square error (RMSE) and normalized mean square error (ε). These errors are usually used separately, which may hinder a better conclusion on the generalization power of the surrogate models. Finally, the results show that surrogate models based on RBF have a good capacity to generalize unseen data sets of the surrogate model already trained. The training sets for the examples were relatively small and the hypothesis tests showed that the RBFs of the compact support type were capable of predicting cumulative distribution functions statistically similar to the reference model.

The article is organized as follows: Section 5.2 presents a general formulation for surrogate models with RBF and ridge regression, notions of genetic algorithm and definitions of PRESS and GCV. In Section 5.3, the methodology is presented. Section 5.4 presents the numerical examples while Section 5.5 presents the conclusions and final considerations. At the end of the work, some appendices are presented with results obtained for the three numerical examples. These results in the appendices are not the main focus in this paper.

5.2. REGULARIZED SURROGATE MODEL BACKGROUND

Surrogate models consist of a set of modeling techniques that are generally built by adjusting (training) a generalized linear model using a set of training samples. This modeling technique is ideal for significantly reducing the computational time required to solve a physical or mathematical model. Figure 5.1 shows the flowchart using the radial basis function (RBF) technique for training and validation of the surrogate models, also called metamodels. The purple boxes indicate the process used for the training phase. In this phase, the surrogate models (blue box) learn from the training process for then apply the already trained models to predict new output responses in uncertainty quantification problems (orange boxes). Note that the validation samples can be generated by random sampling methods such as Monte Carlo (MC) or can consist of a set of experimental data different from the training set.

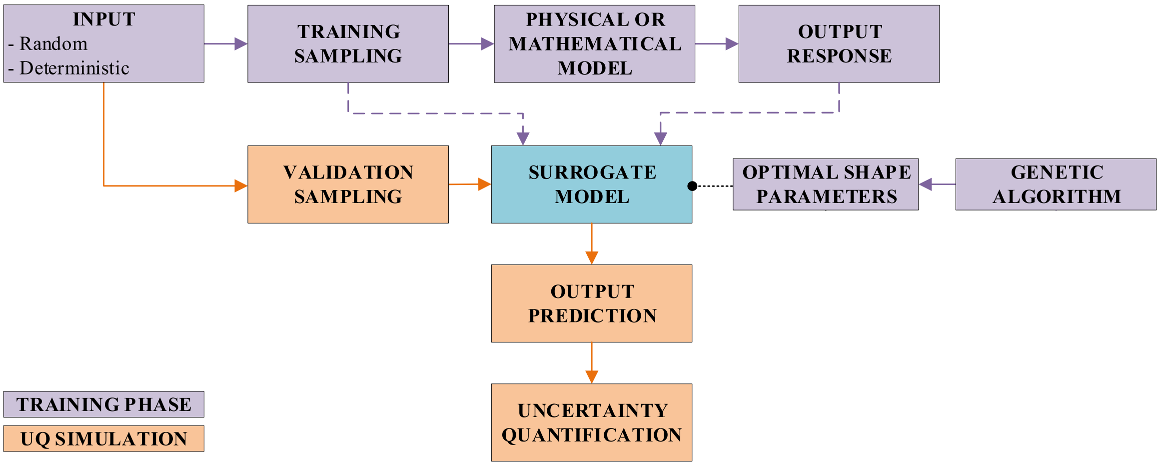


Figure 5.1 – Flowchart with the procedure for training and validation of the surrogate models

Surrogate models are regularly used to construct a predicted value $\hat{y}(.)$ of a given function $y(.)$ through a generalized linear model:

$$\hat{y}(\xi) = \sum_{j=1}^{m} \omega_j h_j(\xi) \tag{5.1}$$

in which the m regressors $h_j(\xi)$ are fixed functions of the variables $\xi \in \mathbb{R}^k$ with $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots \xi_k)$ corresponding to the vector of input variables and k is the number of input variables. The m unknown values are the $\boldsymbol{\omega}_j$ coefficients determined through a regression process based on the training set $\mathfrak{I} = \{(\xi_i, y_i) | \; i = 1, \ldots p\}$ being the output $y_i \in \mathbb{R}$ in the global domain (Ω). The quality of the fitting/training process is assessed by a set of unseen samples from the model or by a cross-validation procedure that divides the sample set into training samples and validation samples (data not known during training).

5.2.1. Radial Basis Function (RBF)

In the present article, the proposed surrogate models are given through a rewriting of Equation (5.1). The proposed surrogate model is given by:

$$\hat{y}(\xi) = \sum_{i=1}^{m} \omega_i \Psi_i(\xi, \rho) \tag{5.2}$$

in which $\Psi_i\ (\boldsymbol{\xi}, \rho)$ are the radial basis function (RBF) with a given length (or support) ρ, and ω_i their respective coefficients or weights. The RBFs considered in this work are not related to a neural network. The RBF is used here as a regression model in which the weights are determined by least squares. Equation (5.2) can also be expressed as:

$$\mathbf{y} = \mathbf{\Psi}\boldsymbol{\omega} + \mathbf{e} \tag{5.3}$$

in which **y** and **e** are *p*-dimensional vectors of the known output set and the errors, respectively; $\boldsymbol{\omega}$ is the *m*-dimensional vector of the unknown coefficients. The $p \times m$ design matrix $\mathbf{\Psi}$ contains the response of the *m* regressors obtained for the already known *p* input data. The linear system can be expressed as:

$$\begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{Bmatrix} = \begin{bmatrix} \Psi_1\left(\left\|\xi_1 - \xi_1^c\right\|\right) & \Psi_2\left(\left\|\xi_1 - \xi_2^c\right\|\right) & \cdots & \Psi_m\left(\left\|\xi_1 - \xi_m^c\right\|\right) \\ \Psi_1\left(\left\|\xi_2 - \xi_1^c\right\|\right) & \Psi_2\left(\left\|\xi_2 - \xi_2^c\right\|\right) & \cdots & \Psi_m\left(\left\|\xi_2 - \xi_m^c\right\|\right) \\ \cdots & \cdots & \cdots & \cdots \\ \Psi_1\left(\left\|\xi_p - \xi_1^c\right\|\right) & \Psi_2\left(\left\|\xi_p - \xi_2^c\right\|\right) & \cdots & \Psi_m\left(\left\|\xi_p - \xi_m^c\right\|\right) \end{bmatrix} \begin{Bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{Bmatrix} + \begin{Bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{Bmatrix} \tag{5.4}$$

in which the regressor Ψ_i is distinguished by its center ξ_i^c which corresponds to a combination of the training set $\Im$. In this work, the centers were established among the training samples to ensure efficiency ($m = p$). The training inputs (ξ) were normalized following a standard normal distribution (mean zero and standard deviation equal to 1). In Equation (5.4), Ψ_i processes the Euclidean norm (Euclidean distance) of each input data ξ to its respective center ξ_i^c trough the general formulation:

$$\Psi_i = \psi\left(\frac{\left\|\xi - \xi_i^c\right\|}{\rho}\right) \tag{5.5}$$

in which ψ is a nonlinear function that decreases monotonically with increasing of ξ. This type of function is called RBF, and in the literature there are several types such as gaussian, multiquadric, logarithmic to name a few [20, 64, 66]. On the other hand, this article focuses on three specific types of RBF: gaussian, wendland-c2 and Almeida-Evangelista Jr. The reason of choosing these functions is due to the representativeness of the RBF models with global and local support. In addition, a new RBF is used in uncertainty quantification process. This new RBF is so-called Flat-top RBF from Almeida Evangelista Jr. being described later.

The gaussian function is a global classical RBF whose support can span the entire domain [22, 65]. This function presents several variations in its general formulation, being considered for this work the one presented as:

$$\psi = e^{-\left(\frac{\|\xi - \xi^c\|}{\rho}\right)^2} \tag{5.6}$$

The formulation of the RBF Wendland-C2 presented by [23] is classified as a compact support RBF due to its null values outside the support region ($\|\xi - \xi^c\| > \rho$). The respective equation of the RBF Wendland-C2 is given as:

$$\psi = \begin{cases} \left(1 - \frac{\|\xi - \xi^c\|}{\rho}\right)^4 \left(1 + 4\frac{\|\xi - \xi^c\|}{\rho}\right), & 0 \le \frac{\|\xi - \xi^c\|}{\rho} \le 1 \\ 0 & \frac{\|\xi - \xi^c\|}{\rho} > 1 \end{cases} \tag{5.7}$$

The Flat-top RBF of Almeida-Evangelista Jr. consists of a new local support RBF with behavior similar to Wendland-C2 and little known in the literature. However, the main difference is that this RBF has a greater number of parameters than the other RBFs, but presenting a greater efficiency. The equation is defined as:

$$\psi = \begin{cases} 1, & \text{if } \frac{\|\xi - \xi^c\|}{\rho} \le \beta_0 \\ \left(1 - \left(\frac{\frac{\|\xi - \xi^c\|}{\rho} - \beta_0}{1 - 2\beta_0}\right)^{\kappa}\right)^{\kappa}, & \text{if } \beta_0 < \frac{\|\xi - \xi^c\|}{\rho} < (1 - \beta_0) \\ 0, & \text{if } (1 - \beta_0) \le \frac{\|\xi - \xi^c\|}{\rho} \le 1 \text{ or } \frac{\|\xi - \xi^c\|}{\rho} > 1 \end{cases} \tag{5.8}$$

in which, κ is a shape parameter indicating the smoothness of the curve that interconnects the flat regions of the approximation function and β_0 is the Flat-top size parameter with an interval of $0 \le \beta_0 \le 0.50$. The parameters ρ, κ and β_0 are determined using a genetic algorithm and will be presented later. The range of values of the RBFs parameters used for scanning the genetic algorithm were defined as being $0.0001 \le \rho \le 50$, $0 \le \kappa \le 6$ and $0 \le \beta_0 \le 0.499$.

The coefficients $\boldsymbol{\omega}$ are determined with the regularization process that consists of adding a penalty term to the criterion for minimizing the sum of squared errors [18]. Essentially, the added term, called the regularization parameter λ, consists of a positive number added to the design matrix $\boldsymbol{\Psi}$

that can be ill-conditioned [20, 19]. Thus, in order to determine the coefficients $\boldsymbol{\omega}$, it is necessary minimize the cost function $E = \boldsymbol{e}^T\boldsymbol{e} + \lambda\boldsymbol{\omega}^T\boldsymbol{\omega}$, in which is the sum of squared errors over the training set $\mathfrak{I}$. This leads to the solution provided by a least squares procedure given by:

$$\boldsymbol{\omega} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi} + \lambda\mathbf{I}_p\right)^{-1}\boldsymbol{\Psi}^T\mathbf{y} \tag{5.9}$$

in which $\mathbf{I}_p$ is an identity matrix of order $p \times p$. The parameter λ is also determined using a genetic algorithm. For some of RBF models, zero-order regularization was also considered, described here as surrogate models “without regularization” or “non-regularized regression” (λ = 0).

5.2.2. Genetic Algorithm

The genetic algorithm is a probabilistic algorithm that is a part of a broad group of optimization algorithms [117, 118]. The optimization process is based on Darwin’s theory of evolution and genetic recombination with several applications in optimization of engineering design [159, 160, 161]. This algorithm produces new individuals that best adapt to environmental conditions using selection, crossover and mutation operators. The evolution of populations causes an optimization of the spaces that will be scanned by the algorithm. The procedure ends when the best individual adapts to the environmental conditions [161]. This individual is then the ideal solution for the problem to be optimized. The schematic of the genetic algorithm is shown in Figure 5.2.

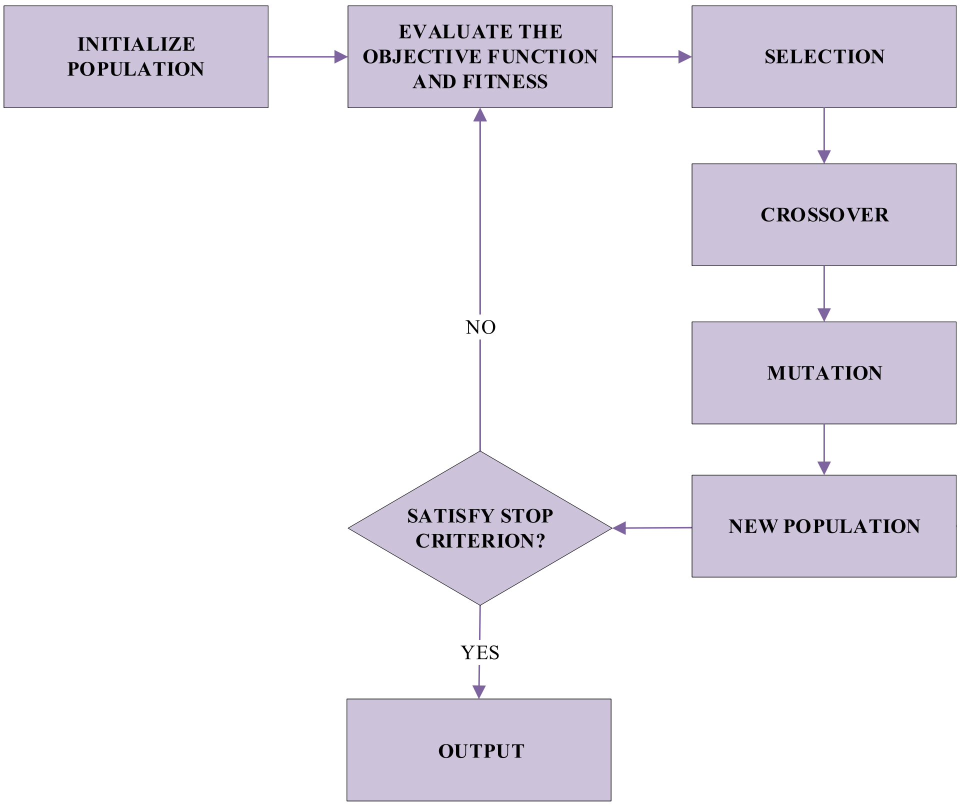


Figure 5.2 – Flowchart of the genetic algorithm to determine the shape parameters of the RBFs

The procedure used by the genetic algorithm generally requires three important points, which are: encoding the variables in chromosomes (usually binary strings), an objective function and operators that define the sampling method [118]. In Figure 5.2, the algorithm starts with an initial population in which chromosomes are used to define the solution space and the ideal solution is obtained by modifying the chromosomes of this population. Then, it is used an objective function that will be optimized to its maximum or minimum value, depending on the objective of the algorithm. Thus, each chromosome is then a solution of this function [118]. Finally, it is important to define the operators that define the sampling method, that is, how the result space will be scanned [118]. The three operators are: selection, crossover and mutation.

In Figure 5.2 the selection corresponds to defining certain chromosomes of the population for the reproduction step and is obtained with the crossover operator. Another important point is the reproduction, and during reproduction, random chromosomes are chosen to create a new chromosome. Finally, the mutation is then used to make small random changes in a chromosome, that is, there is an addition of a variety in the population [118]. Other details about the formulations and functionalities of genetic algorithm are easily found in the literature [150, 119, 162, 120].

The present work considers the use of genetic algorithm functions from MATLAB™ to minimize two objective functions. The first objective function consists of the prediction residual error sum of squares (PRESS); and the second objective function is the generalized cross-validation (GCV). The minimization of the PRESS and GCV functions aims to determine the optimal shape parameters of the RBFs of the surrogate models presented in the purple box of Figure 5.1. It is also important to note that in each surrogate model with RBF, two different situations were considered for the PRESS, being built models with RBF parameters obtained from the minimization of the PRESS without regularization ($\lambda = 0$); and a second situation with parameters of the RBFs obtained from the minimization of the PRESS in models with regularized regression ($\lambda \neq 0$). On the other hand, for the minimization of the GCV, only surrogate models with regularized regression ($\lambda \neq 0$) were addressed. Thus, in this work, the use of GCV is only a complementary analysis, not being the focus of a robust analysis of the GCV.

5.2.3. Indicators of PRESS and GCV

The surrogate models with RBF gaussian (GA), Wendland-C2 (WE-C2) and Almeida-Evangelista Jr. (AEJ-FT) were then trained with the provided training set. During training and cross-validation, the parameters of each RBF, without regularization ($\lambda = 0$) and with regularization ($\lambda \neq 0$), are then determined by optimizing the PRESS of the leave-one-out cross-validation (LOOCV). This procedure was used to measure the prediction of surrogate models and to ensure that the models were not overfitted during the training process. The LOOCV is a type of *S*-folds with the number of folds equal to the number of available dada (S = p). The method takes a sample from the initial sample set (sample disregarded for training) to be predicted by the trained model with the other p - 1 training data. This procedure is then employed for the p data and the performance quality, known as PRESS, is then calculated for each model as:

$$PRESS = \sum_{i=1}^{p} (\hat{y}_{i,-i} - y_i)^2 \tag{5.10}$$

in which, $\hat{y}_{i,-i}$ is the predicted response of the sample taken from the training sample set and y_i is the respective known output of this sample. Although LOOCV is the most time-consuming S-fold cross-validation, it can be the most rigorous and recommended for small training samples. Another competitive method to PRESS adopted in regularized surrogate models is the GCV defined as:

$$GCV = \frac{1}{p} \frac{\left\| \mathbf{P}_m \mathbf{y} \right\|^2}{\left(\frac{1}{p} trace\left(\mathbf{P}_m \right) \right)^2} \tag{5.11}$$

with:

$$\mathbf{P}_m = \mathbf{I}_p - \boldsymbol{\Psi} \left(\mathbf{A}_m \right)^{-1} \boldsymbol{\Psi}^T \tag{5.12}$$

and

$$\mathbf{A}_m = \boldsymbol{\Psi}^T \boldsymbol{\Psi} + \lambda \mathbf{I}_m \tag{5.13}$$

in which $\mathbf{P}_m$ is the projection matrix, $(\mathbf{A}_m)^{-1}$ is the variance matrix, $\mathbf{I}_m$ is the identity matrix of order $m \times m$ and λ is the regularization parameter also described in Equation (5.9). The procedure resembles the traditional PRESS; however, it presents a better result for regularized surrogate models [131, 18, 19]. The GCV method is an adaptive method of the traditional PRESS with a lower computational cost. The use of GCV in models with a regularization parameter equal to zero causes, in some cases, much higher GCV values than PRESS, making it uncompetitive in cases of non-regularized models. In this way, for this work, the minimization of the GCV will be considered only for surrogate models with regularization in order to compare it to the regularized PRESS. Details of the algorithm and formulations of PRESS and GCV can be found in the publications of Orr [18, 19]. Furthermore, the process of minimizing the PRESS and the GCV is performed using the genetic algorithm described above. The LOOCV results for the surrogate models with PRESS and the GCV results for the training and validation sample set are presented in appendices E to G.

5.3. METHODOLOGY

The methodology is shown in this section and aims to train surrogate models to predict the output response under an uncertainty quantification framework considering the statistical uncertainty of the input variables. Surrogate models using RBF functions of type GA, WE-C2 and AEJ-FT were trained to predict the maximum deflection of a cantilever beam in a first example, the concrete compressive strength in a second numerical example and the concrete slump test in a last example.

5.3.1. Training sampling

A sampling procedure was used to obtain a set of training samples $\mathfrak{I} = \{(\boldsymbol{\xi}_i, y_i)|i = 1, \dots p\}$ to be used as shown in the purple boxes of Figure 5.1. For the first numerical example was considered Latin Hypercube Sampling (LHS) to generate p training samples in which each interval was equally divided into n_d intervals creating $(n_d)^k$ combinations of the k input variables. Then, a random algorithm selects p combinations among the $(n_d)^k$ without substitution to the training vector $\boldsymbol{\xi}_i$. The LHS method was selected due to its simplicity and efficiency (more details can be found in [73, 74]). However, for the second and third examples, LHS were not considered, but a portion of the experimental data was randomly selected from the total set of experimental results in the literature (see items 5.4.2 and 5.4.3). After the training phase, each set of training samples was used to evaluate the phenomena of the numerical examples and thus obtain the y_i values to complete the training set $\mathfrak{I}$.

The determination of the numbers of training samples is crucial for training efficiency and quality of the results. A larger sample size generally provides better training and generalization of predictions, but it demands high computational cost, since it requires more p-evaluations of the physical or mathematical model. However, a smaller training sample size p generally decreases the quality of predictions due to the insufficient generalization of the trained surrogate model. This behavior is particularly noticeable for a large number of random variables, as is the case of compressive strength, which has a large number of input variables.

In this paper, the number of training samples was obtained through a series of experiments with different training sets of size p to arrive at a minimum size of 15 training samples for the first example, 55 training samples for the second example and 20 training samples for the last example. In this way, in all numerical examples, the number of samples considered for training the RBF models presented a good performance and this behavior is shown in the next items of this paper. These minimum size of training samples were defined having as a starting point the sample size used by the examples in the literature, and later a convergence test was carried out in search of a decrease in the sample size. This paper will not focus on these convergence tests. Furthermore, in the first numerical example, the use of two different cases for the surrogate models was considered during the training phase; the first case corresponds to the use of training samples without adding noise to the training set response, here called case 01; and the second case, called case 02, corresponds to an addition of a noise that follows a normal probability distribution with zero mean and standard deviation equal to 0.10. This noise is added to the output response of the training set and so later, the construction of surrogate model is done. The additional noise serves to simulate the real behavior of the phenomenon, since the results of experimental tests usually present a disturbance in the data which does not appear in computationally simulated models. However, for case 02, in which noise is added, this noise is only added to the response of the training samples, therefore not being added to the validation samples used in the uncertainty quantification process.

5.3.2. Performance evaluation of uncertainty quantification prediction

Once the surrogate models were trained and validated, the uncertainty quantification was performed for the numerical examples. In the first numerical example that corresponds to the maximum deflection of a cantilever beam, MC simulations and also LHS sampling method were considered. On the other hand, for the second and third numerical example, which correspond to the concrete compressive strength and the concrete slump test, a division of the total sample set into samples for training and the remainder for validation was used, as will be seen posteriorly. The MC simulation method consisted of a choice due to its easy implementation, popularity and convenience that meet the objective of this work, in which is not intended to determine what random simulation technique is more efficient. However, any other random generator could have been considered to determine the uncertainty quantification samples from the first numerical example. A set of 5 E+05 samples was generated and considered as a validation of the surrogate models of the first numerical example and thus, predict and provide a sufficient answer to determine the cumulative distribution functions, as illustrated by the orange boxes in Figure 5.1. In the second numerical example, the difference in relation to the first example is only that the data selected for validation are the 975 experimental data from the literature that are then used in the uncertainty quantification (see item 5.4.2). The third example considered 83 experimental samples from the literature for the validation stage (see item 5.4.3).

The performance of each surrogate model was evaluated by comparing its predicted values with a reference solution in the validation phase, being for Example 01 the same samples of 5 E+05, but analytically evaluated in Equation (5.19). These results are referred here as the Monte Carlo Reference Solution (MC^{REF}). However, for Example 02, the same 975 validation data obtained experimentally in the literature constitute the reference model. The Example 03 considers the 83 validation samples from the literature as the reference model. The performance is then compared using the normalized mean square error (ε) to analyze the goodness of fit of the values of the reference models with the predicted values of the surrogate models. The error ε is given by:

$$\varepsilon = \frac{\left\| \mathbf{y}_{Ref} - \hat{\mathbf{y}} \right\|^2}{\left\| \mathbf{y}_{Ref} - \mu_{\mathbf{y}_{Ref}} \right\|^2} \tag{5.14}$$

in which $\mathbf{y}_{Ref}$ is the vector of reference solutions which for Example 01 corresponds to the values of MC^{REF} and for Example 02 and 03 correspond to the values of the reference model which are the 975 samples of the second example and 83 samples given in the third example; $\hat{\mathbf{y}}$ is the vector with the predicted values of surrogate models, $\mu_{\mathbf{y}_{Ref}}$ is the mean value of the reference models, and $\|\cdot\|$ is the Euclidean norm. Another verification regarding the performance and generalization capacity of the surrogate model that was performed for all numerical examples is the mean square error (MSE) defined as:

$$MSE = \frac{1}{p}\sum_{i=1}^{p}\left(\hat{y}_i - y_i\right)^2 \tag{5.15}$$

The root of the MSE error consists of another measure to quantify the generalization error of surrogate models. This error is called the root mean square error (RMSE) and is given by:

$$RMSE = \left[\frac{1}{p}\sum_{i=1}^{p}\left(\hat{y}_i - y_i\right)^2\right]^{1/2} \tag{5.16}$$

The normalized error (ε), the MSE and the RMSE were used in the validation phase as also considered in the training phase. These three errors aim to quantify the generalization capacity of the surrogate model, that is, to verify if the surrogate model presents a good prediction in the training process and also for an unseen data set by the already trained model (validation). Furthermore, the CDFs predicted by the surrogate models were also statistically tested with the solutions of the reference models using the Kolmogorov-Smirnov (KS) two-sample statistical test. The KS test variable is the maximum vertical distance between the compared CDFs:

$$KS = \max\left[F(\hat{y}) - F_{Ref}(y)\right] \tag{5.17}$$

in which $F(\hat{y})$ is the CDF generated by the predictions of the surrogate models and $F_{Ref}(y)$ is the CDF generated with the values of the reference models. The hypothesis H_0 consists of $F(\hat{y}) = F_{Ref}(y)$, being the hypothesis rejected in the case where the two CDFs are not obtained from the same population, and therefore do not present statistical similarity for a given level of significance.

In this paper, a significance level of 5% was used, which corresponds to the probability of rejection of hypothesis H_0 when it is true (P – value).

5.4. NUMERICAL EXAMPLES

This section presents the numerical examples selected for prediction with the surrogate models based in RBFs of type Gaussian, Wendland-C2 and Almeida and Evangelista Jr. The first example consists of a performance function to predict the maximum deflection of a cantilever beam; the second example corresponds to the concrete compressive strength and the third example consists of the analysis of the results of the concrete slump test. Furthermore, in the first numerical example, two different cases of training samples were considered for each of the surrogate models. The first case, described above, corresponds to case 01 in which the response of the training sample set does not present an additional noise; and a second case, so-called case 02, in which there is an additional noise added to the training set response (see item 5.3.1).

5.4.1. Example 01: Cantilever beam (case 01 and case 02)

Consider a cantilever beam with a rectangular section ($b_w \times h_w$) submitted to a distributed load w [10, 123, 30]. The cantilever beam is shown in Figure 5.3.

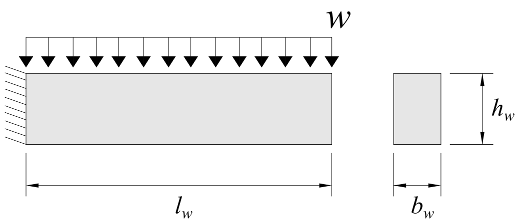


Figure 5.3 – Cantilever beam: a) Longitudinal section, b) Cross section

The serviceability limit state function corresponds to the maximum deflection at the free end of a cantilever beam greater than l_w / 325. The performance function is defined by:

$$G(x) = \frac{l_w}{325} - \frac{w b_w l_w^4}{8EI} \tag{5.18}$$

in which, l_w is the span length, E is the modulus of elasticity and I is the moment of inertia. The variables l_w and E are deterministic with values shown in Table 5.1.

Table 5.1 – Deterministic variables for the Example 01

Var	Unit.	Value
l_w	mm	6.00 E+03
E	MPa	2.60 E+04

The variables w and h_w in Figure 5.3 are rewrite as x_1 and x_2, respectively. The output variable, as previously described, corresponds to the solution of the performance function related to the maximum deflection given in millimeters (mm).

Table 5.2 – Random variables for the Example 01

Var	Unit.	μ	σ	$f(x)$
x_1	N/m²	1000.0	200.0	Normal
x_2	mm	250.0	37.5	Normal

The Equation (5.18) can be reduced by replacing the variables in Tables 5.1 and 5.2 with the final performance function given by:

$$G(x) = 18.46154 - 74769.23\frac{x_1}{x_2^3} \tag{5.19}$$

In the Example 01, two different cases were considered for study, the first one without noise and the second case with the addition of a noise that follows a normal probability distribution of zero mean and standard deviation equal to 0.10. The definition of the outputs with an additional noise (case 02) was given by the equation:

$$G(x)_{Noise} = G(x)(1 + noise) \tag{5.20}$$

For the construction of the surrogate models, 15 training samples were adopted and for the validation stage, a Monte Carlo simulation (MC) was considered with a total of 5E+05 samples (see items 5.3.1 and 5.3.2). The definition of sample sizes for training the RBFs was performed considering the initial sample size as the size adopted by the literature (p = 40). Then, a convergence test was carried out with the objective of get a smaller training sample size (p = 15). Equation (5.19) is a computer simulation model that does not adequately represent the real behavior of the phenomenon. In this way, it is conventional to add noise to the simulated models. The mean and the standard deviation values are shown in Table 5.3 for the reference data set (MC^{REF}) and the 15 training samples of case 01 and case 02. The respective training samples were obtained by LHS sampling. The results for each of these cases and the types of surrogate models are showed in the next sections.

Table 5.3 – Mean and standard deviation of reference data set for Example 01

Case	μ	σ
Case 01 (15 samples)	13.06	3.37
Case 02 (15 samples)	12.76	3.48
MC^{REF}	12.89	3.48

Case 01:

The case 01 corresponds to the use of training samples without considering an additional noise. The parameters obtained in the training of the surrogate models were determined using a genetic algorithm in order to minimize the PRESS and the GCV, and these parameters are shown in Table 5.4. In the determination of the parameters of the surrogate models with the use of a genetic algorithm, it was considered for the case 01 and case 02, exhaustive numerical experiments were done to reduce the computational time and thus obtain a population size between 30 and 50 and a generation size between 40 and 200 to optimize PRESS and GCV. The other parameters of the genetic algorithm are default functions from the MATLAB™. The mutation, for example, considers the "mutationgaussian" function and the "crossoverscattered" function is used for the crossover. Results with larger populations also indicated the same result, but with prohibitive computational cost, especially for PRESS. The results of the PRESS and GCV optimization graphs are shown in appendix D and are not the main focus of this work.

Table 5.4 – Values of the parameters of case 01 for Example 01

Surrogate model	**Parameters**	**PRESS**		**GCV**
		RBF ($\lambda = 0$)	RBF ($\lambda \neq 0$)	RBF ($\lambda \neq 0$)
RBF (GA)	λ	-	0.0000	0.0101
	ρ	2.1250	2.1250	2.0420
	Value	0.5751	0.5751	0.4401
RBF (WE-C2)	λ	-	0.0000	0.0100
	ρ	35.5868	41.3292	5.4430
	Value	0.8550	0.8525	0.4191
RBF (AEJ-FT)	λ	-	0.0000	0.0100
	ρ	1.9142	1.9141	2.3220
	β_0	9.899 E-05	9.899 E-05	0.0842
	κ	2.1208	2.1212	4.6576
	Value	1.8085	1.8085	0.1388

It can be seen in Table 5.4 that the parameters obtained by minimizing the PRESS for the surrogate models were similar and the regularization did not contribute to the minimization of the PRESS. However, it is important to note that the algorithm used allows the scan of the regularized minimum PRESS to also pass through the same combinations of parameters of the non-regularized PRESS, and therefore, for this situation, it is obtained PRESS values similar to those of the non-regularized algorithm. Thus, the presence of regularization did not cause a decrease in the magnitude of PRESS. This behavior only occurs in this example and demonstrates the difficulty in minimizing PRESS for a training set size p = 15. On the other hand, these results should not be despised, as they allowed the construction of surrogate models based on RBFs with good generalization capacity. In the case of the GCV, it is observed that the results obtained present lower values than those of the PRESS. However, this result is only possible for the GCV with the presence of the regularization parameter. Furthermore, in the GCV without the presence of the regularization parameter, the results of the minimum GCV are much higher than the regularized GCV and, therefore, are not relevant (GCV > 100) and, as previously described, are not part of the scope of this paper. It is also noted that the RBF (AEJ-FT) showed the lowest GCV value which is also lower than the PRESS results for the surrogate models.

An additional observation to be mentioned here is the verification of the computational cost. This observation will be made superficially, as this paper is not focused on the study of computational cost. In the determination of the RBFs parameters with the minimization of PRESS and GCV, it was noticed that the GCV presented a smaller computational effort when compared to the PRESS. This behavior was expected, since the GCV has a faster convergence. In the case of RBF (AEJ-FT), this new RBF has a greater number of parameters to be determined by PRESS and GCV. This greater number of parameters did not cause a significant increase in computational time when compared to traditional RBFs. The results of the respective CDFs of the RBFs (GA) of case 01 are shown in Figure 5.4.

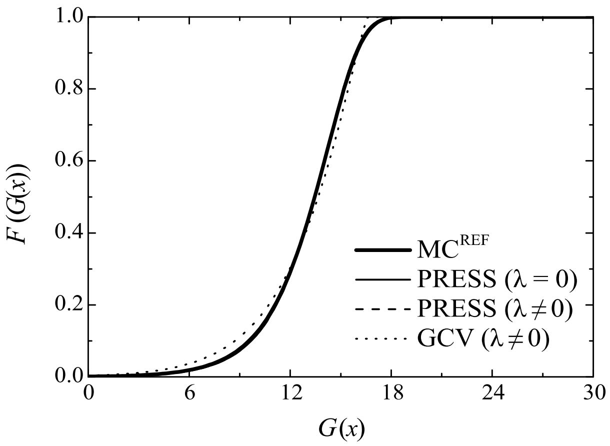


Figure 5.4 – Comparison of cumulative distribution functions (CDF) for RBF (GA) of case 01 of Example 01

In Figure 5.4, it is observed that the CDFs built by all RBF (GA) models show a goody accuracy, and the CDFs built with parameters obtained from the PRESS ($\lambda = 0$ and $\lambda \neq 0$) are very similar and consequently the difference is insignificant in relation to the reference model. This behavior was expected since the parameters of the models with PRESS are quite similar. However, in the CDF constructed to the GCV model (dotted line), it is noted that there is a significant difference between the lower tails of the RBF (GA) trained with the parameters of the regularized GCV in relation to the reference model. This difference between the CDFs is clear in the lower tails, but does not affect the prediction result of the surrogate model. The results obtained for the RBF (WE-C2) of case 01 are shown in Figure 5.5.

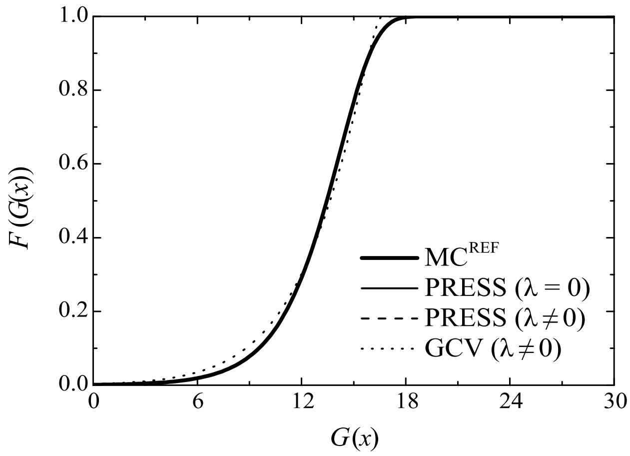


Figure 5.5 – Comparison of cumulative distribution functions (CDF) for RBF (WE-C2) of case 01 of Example 01

The behavior of the CDFs of the surrogate models of RBF (WE-C2) built with parameters derived from the minimization of the PRESS is similar to that of the RBF (GA). For the RBFs (WE-C2) there is a good accuracy of all surrogate models with only a small difference in lower tails for the surrogate model built with the parameters derived from the minimization of the regularized GCV. This difference in lower tails is insignificant and does not affect the prediction result of the surrogate model, as can be seen later in the KS test. For the RBF (AEJ-FT), the results of the CDFs constructed with the parameters obtained from the minimization of the PRESS and the GCV are shown in Figure 5.6.

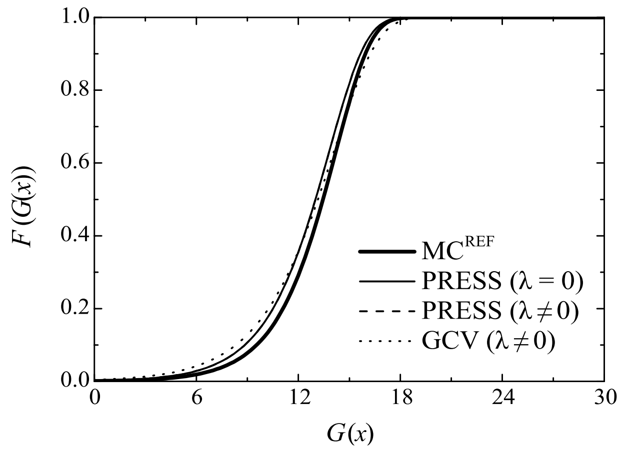


Figure 5.6 – Comparison of cumulative distribution functions (CDF) for RBF (AEJ-FT) of case 01 of Example 01

Figure 5.6 shows that the results has a difference between the lower tails of the predicted model's compared to the reference model. This difference between the tails does not affect the final behavior of the surrogate models, as can be seen in the KS hypothesis test. Furthermore, when comparing the results of the CDFs in relation to the surrogate models with RBF (GA) and RBF (WE-C2), it is noted that the CDF to the GCV model of the RBF (AEJ-FT) presents the worst approximation. However, when looking at Table 5.4, it appears that the surrogate model of RBF (AEJ-FT) with regularized GCV shows the lowest value of GCV for case 01, that is, it is concluded that not always the lowest magnitude of GCV will cause the best approximation of the CDFs. The results of the KS hypothesis test are shown in Table 5.5.

Table 5.5 – Results of the KS Statistical Hypothesis Test for Case 01 of Example 01

Surrogate models	Method	λ	P-Value	KS	Result
RBF (GA)	PRESS	RBF ($\lambda = 0$)	1.00	0.01	H_0
		RBF ($\lambda \neq 0$)	1.00	0.01	H_0
	GCV	RBF ($\lambda \neq 0$)	0.86	0.05	H_0
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	1.00	0.01	H_0
		RBF ($\lambda \neq 0$)	1.00	0.01	H_0
	GCV	RBF ($\lambda \neq 0$)	0.96	0.04	H_0
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	0.52	0.07	H_0
		RBF ($\lambda \neq 0$)	0.53	0.07	H_0
	GCV	RBF ($\lambda \neq 0$)	0.59	0.07	H_0

H_0: predictions of the surrogate models and reference MC solution are obtained from the same population
H_1: predictions of the surrogate models and reference MC solution are obtained from different populations

In Table 5.5, it is noted that the KS hypothesis is not rejected and, therefore, the CDFs of the surrogate models are part of the same population as the CDF of the MC^{REF} model, and this behavior is expected, as observed in the CDFs of the previous figures. The CDFs from RBF (GA) and RBF (WE-C2) models presented a better approximation in relation to the CDF of the reference solution, mainly for models with PRESS. The RBF (AEJ-FT) showed a greater difference from the KS measure, mainly with the models with PRESS. It is also important to emphasize that the threshold value for the KS is 0.12 for a significant level of 5% as an indication of non-rejection of the null hypothesis.

<u>Case 02 (Additional noise):</u>

The case 02 corresponds to the use of training samples with the addition of a perturbation to the training data output that follows a normal probability distribution of zero mean and standard deviation of 0.10. The parameters obtained for training the samples of the surrogate model were obtained by using a genetic algorithm in order to minimize the PRESS and the GCV. These parameters are shown in Table 5.6.

Table 5.6 – Values of the parameters of case 02 for Example 01

Surrogate models	Parameters	PRESS		GCV
		RBF ($\lambda = 0$)	RBF ($\lambda \neq 0$)	RBF ($\lambda \neq 0$)
RBF (GA)	λ	-	0.0123	0.0101
	ρ	0.8693	3.5745	2.5567
	Value	10.0874	5.8792	1.9446
RBF (WE-C2)	λ	-	0.0000	0.0101
	ρ	6.7171	6.7159	8.0476
	Value	6.0742	6.0742	2.0166
RBF (AEJ-FT)	λ	-	0.0000	0.0100
	ρ	5.2089	5.2186	2.9198
	β_0	0.0625	0.0623	0.3618
	κ	1.8525	1.8556	0.2756
	Value	3.7258	3.7230	0.5648

It is noted in Table 5.6 that the addition of the noise to the output of the training set allowed to obtain a lower magnitude of PRESS for the RBF (GA) with regularization when compared to the result of the RBF (GA) without regularization. On the other hand, for the minimum values of PRESS with regularization of RBF (WE-C2) and RBF (AEJ-FT) it was observed that there was no significant change when comparing the magnitudes of the RBF (WE-C2) and RBF (AEJ-FT) without regularization. For the minimization of the GCV with regularization, it was observed magnitudes of GCV lower than half the value of the PRESS with regularization. The computational time required to determine the RBFs parameters by minimizing the GCV was smaller than the time of PRESS minimization, as occurred in case 01. In the case of RBF (AEJ-FT), this new RBF has a greater number of parameters to be determined by PRESS and GCV. This greater number of parameters did not cause a significant increase in computational time when compared to traditional RBFs. The results of the respective CDFs of the RBFs (GA) for the case 02 are shown in Figure 5.7.

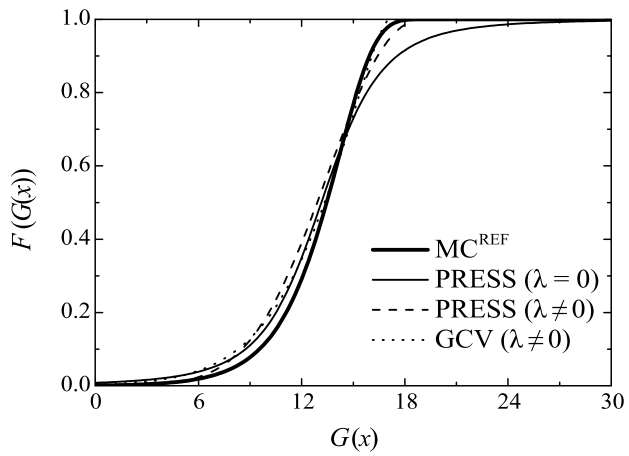


Figure 5.7 – Comparison of cumulative distribution functions (CDF) for RBF (GA) of case 02 of Example 01

Figure 5.7 shows a significant difference between the CDFs obtained from the predicted models with PRESS in relation to the CDF of the reference model (thicker solid line). This result was caused by the addition of noise to the output response of the training set in order to change the parameters used for the prediction of the surrogate models. In Figure 5.7, the parameters derived from the minimization of the GCV allowed a better accuracy of the RBF (GA). On the other hand, the CDF obtained by minimizing the PRESS without the regularization parameter showed a very significant difference in the upper tail of the CDF, and consequently a lower adherence in relation to the reference model. The results of the CDFs for the RBF (WE-C2) models of case 02 are shown in Figure 5.8.

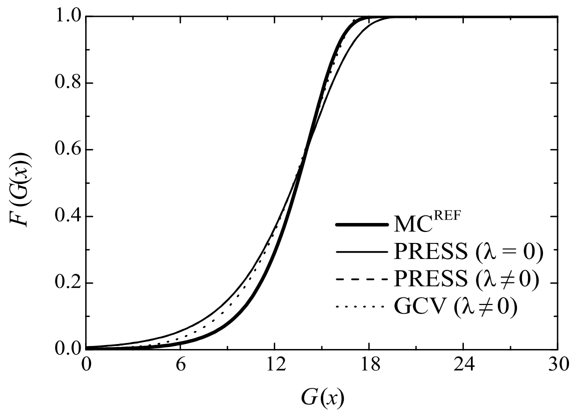


Figure 5.8 – Comparison of cumulative distribution functions (CDF) for RBF (WE-C2) of case 02 of Example 01

The CDFs constructed from the predicted models with the PRESS without regularization and PRESS with regularization are very similar. The results with PRESS showed a lower accuracy than the CDF constructed by the RBF (WE-C2) obtained by minimizing the GCV with regularized regression. Moreover, the CDFs obtained from the PRESS present a lower accuracy than the CDFs from the RBF (GA) of case 01, and this behavior is also explained by the addition of noise to the surrogated models. The CDF results for the surrogate models based in the RBF (AEJ-FT) with a noise at the output response of the training set are shown in Figure 5.9.

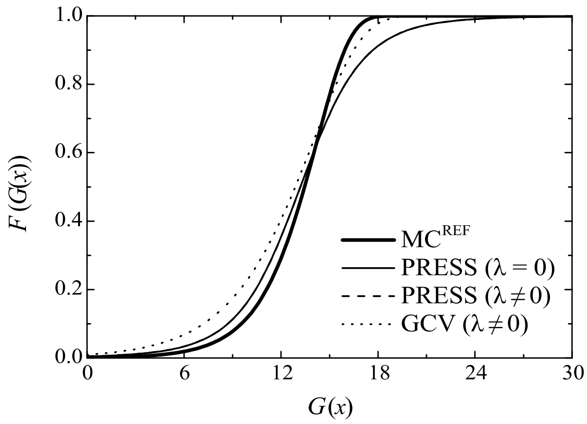


Figure 5.9 – Comparison of cumulative distribution functions (CDF) for RBF (AEJ-FT) of case 02 of Example 01

Figure 5.9 shows that the accuracy of the surrogate models with RBF (AEJ-FT) is lower than the previous models both in case 02 and in case 01, showing a significant difference in the tails of the surrogate models. For the predicted models with the parameters obtained from the PRESS minimization without and with regularization, the greatest difference in the CDFs occurred in the upper tail, that is, for values of $G(x) > 15$; and for the model built with the parameters of the minimization of the regularized GCV, the difference between the tails was in the lower tail. The results of the KS hypothesis statistical test for case 02 are shown in Table 5.7.

Table 5.7 – Results of the KS Statistical Hypothesis Test for Case 02 of Example 01

Surrogate model	Method	λ	P-Value	KS	Result
RBF (GA)	PRESS	RBF ($\lambda = 0$)	0.13	0.10	H_0
		RBF ($\lambda \neq 0$)	0.16	0.10	H_0
	GCV	RBF ($\lambda \neq 0$)	0.69	0.06	H_0
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	0.34	0.08	H_0
		RBF ($\lambda \neq 0$)	0.34	0.08	H_0
	GCV	RBF ($\lambda \neq 0$)	0.62	0.07	H_0
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	0.07	0.11	H_0
		RBF ($\lambda \neq 0$)	0.07	0.11	H_0
	GCV	RBF ($\lambda \neq 0$)	0.05	0.12	H_0

H_0: predictions of the surrogate models and reference MC solution are obtained from the same population
H_1: predictions of the surrogate models and reference MC solution are obtained from different populations

In Table 5.7, it is observed that the surrogate models built with parameters obtained from the minimization of the GCV presented the highest *P* values for the RBFs (GA) and RBFs (WE-C2). However, for the RBF (AEJ-FT), there was no significant difference in the magnitude of *P* between the PRESS and GCV models. It is also important to emphasize that the threshold value for the KS is 0.12 for a significant level of 5% as an indication of non-rejection of the null hypothesis. Thus, the results predicted by the surrogate models are part of the same population as the reference MC. Furthermore, it is noted that the results of the hypothesis test for the RBF(AEJ-FT) in relation to the RBF (GA) and RBF (WE-C2) presented the highest magnitudes of the KS measure.

Prediction of MSE, RMSE and normalized error (ε):

The verification of errors in the results predicted by the RBFs was done in two ways, the first with MSE and RMSE; and the second verification performed by the normalized mean square error (ε) of the results predicted of the RBFs with the reference model. The results of the MSE and RMSE are presented in Table 5.8.

Table 5.8 – MSE and RMSE error for case 01 of Example 01

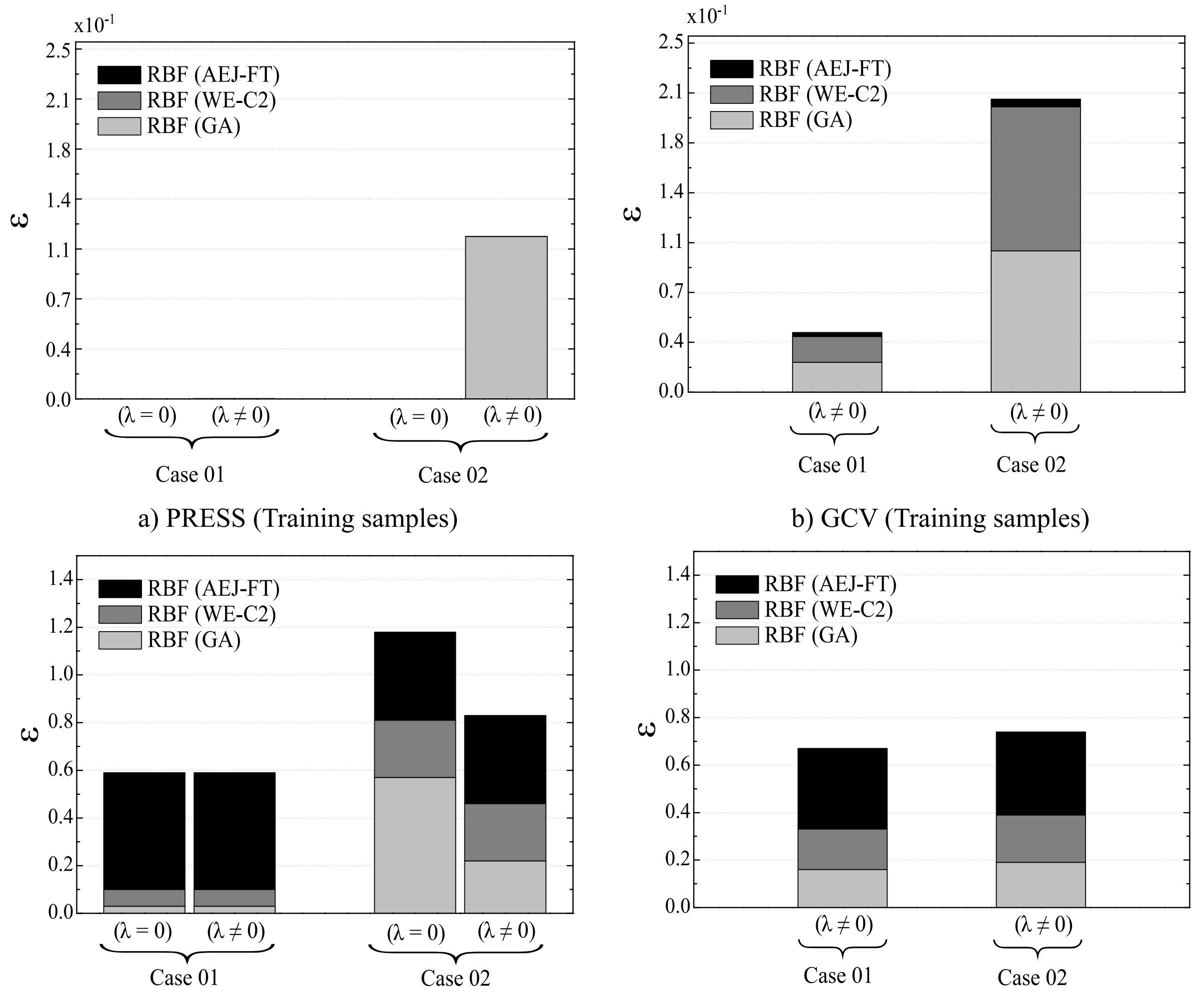
Surrogate model	Method	λ	Training		Validation	
			MSE	RMSE	MSE	RMSE
RBF (GA)	PRESS	RBF ($\lambda = 0$)	2.05 E-03	4.53 E-02	3.74 E-01	6.12 E-01
		RBF ($\lambda \neq 0$)	2.05 E-03	4.53 E-02	3.74 E-01	6.12 E-01
	GCV	RBF ($\lambda \neq 0$)	2.23 E-01	4.72 E-01	1.90 E+00	1.38 E+00
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	4.55 E-08	2.13 E-04	8.09 E-01	8.99 E-01
		RBF ($\lambda \neq 0$)	1.56 E-05	3.95 E-03	8.00 E-01	8.94 E-01
	GCV	RBF ($\lambda \neq 0$)	1.91 E-01	4.37 E-01	2.01 E+00	1.42 E+00
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	3.76 E-19	6.13 E-10	5.96 E+00	2.44 E+00
		RBF ($\lambda \neq 0$)	3.02 E-17	5.50 E-09	5.94 E+00	2.44 E+00
	GCV	RBF ($\lambda \neq 0$)	2.94 E-02	1.71 E-01	4.06 E+00	2.01 E+00

The MSE and RMSE errors indicate a good approximation of the sample set in the training stage with a non-significant increase in the error magnitude during the validation phase. In the validation phase, the values of MSE and RMSE indicate a good capacity to approximate unseen data from the surrogate models already trained. The RBFs with gaussian and wendland-C2 and with non-regularized PRESS showed the lowest RMSE error values and consequently a better efficiency when compared to RBF (AEJ-FT). However, in the training phase, the models based on the RBF (AEJ-FT) present practically zero errors. The presence of regularization in the surrogate models with PRESS did not cause a decrease in errors, since the parameters optimized by the regularized PRESS are practically the same as those of PRESS without regularization. Models with RBF (AEJ-FT) and PRESS showed similar MSE and RMSE error magnitudes. This behavior was expected, as the values obtained from PRESS are also similar. This, however, does not indicate that an overfitting occurred, but rather, it demonstrates that in this Example 01, the RBFs presented greater difficulty in determining the PRESS with very small sample size ($p = 15$). The RBF (AEJ-FT) with GCV was the only one that has a lower indicator than that obtained for the PRESS with regularization in the validation phase. The processing time will be given as a function of the number of bases functions to be built by the surrogate model during the training phase. Each base function was built with one training sample. In this way, the processing time can be determined as the number of evaluations performed during the training phase. As the number of samples is similar for all RBFs, it can be stated that the computational effort is similar for all RBFs in case 01. The results of the MSE and RMSE errors for case 02 are shown in Table 5.9.

Table 5.9 – MSE and RMSE error for case 02 of Example 01

Surrogate model	Method	λ	Training		Validation	
			MSE	RMSE	MSE	RMSE
RBF (GA)	PRESS	RBF ($\lambda = 0$)	1.47 E-17	3.84 E-09	6.91 E+00	2.63 E+00
		RBF ($\lambda \neq 0$)	1.28 E+00	1.13 E+00	2.62 E+00	1.62 E+00
	GCV	RBF ($\lambda \neq 0$)	1.12 E+00	1.06 E+00	2.25 E+00	1.50 E+00
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	2.31 E-15	4.81 E-08	2.93 E+00	1.71 E+00
		RBF ($\lambda \neq 0$)	4.59 E-15	6.78 E-08	2.93 E+00	1.71 E+00
	GCV	RBF ($\lambda \neq 0$)	1.14 E+00	1.07 E+00	2.45 E+00	1.56 E+00
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	8.82 E-15	9.39 E-08	4.45 E+00	2.11 E+00
		RBF ($\lambda \neq 0$)	2.99 E-14	1.73 E-07	4.44 E+00	2.11 E+00
	GCV	RBF ($\lambda \neq 0$)	6.18 E-02	2.49 E-01	4.24 E+00	2.06 E+00

The presence of noise in the output response of case 02 did not cause an obstacle to the generalization of the surrogate models. Note that the errors MSE and RMSE are practically null in the training stage and for the validation stage they have significantly small errors. The presence of the regularization parameter allowed for PRESS smaller errors only for the RBF (GA) in the validation phase. Models with RBF (AEJ-FT) and PRESS showed similar MSE and RMSE error magnitudes. This behavior was expected, as the values obtained from PRESS are also similar. This, however, does not indicate that an overfitting occurred. The results obtained for the model with GCV indicated a better accuracy when compared to the other surrogate models. The computational effort is similar for all RBFs because both have the same 15 samples to be evaluated during the training phase. The normalized mean square error analysis of the predicted results compared to the reference MC model is shown in Figure 5.10.



a) PRESS (Training samples)

b) GCV (Training samples)

c) PRESS (Validation samples)

d) GCV (Validation samples)

Figure 5.10 – Normalized mean square error (ε) of the surrogate models in Example 01

In the training stage, the normalized error is practically null for surrogate models with parameters optimized by PRESS. The only exception is for the RBF (GA) which has an error of approximately 0.11 and although higher compared to the others is not significant. The surrogate models with parameters optimized by the GCV indicated lower normalized errors for the RBF (AEJ-FT) during the training process. The presence of a noise in the case 02 caused an increase in the value of the prediction error of the models with parameters obtained from the minimization of the PRESS in the validation step when compared to case 01.On the other hand, for the RBF (AEJ-FT), it is noted that there was a small decrease in the error. In the models predicted with GCV, it is observed that the noise in case 02 had a not very significant increase in error when compared to case 01 in the validation phase. The case 02 with regularized GCV presented lower normalized errors compared to case 02 with PRESS and regularization in the validation stage.

5.4.2. Example 02: Concrete compressive strength (f_c)

The concrete compressive strength data correspond to the University of California, Irvine (UCI) machine learning repository obtained in [163]. The UCI data [163] are experimental data from the paper of Yeh [164], and this sample set was used in several publications [165, 166]. The paper of Yeh [164] considered a total of 727 samples distributed in approximately 72% as training sampling and 28% as validation sampling in neural network surrogate model. Posteriorly, Yeh added more experimental results and this sample set was used in other publications [166]. In the publication of Wu and coauthors [166] new approaches with active learning for regression (ALR) were proposed based on a greedy sampling (GS). These approaches aimed to reduce the number of samples from the training set by selecting the most beneficial ones rather than a random selection [166]. To verify the efficiency of the approaches adopted by the authors, extensive experiments were done in several benchmark problems from the literature and the results demonstrated a good efficiency of the proposed methods, mainly related to a better generalization. Among the examples used by the authors, the concrete compressive strength presented a total number of 1030 of samples in which 60 samples were used for training the surrogate models. The mean and standard deviation values of the 1030 samples used by Wu and coauthors [166] as a function of specific weight and age are shown in Table 5.10.

Table 5.10 – Variables of concrete compressive strength

Var	**Unit.**	μ	σ
Cement	kg/m³	281.17	104.51
Blast furnace slag	kg/m³	73.90	86.28
Fly ash	kg/m³	54.19	64.00
Water	kg/m³	181.57	21.36
Superplasticizer	kg/m³	6.20	5.97
Coarse aggregate	kg/m³	972.92	77.75
Fine aggregate	kg/m³	773.58	80.18
Age	day	45.66	63.17
Compressive strength (f_c)	MPa	35.82	16.71

The construction of surrogate models used a total of 8 input variables with an output variable that corresponds to the concrete compressive strength in MPa. These input variables showed in Table 5.10 correspond to an experimental data obtained from the UCI repository. The surrogate models based on RBFs used 55 samples for training in order to be competitive with the results obtained by Wu and coauthors [166] with the rest of the 1030 samples, that is, a total of 975 samples for the validation of the surrogate models.

<u>Uncertainty quantification:</u>

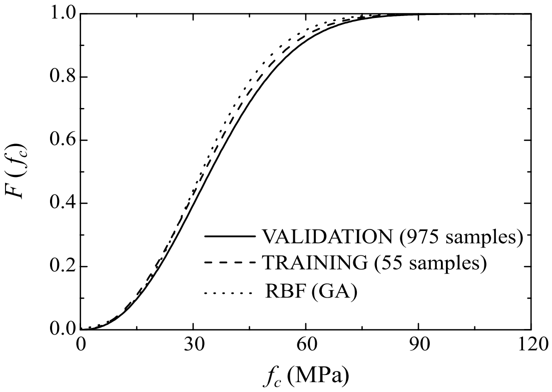
The surrogate models based on RBFs used 55 samples for the training process. The size of the training sets considered for the RBFs was obtained through a convergence test. The convergence started with a sample size similar to that adopted by Wu and coauthors [166] (p = 60). The minimum sample size p = 55 was found at the end of the convergence. The convergence process is already been extensively explored by the literature and consequently will not be studied in Example 02. The training sampling was randomly selected from 1030 samples of concrete compressive strength. These 55 samples have a natural noise and therefore it is not necessary to add a noise to the output response of the training samples.

The parameters obtained for training the samples of the surrogate models were acquired by a genetic algorithm in order to minimize the PRESS and the GCV, and these parameters are shown in Table 5.11. Moreover, to determine the parameters of the surrogate models using the genetic algorithm, exhaustive numerical experiments were considered to reduce computational cost and thus obtain a population size between 15 and 30 and a generation size between 30 and 50 to optimize the PRESS and the GCV. The other parameters of the genetic algorithm are default functions from the MATLAB™. The mutation, for example, considers the "mutationgaussian" function and the "crossoverscattered" function is used for the crossover. Results with larger populations also indicated the same result, but with a prohibitive computational cost, especially for the PRESS. The results of the PRESS and the GCV optimization graphs are shown in appendix D and are not the main focus of this work.

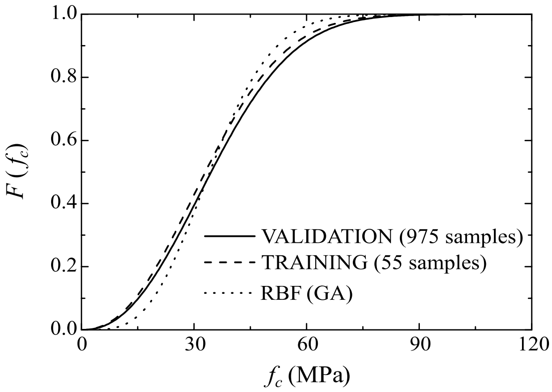
Table 5.11 – Values of the parameters for Example 02

Surrogate model	**Parameters**	**PRESS**		**GCV**
		RBF (λ = 0)	RBF ($\lambda \neq 0$)	RBF ($\lambda \neq 0$)
RBF (GA)	λ	-	0.0224	0.0117
	ρ	0.8120	1.7526	1.8984
	Value	83.0099	64.8688	70.9491
RBF (WE-C2)	λ	-	0.0515	0.0157
	ρ	2.5685	5.4023	4.4522
	Value	78.7411	67.5372	73.6126
RBF (AEJ-FT)	λ	-	0.0380	0.0271
	ρ	4.4379	2.6994	2.1825
	β_0	0.0148	0.1422	0.0873
	κ	2.1557	1.3300	1.8912
	Value	90.6264	65.0070	66.4538

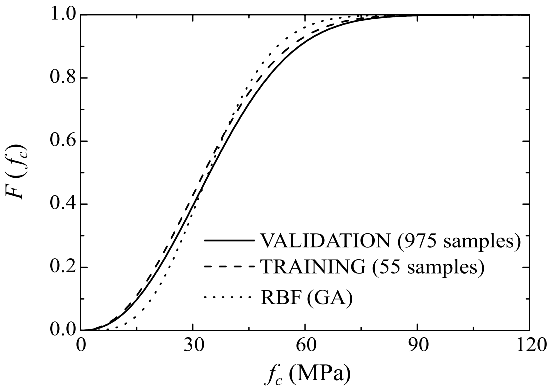
The results of the parameters of the surrogate models show that the presence of the regularization parameter decreases the magnitude of the PRESS by 14 to 28% when compared to the result of the non-regularized PRESS. In the case of the GCV, it is noted that for the RBF (GA) and RBF (WE-C2), the result obtained from the regularized GCV is greater than that of PRESS with ridge regression. On the other hand, the result of the GCV for the RBF (AEJ-FT) has a magnitude very similar to that of the regularized PRESS. The verification of the computational cost for determining the parameters of the RBFs by minimizing the PRESS and the GCV is similar to Example 01. The models with GCV presented a faster convergence when compared to the models with PRESS. In this way, the models with GCV showed a lower computational effort. The results of the CDFs built for the RBFs (GA) are shown in Figure 5.11.



a) PRESS, λ = 0



b) PRESS, λ ≠ 0



c) GCV, λ ≠ 0

Figure 5.11 – Comparison of cumulative distribution functions (CDF) for RBF (GA) of Example 02

In Figure 5.11, there are three CDFs of different formats for each graph, one with a solid line, another with a dashed line and the last with a dotted line. The CDF represented by a solid line corresponds to the response obtained for the 975 experimental data that are considered as a reference model. The dashed line corresponds to the CDF constructed with the experimental results of the 55 samples used for training the RBFs, and the dotted curve corresponds to the CDF of the prediction of the RBFs for the 975 samples of the reference model. It is also noted that among the three surrogate models, the model in which the parameters obtained from the minimization of the non-regularized PRESS showed the best approximation of the reference model (solid line). For the GCV and the regularized PRESS, it was observed that there is no significant difference between the lower and upper tails of the CDFs of the surrogate models in relation to the CDF of the reference model. The results of the CDFs of the surrogate models based on RBFs (WE-C2) are shown in Figure 5.12.

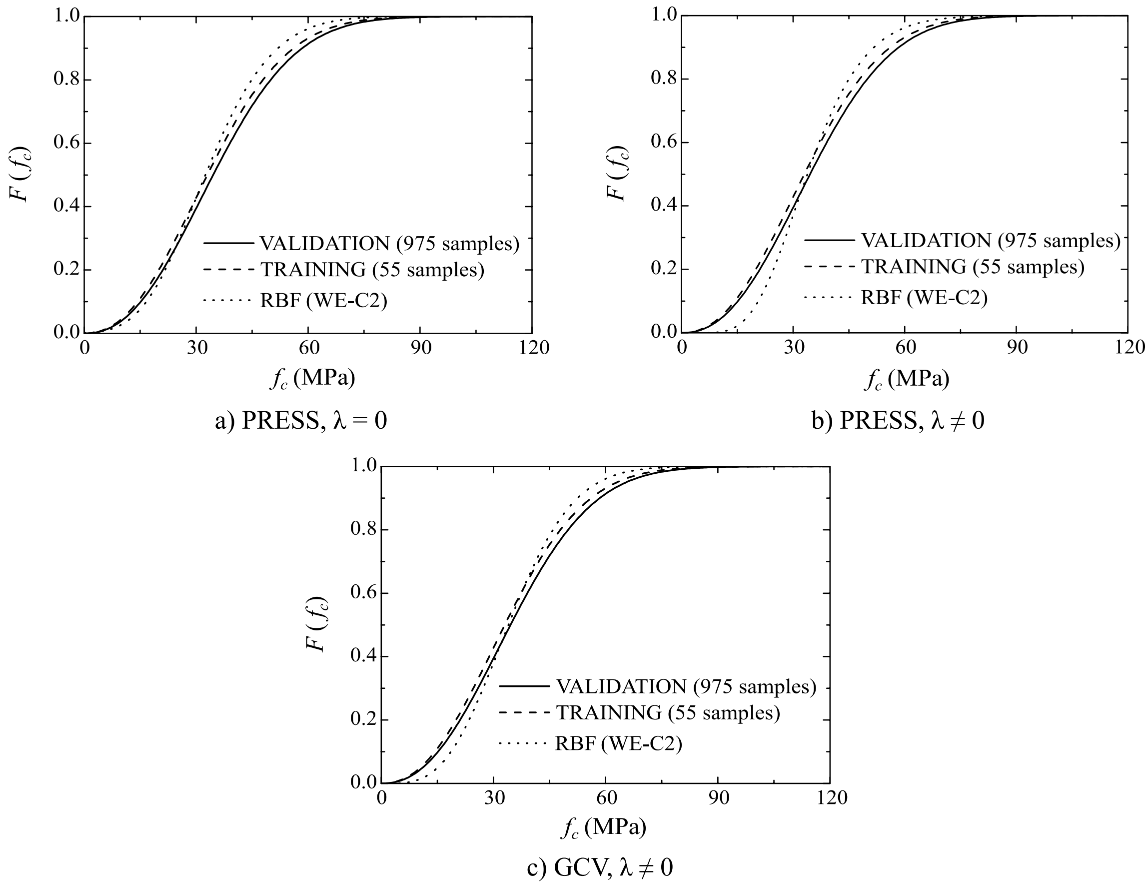
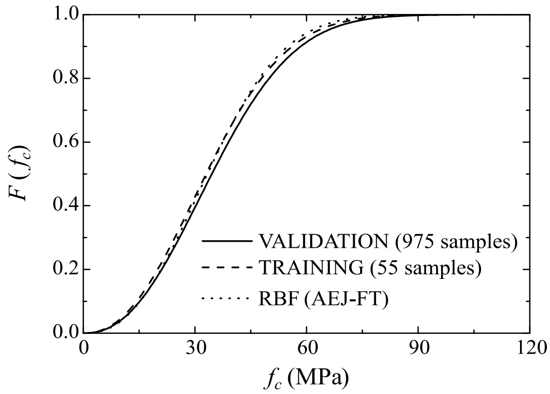
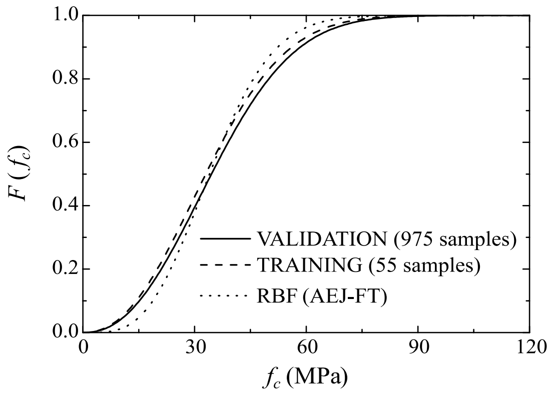


Figure 5.12 – Comparison of cumulative distribution functions (CDF) for RBF (WE-C2) of Example 02

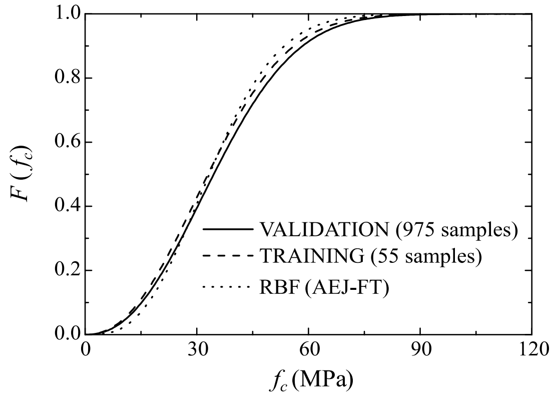
The behavior of the CDF of the surrogate model of RBF (WE-C2) built with parameters derived from the minimization of the PRESS with regularization has a greater difference in relation to the reference model (solid line), being this approximation inferior to that obtained for the RBF (GA) model with regularized PRESS. For the model with GCV, there is a better approximation to the reference model than the PRESS with regularization. On the other hand, the CDF obtained from the GCV presents a lower approximation than that obtained from the non-regularized PRESS. The differences in the lower tails are clear for models with regularization for both PRESS and GCV. The CDF results for the RBF (AEJ-FT) are shown in Figure 5.13.



a) PRESS, λ = 0



b) PRESS, λ ≠ 0



c) GCV, λ ≠ 0

Figure 5.13 – Comparison of cumulative distribution functions (CDF) for RBF (AEJ-FT) of Example 02

The result of Figure 5.13a shows that the RBF (AEJ-FT) with parameters obtained from the minimization of the PRESS without regularization presents the best approximation among all the models with RBF (AE-FT). However, when observing the CDF built for the regularized PRESS, it was noticed a smaller approximation of the CDF in relation to the reference model. The model with parameters obtained with the minimization of the regularized GCV has an accuracy superior to that of the CDF with regularized PRESS. It is also noted that the CDF obtained from the PRESS without regularization is very similar to the CDF of the 55 training samples. The results of the KS hypothesis test for all the surrogate models are shown in Table 5.12.

Table 5.12 – Results of the KS Statistical Hypothesis Test of Example 02

Surrogate model	Method	λ	P-Value	KS	Result
RBF (GA)	PRESS	RBF ($\lambda = 0$)	0.58	0.07	H_0
		RBF ($\lambda \neq 0$)	0.55	0.07	H_0
	GCV	RBF ($\lambda \neq 0$)	0.64	0.07	H_0
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	0.33	0.08	H_0
		RBF ($\lambda \neq 0$)	0.22	0.09	H_0
	GCV	RBF ($\lambda \neq 0$)	0.60	0.07	H_0
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	0.98	0.04	H_0
		RBF ($\lambda \neq 0$)	0.57	0.07	H_0
	GCV	RBF ($\lambda \neq 0$)	0.77	0.06	H_0
Training (55 samples)	-	-	0.99	0.04	H_0

H_0: predictions of the surrogate models and reference model solution are obtained from the same population
H_1: predictions of the surrogate models and reference model solution are obtained from different populations

The KS test showed that the results predicted by the RBFs that were used in the construction of the CDFs are from the same population as the 975 experimental data. For the surrogate models, it was also noted that the best approximation is obtained for the RBF (AEJ-FT) with non-regularized PRESS. In the RBFs (GA) and RBFs (WE-C2), the approximations of the models with regularized GCV showed higher *P* values than those obtained by the PRESS. It is also important to emphasize that the threshold value for the KS is 0.12 for a significant level of 5% as an indication of non-rejection of the null hypothesis.

Prediction of MSE, RMSE and normalized error (ε)

The verification of the errors of the results predicted by the RBFs is performed in two ways, the first being through the comparison of the MSE and RMSE with the results of Wu and coauthors [166]; and the second verification performed by the normalized mean square error (ε) of the results of the prediction of the RBFs with the reference model. The MSE and RMSE results are presented in Table 5.13. Wu and coauthors [166] use approaches with ALR and GS, the models proposed by the authors are the iGs and the GSy and the GSx model is from the literature, but the results was obtained by the authors themselves. The models GSy and iGs are based on active learning for regression in order to select the most propitious samples rather than a random selection. The first, GSy, so-called greedy sampling on the output, selects new samples to enhance the variety at the output and the second technique called improved greedy sampling (iGS) on both inputs and outputs, selects new samples for the input and output spaces [166]. The GSx method, so-called greedly sampling on the inputs, was developed by Yu and Kim [167] and also adopted by [166]. The results of GSx was obtained by Wu and coauthors [166]. This method consists of selecting the first sample as being the closest to the center of the sample set and in each new iteration a new sample positioned as far as possible from the previously selected ones is then added in order to guarantee the greatest diversity of the selected samples [166]. These methods used by Wu and other authors can be better understood in the literature [166]. For [166] the results showed a better performance for 60 training samples of the iGs, GSx and GSy models. On the other hand, the results obtained for the RBFs required only 55 samples to build the surrogate models.

Table 5.13 – MSE and RMSE error of Example 02

Surrogate model	Method	λ	Training		Validation	
			MSE	RMSE	MSE	RMSE
RBF (GA)	PRESS	RBF ($\lambda = 0$)	7.14 E-13	8.45 E-07	183.68	13.55
		RBF ($\lambda \neq 0$)	4.36 E+01	6.61 E+00	95.85	9.79
	GCV	RBF ($\lambda \neq 0$)	4.20 E+01	6.48 E+00	95.78	9.79
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	3.66 E-15	6.05 E-08	139.46	11.81
		RBF ($\lambda \neq 0$)	5.36 E+01	7.32 E+00	100.57	10.03
	GCV	RBF ($\lambda \neq 0$)	3.59 E+01	5.99 E+00	93.99	9.69
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	2.32 E-11	4.82 E-06	124.82	11.17
		RBF ($\lambda \neq 0$)	4.72 E+01	6.87 E+00	99.82	9.99
	GCV	RBF ($\lambda \neq 0$)	3.69 E+01	6.08 E+00	110.95	10.53
GSy (a) ($p = 10$)	-	-	-	-	235.62 (b)	15.35
GSy (a) ($p = 30$)	-	-	-	-	137.36 (b)	11.72
GSy (a) ($p = 60$)	-	-	-	-	122.99 (b)	11.09
iGS (a) ($p = 10$)	-	-	-	-	203.35 (b)	14.26
iGS (a) ($p = 30$)	-	-	-	-	133.63 (b)	11.56
iGS (a) ($p = 60$)	-	-	-	-	119.25 (b)	10.92
GSx (a) ($p = 10$)	-	-	-	-	359.10 (b)	18.95
GSx (a) ($p = 30$)	-	-	-	-	131.79 (b)	11.48
GSx (a) ($p = 60$)	-	-	-	-	124.99 (b)	11.18

(a) Results extracted from Wu and coauthors [166]. GSx: greedy sampling on the inputs; GSy: greedy sampling on the outputs; iGS: improved greedy sampling on both inputs and output.
(b) Results obtained by the Equation (5.15).

The MSE and RMSE errors of the RBF-based models are practically null for the surrogate models with parameters optimized by the non-regularized PRESS during the training phase. On the other hand, still in the training phase, the values of MSE and RMSE tend to increase due to the presence of regularization parameter in the PRESS. The RMSE values indicate that the GCV presented values close to the PRESS with regularization. In the validation phase, the presence of regularization was fundamental to reduce the errors in the prediction of the concrete compressive strength, with the smallest errors for the models with regularized PRESS. However, for the RBFs with PRESS without regularization, the presence of a greater residual error of RMSE was noticed. The ε error results are shown in Figure 5.14. An additional observation to be mentioned here is the verification of the computational cost. The referenced methods in the literature do not make a study of the CPU time for Example 02. A viable solution to verify the processing time will be given as a function of the number of samples to build the bases functions during the training phase. The surrogate models with RBF (GA, WE-C2 and AEJ-FT) shows a reduction in the number of samples in approximately 8% when compared to the methods GSy, iGS and GSx. This occurs because the methods with RBFs need to evaluate a smaller training set during the training phase (p = 55), and the methods in the literature take longer to evaluate p = 60. On the other hand, the sets of p = 10 and p = 30 used by the literature in ridge regression models should present lower computational times than RBFs. This occurs because the number of samples is much smaller than the number of RBFs samples.

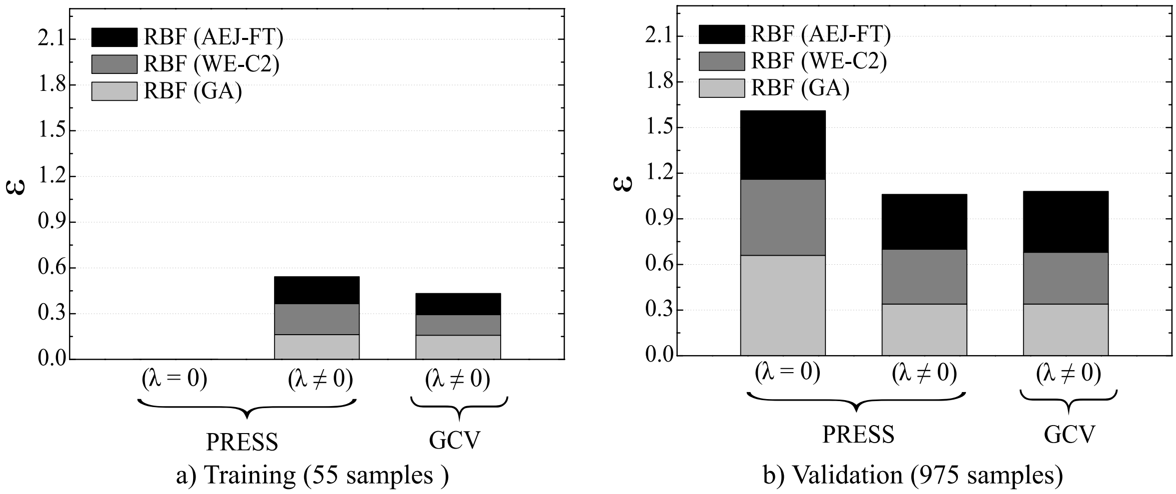


Figure 5.14 – Normalized mean square error (ε) of the surrogate models in Example 02

In Figure 5.14a, it is shown that the normalized error is practically null for all RBFs with parameters optimized by the non-regularized PRESS in the training process. The normalized errors of the PRESS with regularization indicate a lower accuracy when compared to the error results for the GCV in Figure 5.14a. In this case, the GCV showed a better behavior in the training phase. In the validation step, it is noted that the normalized errors present magnitudes higher than those obtained in the training in which the normalized error is given for the RBF (WE-C2). Furthermore, in Figure 5.14b, it is observed that the RBF (AEJ-FT) presents errors similar to the RBF (WE-C2).

5.4.3. Example 03: Concrete slump test ($\boldsymbol{s_p}$)

The data considered for the concrete slump test are from the UCI repository [168] and correspond to an initial publication of Yeh [169]. In this first paper, the author verified the behavior of the properties of the fresh concrete through a study of the high-performance concrete. The surrogate models used were the second order polynomial regression and artificial neural networks with a total of 78 samples of different mixtures and proportions of materials. The construction of the surrogate models in the training phase was done with the use of 74% of the samples for training stage and 26% of the samples for the validation process. The results obtained showed that models with artificial neural networks had a better capacity to predict and generalize than surrogate models with second-order polynomials in predicting the slump flow [169]. On the other hand, it is important to emphasize that the size for the training set used by the author is considered relatively large. Other publications also considered this slump test dataset in which more samples were added to be trained and validated by other surrogate models [170, 171, 166]. The paper of Wu and coauthors

[166] has already been presented in Example 02 and among one of the examples used by the authors, there are the results of 103 slump test samples that are used in the surrogate models. In this paper, the authors consider several training sets ranging from 8 to 20 training samples. The best results obtained by Wu and coauthors [166] showed a good efficiency of the methods (see Example 02 for more details on the work of Wu and coauthors). The values of Wu and coauthors are shown in Table 5.14.

Table 5.14 – Variables of concrete slump test

Var	Unit.	μ	σ
Cement	kg/m³	229.89	78.88
Blast furnace slag	kg/m³	77.97	60.46
Fly ash	kg/m³	149.01	85.42
Water	kg/m³	197.17	20.21
Superplasticizer	kg/m³	8.54	2.81
Coarse aggregate	kg/m³	883.98	88.39
Fine aggregate	kg/m³	739.61	63.34
Slump (s_p)	cm	18.05	8.75

The first seven variables are properties of fresh concrete and were used by Wu and coauthors [166] as input variables for training the surrogate models. The output response of the 103 samples consists of the slump given in centimeters. In the present work, the same training sample size used by Wu and coauthors [166] is considered which is 20 samples. The size of the training sets considered for the RBFs was obtained through a convergence test. The convergence started with a sample size similar to that adopted by Wu and coauthors [166] (p = 20). However, in the convergence test it was not possible to find a sample size smaller than p = 20. The remainder of the samples, which is 83 samples, is used as validation to verify the generalization capacity of the surrogate models based on RBFs. In addition, a comparison of the results from the RMSE with those obtained in the paper of Wu and coauthors [166] is considered.

Uncertainty quantification

The surrogate models based on RBF were built using 20 training samples randomly selected from the 103 samples of the concrete slump test. These 20 samples have a natural noise and, therefore, it is not necessary to add noise to the output response of the training samples. The parameters obtained for training the samples of the surrogate models were acquired with a genetic algorithm in order to minimize the PRESS and the GCV, and these parameters are shown in Table 5.15. The determination of the parameters of the surrogate models with the use of a genetic algorithm considered exhaustive numerical experiments to reduce computational cost. The population size in the genetic algorithm was 30 and the maximum generation number was 40 for optimizing the PRESS and the GCV. Other larger population sizes were used, but they indicated prohibitive computational cost. The other parameters of the genetic algorithm are default functions from the MATLAB™. The mutation, for example, considers the “mutationgaussian” function and the “crossoverscattered” function is used for the crossover. The results of the PRESS and GCV optimization plots are shown in appendix D of this paper. It is important to emphasize that these optimization graphs are not the main focus of this work.

Table 5.15 – Values of the parameters for Example 03

Surrogate model	Parameters	PRESS		GCV
		RBF ($\lambda = 0$)	RBF ($\lambda \neq 0$)	RBF ($\lambda \neq 0$)
RBF (GA)	λ	-	0.3547	0.9989
	ρ	0.5999	1.0740	1.0517
	Value	48.6262	29.2473	47.5401
RBF (WE-C2)	λ	-	0.3186	0.9989
	ρ	2.0308	2.7764	2.6514
	Value	40.6427	29.2248	47.1380
RBF (AEJ-FT)	λ	-	0.9456	0.0130
	ρ	22.7980	2.4548	2.0589
	β_0	0.0026	0.4779	0.3143
	κ	0.4645	5.9998	0.6936
	Value	25.8691	24.5706	17.3928

The regularization process caused a decrease in the PRESS value for RBF (GA) and RBF (WE-C2) surrogate models in which the magnitude of the PRESS had a reduction of 40% and 28%, respectively, when these are compared to the PRESS values of the models without regularization. On the other hand, the same does not occur with the RBF (AEJ-FT) which did not have a significant difference between the PRESS values. In the case of the parameters of the RBFs optimized by the minimization of the GCV, it can be observed that the value of the GCV is superior to the value obtained for the regularized PRESS. This behavior indicates that the generalization capacity of the GCV tends to be lower for the RBF (GA) and RBF (WE-C2). The RBF (AEJ-FT) showed a different behavior in the minimization of the GCV in which the result obtained from the GCV is lower than the value of the regularized PRESS. Thus, initially because the GCV is smaller for the RBF (AEJ-FT), the respective surrogate model tends to present a greater capacity to predict the unseen dataset of the already trained model. The verification of the computational cost for determining the parameters of the RBFs by minimizing the PRESS and the GCV is similar to Example 02. The models with GCV presented a faster convergence when compared to the models with PRESS. In this way, the models with GCV showed a lower computational effort. The results of the CDFs built for the RBFs (GA) are shown in Figure 5.15.

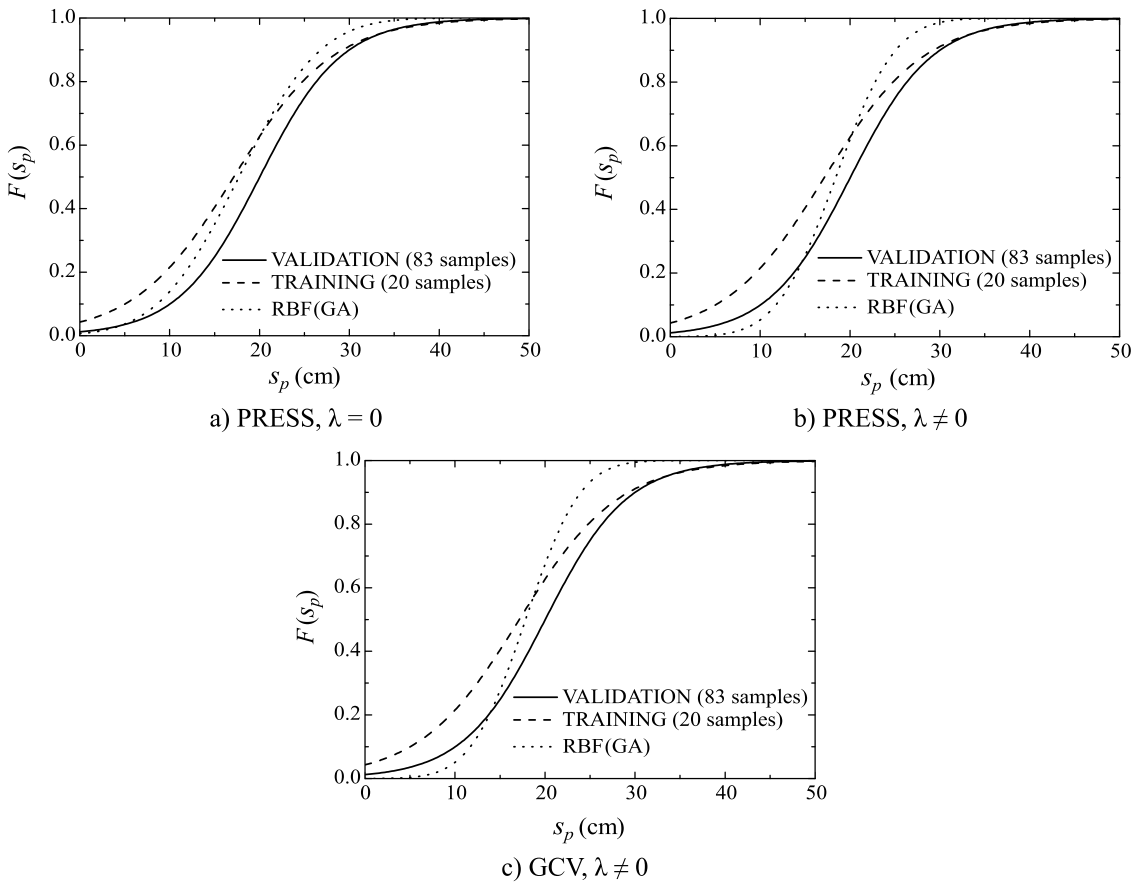


Figure 5.15 – Comparison of cumulative distribution functions (CDF) for RBF (GA) of Example 03

In Figure 5.15, there are three CDFs of different formats for each graph, one with a solid line, another with a dashed line and the last with a dotted line. The CDF represented by a solid line corresponds to the response obtained for the 83 samples that are considered in the validation phase. The dashed line corresponds to the CDF constructed with the experimental results of the 20 samples used for training the RBFs, and the dotted curve corresponds to the CDF of the prediction of the RBFs for the 83 validation samples. The results show that the surrogate models based on the RBF (GA) present CDFs that are distant from the validation CDF (solid line). The CDFs built for the RBF (GA) with regularization showed a greater distance for the upper tails. In the model with RBF (GA) and non-regularized PRESS, a greater distance is observed in the intermediate region of the CDF as well as in the upper tail. Furthermore, it is noted that models with RBF (GA) are not efficient to predict the validation CDF. The results of the CDFs of the surrogate models with RBF (WE-C2) are shown in Figure 5.16.

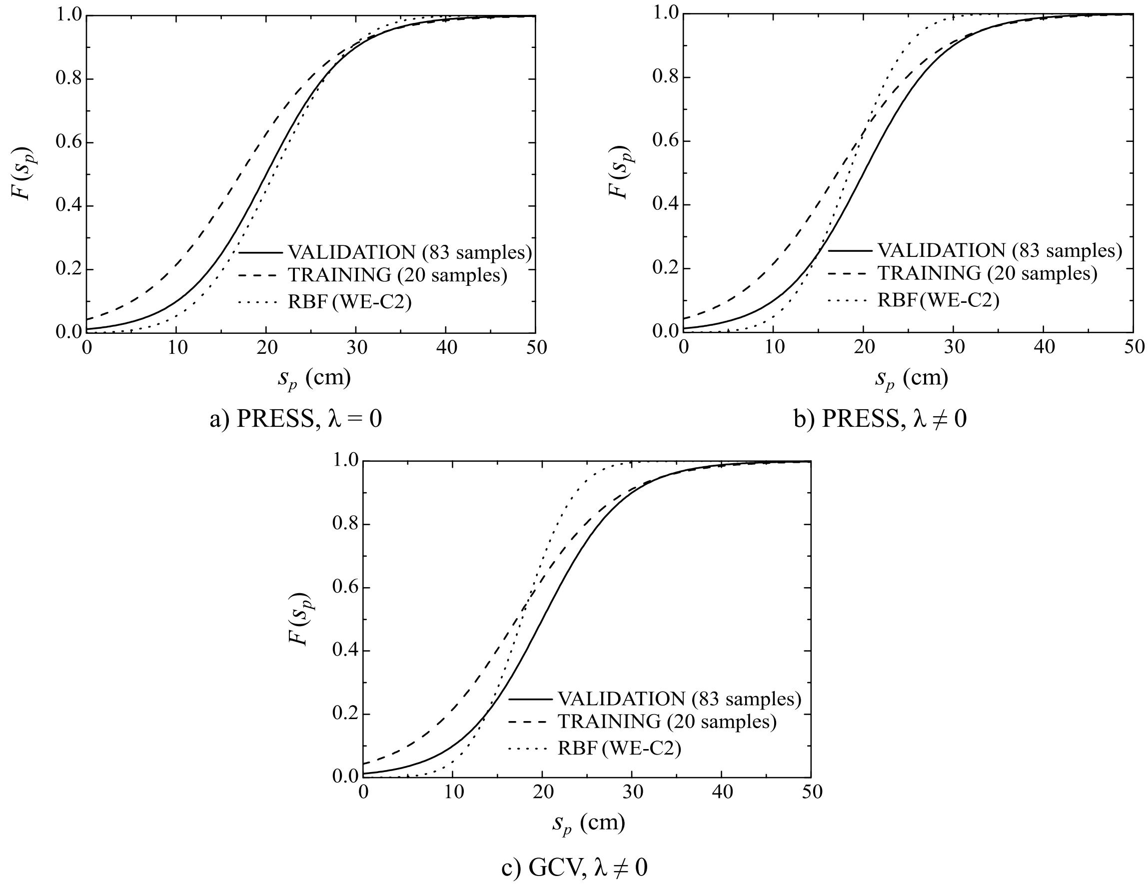
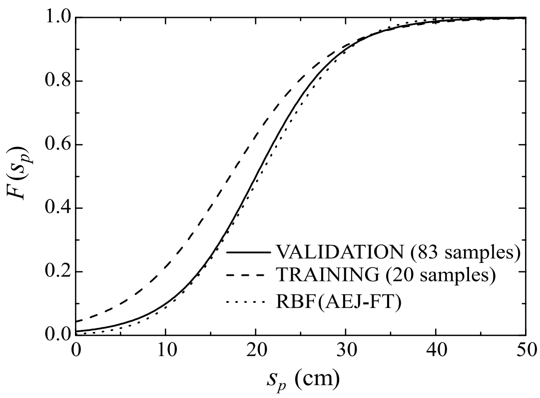
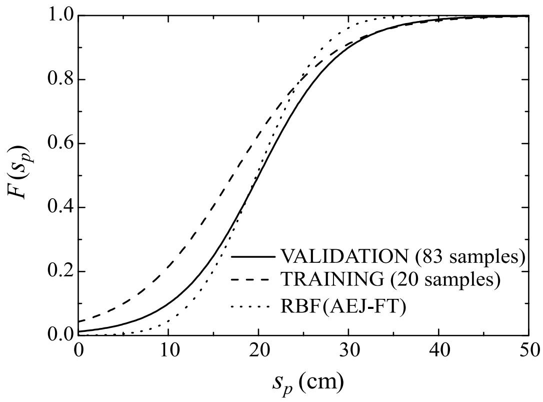


Figure 5.16 – Comparison of cumulative distribution functions (CDF) for RBF (WE-C2) of Example 03

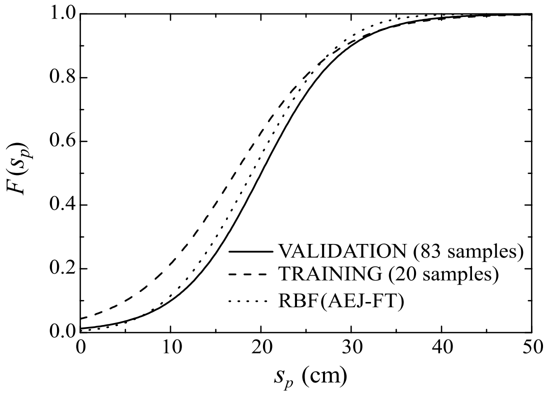
The CDF of the RBF (WE-C2) built with parameters derived from the minimization of the non-regularized PRESS presents a smaller distance in relation to the validation CDF (continuous line) and consequently a better approximation of the uncertainties of the reference model. Another important observation is the CDF of the training sample set that presented a greater distance in relation to the validation CDF. This behavior was expected due to the fact that the CDF was built with only 20 samples. This demonstrate the difficulty present in the approximation of the CDFs of the concrete slump test. Furthermore, it is noted that the presence of regularization in PRESS and GCV did not guarantee a better prediction of slump uncertainties. The results of the CDFs constructed from the surrogate models based on the RBF (AEJ-FT) are shown in Figure 5.17.



a) PRESS, λ = 0



b) PRESS, λ ≠ 0



c) GCV, λ ≠ 0

Figure 5.17 – Comparison of cumulative distribution functions (CDF) for RBF (AEJ-FT) of Example 03

Figure 5.17 shows that the RBF (AEJ-FT) has a high capacity to predict the behavior of the CDF built with the validation data (reference model). The CDF obtained from the non-regularized PRESS showed the smallest distances among all the surrogate models and, consequently did not present a statistical difference in relation to the CDF constructed from the 83 samples (continuous line). The CDFs obtained from the surrogate models based on the RBF (AEJ-FT) with parameters optimized by the PRESS and GCV with regularization were also efficient, but with a lower approximation than the PRESS without regularization. It is important to emphasize that the RBF (AEJ-FT) is the only surrogate model that was able to reproduce the slump uncertainties with both the PRESS and the GCV. These results can be better analyzed through the KS statistical test showed in Table 5.16.

Table 5.16 – Results of the KS Statistical Hypothesis Test of Example 03

Surrogate model	Method	λ	P-Value	KS	Result
RBF (GA)	PRESS	RBF ($\lambda = 0$)	0.03	0.13	H_1
		RBF ($\lambda \neq 0$)	0.00	0.15	H_1
	GCV	RBF ($\lambda \neq 0$)	0.00	0.20	H_1
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	0.82	0.06	H_0
		RBF ($\lambda \neq 0$)	0.00	0.16	H_1
	GCV	RBF ($\lambda \neq 0$)	0.00	0.21	H_1
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	1.00	0.03	H_0
		RBF ($\lambda \neq 0$)	0.52	0.07	H_0
	GCV	RBF ($\lambda \neq 0$)	0.79	0.06	H_0
Training (20 samples)	-	-	0.00	0.16	H_1

H_0: predictions of the surrogate models and reference model solution are obtained from the same population
H_1: predictions of the surrogate models and reference model solution are obtained from different populations

The KS test showed that the uncertainties of the slump test are more difficult when compared to the other examples studied, for example, in relation to Example 02, which studied the concrete compressive strength. The results predicted by the RBF (AEJ-FT) show to be from the same population as the validation samples (reference model). These results can also be confirmed by the high magnitudes of the *P* value for the RBF (AEJ-FT). However, the traditional RBFs of type gaussian and wendland-C2 are not able to predict the slump uncertainties, being only visible a good approximation for the RBF (WE-C2) with shape parameters optimized by the non-regularized PRESS. In this example, as well as in the previous example, the RBF (AEJ-FT) showed good accuracy due to the data not being randomly generated, but consisting of experimental data from the literature. However, in the first example, it was noted greater difficulty in predicting the RBF AEJ-FT in models with noise subsequently added to the output response. It is also important to emphasize that the threshold value for the KS is 0.12 for a significant level of 5% as an indication of non-rejection of the null hypothesis.

Prediction of MSE, RMSE and normalized error (ε)

The verification of the errors of the results predicted by the RBFs is performed in two ways, the first being by comparing the MSE and RMSE results extracted from Wu and coauthors [166]; and a second verification performed by the normalized mean square error (ε) of the RBFs prediction results with the slump reference model. The MSE and RMSE results are shown in Table 5.17. The MSE and RMSE results were determined both in the training and validation phase in order to verify the generalization capacity of each surrogate model. The MSE results for the surrogate models of Wu and coauthors [166] were obtained with the Equation (5.15).

Table 5.17 – MSE and RMSE error of Example 03

Surrogate model	Method	λ	Training MSE	Training RMSE	Validation MSE	Validation RMSE
RBF (GA)	PRESS	RBF (λ = 0)	5.42 E-17	7.36 E-09	85.51	9.25
		RBF (λ ≠ 0)	2.30 E+01	4.79 E+00	57.20	7.56
	GCV	RBF (λ ≠ 0)	2.67 E+01	5.17 E+00	58.49	7.65
RBF (WE-C2)	PRESS	RBF (λ = 0)	5.51 E-16	2.35 E-08	81.54	9.03
		RBF (λ ≠ 0)	2.14 E+01	4.62 E+00	56.57	7.52
	GCV	RBF (λ ≠ 0)	2.57 E+01	5.07 E+00	58.30	7.64
RBF (AEJ-FT)	PRESS	RBF (λ = 0)	4.28 E-09	6.54 E-05	124.35	11.15
		RBF (λ ≠ 0)	9.07 E+00	3.01 E+00	71.04	8.43
	GCV	RBF (λ ≠ 0)	2.42 E+00	1.56 E+00	91.87	9.58
GSy [(a)] (p = 8)	-	-	-	-	230.74 [(b)]	15.19
GSy [(a)] (p = 12)	-	-	-	-	75.52 [(b)]	8.69
GSy [(a)] (p = 20)	-	-	-	-	52.13 [(b)]	7.22
iGS [(a)] (p = 8)	-	-	-	-	230.74 [(b)]	15.19
iGS [(a)] (p = 12)	-	-	-	-	77.62 [(b)]	8.81
iGS [(a)] (p = 20)	-	-	-	-	50.27 [(b)]	7.09
GSx [(a)] (p = 8)	-	-	-	-	235.01 [(b)]	15.33
GSx [(a)] (p = 12)	-	-	-	-	79.21 [(b)]	8.90
GSx [(a)] (p = 20)	-	-	-	-	55.95 [(b)]	7.48

[(a)] Results extracted from Wu e coauthors [166]. GSx: greedly sampling on the inputs; GSy: greedy sampling on the outputs; iGS: improved greedy sampling on both inputs and output.
[(b)] Results obtained by the Equation (5.15).

In the training phase of the surrogate model based on RBFs, the error MSE and RMSE are practically null for the RBFs with parameters obtained from the optimization of the PRESS without regularization. On the other hand, during the validation phase, the MSE and RMSE errors presented magnitudes higher than those of the training phase. Although this increase in errors occurs in the validation phase, this does not mean that the surrogate models have a low capacity for generalization, since, as verified in the uncertainty quantification process, CDFs with RBF (AEJ-FT) and RBF (WE-C2) with non-regularized PRESS showed a good approximation capacity of the CDFs. The presence of regularization led an improvement in the generalization capacity of the surrogate models. In the training phase there was an increase in the magnitude of the MSE and RMSE, but a decrease in the errors during the validation phase. The MSE and RMSE errors of Wu and coauthors [166] are for three different training sample sizes where the smallest errors are only for the size p = 20. The smallest MSE and RMSE errors in the literature are for the surrogate models with iGs and GSy. The results of the RBFs, although do not present minor errors, are competitive with those showed for iGs and GSy. In addition, all the surrogate models showed a good capacity to predict the set of validation sample properly, that is, a good generalization and, consequently, overfitting was avoided. The smallest MSE and RMSE error for the RBFs during the validation phase is presented for the RBF (WE-C2) with parameters obtained from regularized PRESS.

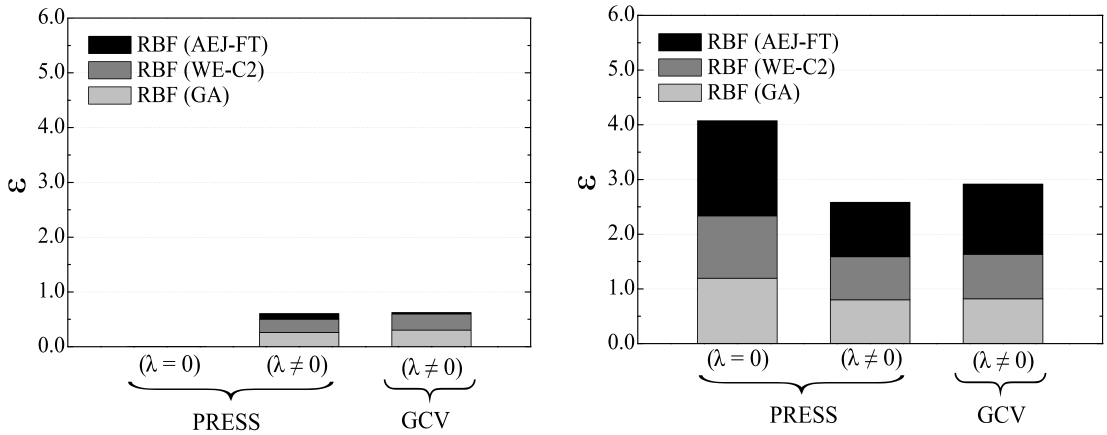
The CPU time study was not verified by the literature methods (GSy, iGS and GSx) for the Example 03. Consequently, the verification will be carried out similarly to Example 02. The surrogate models based on RBF presents a computational effort similar to the literature methods because the number of samples in the training phase is similar (p = 20) to that adopted by the literature. On the other hand, for the training sets p = 8 and p = 12, the literature shows a lower computational cost when compared to the RBFs (p = 55). This occurs because the smaller sizes of the training set require a smaller number of bases to be constructed in ridge regression models by the literature. Another training set was built for the RBFs with a size p = 12 and the respective results are shown in Table 5.18.

Table 5.18 – MSE and RMSE error for 12 training samples of Example 03

Surrogate model	Method	λ	Training		Validation	
			MSE	RMSE	MSE	RMSE
RBF (GA)	PRESS	RBF ($\lambda = 0$)	6.67 E-19	8.16 E-10	117.19	10.83
		RBF ($\lambda \neq 0$)	2.26 E+01	4.76 E+00	71.43	8.45
	GCV	RBF ($\lambda \neq 0$)	3.58 E+01	5.98 E+00	92.80	9.63
RBF (WE-C2)	PRESS	RBF ($\lambda = 0$)	7.70 E-20	2.77 E-10	87.82	9.37
		RBF ($\lambda \neq 0$)	1.87 E+01	4.32 E+00	71.86	8.48
	GCV	RBF ($\lambda \neq 0$)	3.39 E+01	5.82 E+00	95.80	9.79
RBF (AEJ-FT)	PRESS	RBF ($\lambda = 0$)	4.04 E-11	6.35 E-06	113.01	10.63
		RBF ($\lambda \neq 0$)	1.46 E+01	3.82 E+00	106.16	10.30
	GCV	RBF ($\lambda \neq 0$)	9.95 E-01	9.97 E-01	109.07	10.44
GSy [(a)] (p = 8)	-	-	-	-	230.74 [(b)]	15.19
GSy [(a)] (p = 12)	-	-	-	-	75.52 [(b)]	8.69
GSy [(a)] (p = 20)	-	-	-	-	52.13 [(b)]	7.22
iGS [(a)] (p = 8)	-	-	-	-	230.74 [(b)]	15.19
iGS [(a)] (p = 12)	-	-	-	-	77.62 [(b)]	8.81
iGS [(a)] (p = 20)	-	-	-	-	50.27 [(b)]	7.09
GSx [(a)] (p = 8)	-	-	-	-	235.01 [(b)]	15.33
GSx [(a)] (p = 12)	-	-	-	-	79.21 [(b)]	8.90
GSx [(a)] (p = 20)	-	-	-	-	55.95 [(b)]	7.48

[(a)] Results extracted from Wu e coauthors [166]. GSx: greedly sampling on the inputs; GSy: greedy sampling on the outputs; iGS: improved greedy sampling on both inputs and output.
[(b)] Results obtained by the Equation (5.15).

The training set with only 12 samples demonstrates a competitive approach to the surrogate models with iGS, GSy and GSx. Although the results of the RBFs in both training and validation are not superior to the literature, it is possible to notice that the surrogate models based on RBFs, mainly the RBF (GA) and RBF (WE-C2) present low magnitudes of RMSE and MSE error in the training phase. The results of errors in the validation phase are less competitive for the RBF (AEJ-FT). Furthermore, the presence of the regularization parameter continues to decrease MSE and RMSE errors in the validation phase. In case we consider a smaller training set for the RBFs (p = 12), the time needed to construct the bases function can present a processing time similar to the same set p = 12 used by the ridge regression models of the literature. This computational cost is superficially observed, since the focus of this work is not to compare the CPU time spent by the models. The normalized mean square error results are shown in Figure 5.18.



a) Training (20 samples) b) Validation (83 samples)

Figure 5.18 – Normalized mean square error (ε) of the surrogate models in Example 03

In the training phase, it is shown that the error is practically null for the models with parameters obtained from non-regularized PRESS. The validation phase shows the opposite in that the largest normalized errors are for the non-regularized PRESS. The regularization process in the surrogate models led to an improvement in the generalization capacity of the surrogate models. In the models with regularization, it was noticed that the RBF (AEJ-FT) with regularization (PRESS and GCV) presented smaller magnitudes of error during the training phase. In the validation step, the smallest normalized errors occurred for the RBF (WE-C2).

5.5. CONCLUSIONS

The paper developed a strategy for the construction of surrogate models based on RBFs with parameters obtained using a genetic algorithm through the minimization of PRESS and GCV, considering the use of regularization for both PRESS and GCV. Three different types of surrogate models were then employed: RBF (GA), RBF (WE-C2) and RBF (AEJ-FT). These surrogate models were trained to predict the behavior of the maximum deflection of a cantilever beam, the concrete compressive strength and the concrete slump. For the maximum deflection of the beam, thousands of values generated from Monte Carlo simulation were predicted and for the compressive strength and concrete slump, a set of experimental data was predicted, used as validation, and in all examples, was then quantified the uncertainty. The uncertainty quantification process was performed by determining the cumulative distribution functions (CDFs) of the output response of the three examples.

The uncertainty quantification was able to demonstrate the effect of the uncertainty present in the approximation of the numerical examples. The uncertainty of the input variables produced cumulative distributions showing the high uncertainty, mainly for the prediction of the maximum deflection of the beam. This uncertainty was higher for the case in which the training samples present an additional noise (case 02). In the approximation of the concrete compressive strength, it was noticed that the surrogate models with RBFs were able to predict the behavior of the compressive strength. The traditional models based on gaussian and wendland-C2 RBFs showed good accuracy for Examples 01 and 02, with better approximations in relation to the reference solutions of each numerical example. In the third example, a high uncertainty was observed for the prediction of the concrete slump in which only the RBFs of type AEJ-FT were able to adequately reproduce the behavior of the reference solution. The compact support surrogate models in all the examples presented the best approximations for the CDFs in relation to the reference solutions. In most numerical examples, it is shown that RBF Almeida Evangelista Jr. flat-top (AEJ-FT) is competitive with other surrogate models based on traditional RBFs and also in relation to the solutions of other surrogate models in the literature. The KS statistical tests showed that the surrogate models with parameters obtained from the PRESS without regularization presented the best approximation. However, it was also noted that the GCV with regularization allowed to obtain significant approximations in the numerical examples.

The MSE and RMSE errors indicated that the presence of regularization was of fundamental importance for the regression of the surrogate models. Models with PRESS and regularization showed a better generalization capacity with a significant decrease in MSE and RMSE errors in the validation phase for most cases analyzed when they were compared to the results of models without the regularization parameter. In addition, it was found that the models with regularization, including the GCV, allowed small errors in the generalization (validation phase) and consequently these surrogate models based on the RBFs proved to be efficient when compared to other surrogate models in the literature, especially in the examples of concrete compressive strength and in the concrete slump test. Another analysis verified in the training and validation phases that avoided overfitting consisted of measuring the normalized mean square error. This error showed that the presence of regularization in the validation stage (unknown data) was fundamental for the regression of the surrogate models, and it was observed that there was a decrease in the error for most of the analyzed cases when the PRESS presented the parameter of regularization. Although the results of the GCV were not better than the PRESS with regularization, they showed good efficiency, mainly because they present shorter computational cost in the optimization of the parameters of the RBFs.

The results showed that the surrogate models proposed with RBFs are efficient and accurate computational procedures to solve engineering problems involving uncertainty quantification of concrete compressive strength, concrete slump and the maximum deflection of a cantilever beam. In these numerical examples, the models with RBFs were able to generalize well the thousands of unseen data from the Monte Carlo simulations for the maximum deflection of the beam and also to predict the behavior of the compressive strength and slump for an unseen data set in the surrogate models. Furthermore, the surrogate models were able to accurately predict the cumulative distribution functions with a relatively small training sample set. Although the examples consider a cantilever beam or the results of problems such as compressive strength with 8 input variables and concrete slump with 7 input variables, the methodology presented in this work, as well as the surrogate models with RBFs, can be applied directly to different materials in other problems related to structural engineering or even to verify the behavior of cementitious materials.

6. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

6.1. CONCLUSIONS

This doctoral thesis built a surrogate modeling strategy for uncertainty quantification of age and time-dependent fracture mechanics problems and also of nonlinear functions selected from literature. In fracture mechanics problems, a study of the energy release rate for concrete structures in the presence of initial crack was considered. Three types of surrogate models were employed: the traditional second order polynomial (P2) and two proposed RBFs (Gaussian and Wendland-C2). The surrogate models were trained to substitute the time integral to predict output values of the energy release rate for different cases of loading ages and acting times. In the study of nonlinear functions, benchmark functions with nonlinearity were selected from the literature for a structural reliability process based on uncertainty quantification. The RBFs models used were the Gaussian and Wendland-C2 and a new compact support RBF was also proposed (Almeida-Evangelista Jr.). In addition, in another results chapter, a strategy for determining the shape parameters with regularized regression and with a genetic algorithm was proposed to the surrogate models of the RBF type. The surrogate models presented in this thesis were trained to predict Monte Carlo samples to the post-processing of uncertainty quantification parameters. The uncertainty quantification was then performed in most cases by determining the probability density functions (PDFs) and cumulative distribution functions (CDFs) of the output response of the proposed examples.

In chapter 02, the surrogate models based on RBF showed a better capacity to approximate the energy release rate than the P2 polynomial models. The results of the cumulative distribution functions (CDFs) indicated that the surrogate models with RBF were able to measure the uncertainty present in the energy release rate at different ages and loading times. The probability distributions showed long tails for more distant prediction times, thus showing a high uncertainty of the energy release rate further away from the time. The mean square errors also showed that the RBFs perform better than the P2 model for both proposed training sets. Mainly, the compact support RBF models (CSRBF) of Wendland-C2 that presented relatively small errors when compared to the other models. The KS statistical test stated that only models with RBF predicted the CDFs statistically similar to the reference solutions for the three cases of ages and loading times adopted. In addition, the use of LOOCV cross-validation proved to be a successful technique for ensure a good generalization capacity of each surrogate model and avoid overfit.

The chapter 03 consisted of a continuation of the uncertainty quantification process of the chapter 02 in which the statistical moments of an example of age and time-dependent fracture mechanics were determined. The results obtained for the means and standard deviations showed that the RBFs of the Wendland-C2 type have a good capacity to approximate these statistical moments. Especially for the largest sample training set (p = 180) which presented relative error magnitudes lower than the set p = 79. However, the statistical moments of kurtosis and skewness showed a difficulty for prediction by the models with RBFs, especially for the moment of skewness that presented a positive skewness behavior, that is, there is a greater concentration of data in the region to the left of the mean in the PDFs. In addition, there was a tendency in the PDFs built from the reference solutions (analytical evaluation of the energy release rate) in which the mean and standard deviation tend to increase for longer elapsed times. This behavior did not occur for the moments of kurtosis and skewness, which remained with a non-significant change for longer times. In the structural reliability analysis, it was found that the probability of failure tends to increase for longer times and consequently the reliability index tends to decrease.

The chapter 04 presented structural reliability analysis based on the uncertainty quantification process in which the results showed a good performance of the surrogate models based on RBF. These models were able to predict the behavior of the probability of failure and also the reliability index of the benchmark problems. The results also showed that the new RBF by Almeida-Evangelista Jr. presented values of reliability index and probability of failure similar to the reference models in the literature. The relative errors showed that the new RBF (AEJ-FT) has a better approximation capacity of the probability of failure when compared to other surrogate models in the literature. Especially when compared to traditional models based on RBFs and classical methods, such as FORM. The models built with RBFs presented, for most of the examples, a training set relatively inferior to those adopted by other surrogate models. Furthermore, the results obtained by the exhaustive Monte Carlo simulations for the reference models showed a high uncertainty of the numerical examples and a large variability of the random variables of each performance function. The surrogate models showed a high generalization capacity, especially the RBF Almeida-Evangelista Jr., in which this surrogate model was able to predict unseen sample set from the already trained surrogate model and consequently reproduce the structural reliability of all the numerical examples.

In chapter 05, surrogate models based on RBF were used with regularized regression and genetic algorithm in selected examples from the literature. The use of a genetic algorithm allowed obtaining optimal shape parameters for the RBFs and consequently, the overfitting of the surrogate models was avoided. The results showed a high generalization capacity of the models with RBF, mainly for the new RBF Almeida-Evangelista Jr. (AEJ-FT) that demonstrated to be competitive with other surrogate models based on RBFs in the uncertainty quantification process with cumulative distribution functions (CDFs). The CDFs showed a high uncertainty in the prediction of the examples, mainly for the example of the concrete slump test that only showed good results for the RBF (AEJ-FT). The surrogate models based on traditional RBFs (Gaussian and Wendland-C2) showed a good generalization ability for the cantilever beam example and for the concrete compressive strength example. The KS statistical test showed that the surrogate models with parameters obtained from the PRESS minimization and without the regularization process present a better fit of the CDFs in relation to the reference solution. On the other hand, the presence of regularization in the examples showed a decrease in MSE and RMSE errors in the validation phase. The presence of regularization in PRESS also showed a decrease in the normalized mean square error in the validation phase, and consequently a better generalization ability of surrogate models with regularization. The presence of regularization, however, does not ensure a better fit of the CDFs, since the surrogate models with non-regularized PRESS are the ones that present a better fit of the CDFs.

In view of the arguments presented, it can be seen that surrogate models based on RBF are efficient and accurate computational strategies for large-scale engineering problems involving uncertainty quantification of time-dependent fracture phenomena and nonlinear functions that require thousands to millions of evaluations of energy release rate integrals and limit state functions. Models with RBF were able to generalize well from thousands of unseen Monte Carlo sampling data to accurately predict the output response of multivariate problems with two, four, and even 10 random variables with very few samples in the training set. Specifically, the new compact support RBF (Almeida-Evangelista Jr.) which proved to be competitive with models with global and local support RBF. Furthermore, the surrogate models with RBF were able to accurately predict the probability density functions and cumulative distribution functions in both time-dependent problems and nonlinear functions with regularized regression. Although the fracture mechanics example consists of a concrete plate with a semi-elliptical surface crack, the methodology presented in this thesis, as well as the surrogate models with RBF, can be applied directly to different materials in other fracture problems that present the energy release rate formulated as a mapping integral.

6.2. RECOMMENDATIONS FOR FUTURE WORK

This section presents some suggestions for future work that can be developed in relation to the RBF-based surrogate modeling process and for time-dependent fracture mechanics problems. The suggestions are:

- Study possible correlations between input variables in the time-dependent fracture mechanics problem, such as correlations between geometric, loading or material parameters.
- Realize a study of the new compact support RBF function with application in other engineering problems, such as, for example, in problems of fracture mechanics of viscoelastic materials;
- Build adaptive samples to be used in the training sets or even consider other samples for training sets of the surrogate models.
- Develop new strategies for RBF-based surrogate models, such as the use of the regression trees in uncertainty quantification;
- Compare the results predicted by the RBFs with the results predicted from other surrogate models in the literature, such as the use of surrogate models with polynomial chaos expansion.
- Build subset simulation models for the new RBF AEJ-FT and use them in other uncertainty quantification problems.

REFERENCES

[1] S. K. Choi, R. V. Grandhi and R. A. Canfield, Reliability-based Structural Design, London: Springer-Verlag, 2007.

[2] A. T. Beck, Confiabilidade e Segurança das Estruturas, Rio de Janeiro: ELSEVIER, 2019.

[3] J. Zhang, "Modern Monte Carlo methods for efficient uncertainty quantification and propagation: A survey," *WIREs Computational Statistics,* vol. 13, no. 5, p. e1539, 2021.

[4] R. E. Melchers, Structural Reliability: Analysis and Prediction, 2th ed., Wiley, 1999.

[5] F. Evangelista Jr. and I. F. Almeida, "Machine Learning RBF-based surrogate models for uncertainty quantification of age and time-dependent fracture mechanics," *Engineering Fracture Mechanics,* vol. 258, p. 108037, 2021.

[6] G. Blatman, B. Sudret and M. Berveiller, "Quasi random numbers in stochastic finite element analysis," *Mécanique & Industries,* vol. 8, pp. 289-297, 2007.

[7] V. Dubourg and B. Sudret, "Meta-model-based importance sampling for reliability sensitivity analysis," *Structural Safety,* vol. 49, pp. 27-36, 2014.

[8] B. Sudret, S. Marelli and J. Wiart, "Surrogate models for uncertainty quantification: An overview," in *11th European Conference on Antennas and Propagation (EUCAP)*, Paris, France, 2017.

[9] L. Faravelli, "Response-Surface Approach for Reliability Analysis," *Journal of Engineering Mechanics, ASCE,* vol. 115, pp. 2763-2781, 1989.

[10] M. R. Rajashekhar and B. R. Ellingwood, "A New Look at the Response Surface Approach for Reliability Analysis," *Structural Safety,* vol. 12, pp. 205-220, 1993.

[11] S. Goswami, S. Ghosh and S. Chakraborty, "Reliability analysis of structures by iterative improved response surface method," *Structural Safety,* vol. 60, pp. 56-66, 2016.

[12] R. L. Hardy, "Multiquadric Equations of Topography and Other Irregular Surfaces," *Journal of Geophysical Research,* vol. 76, pp. 1905-1915, 1971.

[13] M. F. Hussain, R. R. Barton and S. B. Joshi, "Metamodeling: Radial basis functions, versus polynomials," *European Journal of Operational Research,* vol. 138, pp. 142-154, 2002.

[14] G. R. Liu and Y. T. Gu, "Meshfree Shape Function Construction," in *An Introduction to Meshfree Methods and Their Programming*, 2005, pp. 54-144.

[15] E. J. Kansa, "Multiquadrics - a scattered data approximation scheme with applications to computational fluid dynamics II," *Computers & Mathematics with Applications,* vol. 19, pp. 147-161, 1990.

[16] C. Franke and R. Schaback, "Solving partial differential equations by collocation using radial basis functions," *Applied Mathematics and Computation,* vol. 93, pp. 73-82, 1998.

[17] J. Lin, W. Chen and K. Y. Sze, "A new radial basis function for Helmholtz problems," *Engineering Analysis with Boundary Elements,* vol. 36, pp. 1923-1930, 2012.

[18] M. J. L. Orr, "Regularisation in the Selection of Radial Basis Function Centres," *Neural Computation,* vol. 7, no. 3, pp. 1 - 16, 1995.

[19] M. J. L. Orr, Introduction to Radial Basis Function Networks, Technical Report. Institute for Adaptive and Neural Computation, Division of Informatics, Edinburgh University, 1996.

[20] M. J. L. Orr, *Regularised Centre Recruitment in Radial Basis Function Networks,* Research Paper 59, Centre for Cognitive Science, Edinburgh University, 1993.

[21] S. Chen, E. S. Chng and K. Alkadhimi, "Regularized orthogonal least squares algorithm for constructing radial basis function networks," *International Journal of Control,* vol. 64, no. 5, pp. 829 - 837, 1996.

[22] I. P. Schagen, "Interpolation in Two Dimensions – A New Technique," *IMA Journal of Applied Mathematics,* vol. 23, no. 1, pp. 53-59, 1979.

[23] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree," *Advances in Computational Mathematics,* vol. 4, p. 389–396, 1995.

[24] F. Evangelista Jr. and N. Afanador-García, "A polynomial chaos expansion approach to the analysis of uncertainty in viscoelastic structural elements," *DYNA,* vol. 83, no. 199, pp. 172 - 182, 2016.

[25] N. S. Narváez, N. R. Rojas and F. Evangelista Jr., "Reliability analyses of shear strengthened RC beams with externally bonded fiber reinforced polymer," *Materials and Structures*, vol. 53, pp. 1-19, 2020.

[26] S. Mangalathu, J. S. Jeon and R. DesRoches, "Critical uncertainty parameters influencing seismic performance of bridges using Lasso regression," *Earthquake Engineering & Structural Dynamics,* vol. 47, pp. 784 - 801, 2018.

[27] M. A. Hariri-Ardebili and B. Sudret, "Polynomial chaos expansion for uncertainty quantification of dam engineering problems," *Engineering Structures,* vol. 203, p. 109631, 2020.

[28] Q. Zhang, Y. Zhao, K. Kolozvari and L. Xu, "Reliability analysis of reinforced concrete structure against progressive collapse," *Reliability Engineering and System Safety*, vol. 228, p. 108831, 2022.

[29] P. Criel, N. Reybrouck, R. Caspeele and S. Matthys, "Uncertainty quantification of creep in concrete by Taylor expansion," *Engineering Structures* , vol. 153, pp. 334-341, 2017.

[30] X. Li, C. Gong, L. Gu, W. Gao, Z. Jing and H. Su, "A sequential surrogate method for reliability analysis based on radial basis function," *Structural Safety*, vol. 73, pp. 42 -53, 2018.

[31] J. Xu and S. Zhu, "An efficient approach for high-dimensional structural reliability analysis," *Mechanical Systems and Signal Processing*, vol. 122, pp. 152-170, 2019.

[32] Z. P. Bažant and L. Panula, "Pratical prediction of time dependent deformations of concrete," *Materials and Structures, Part I and II,* vol. 11, no. 65, pp. 307-328, 1978.

[33] Z. P. Bažant and S. Baweja, "Justification and refinements of Model B3 for concrete creep and shrinkage 1. statistics and sensitivity," *Materials and Structures,* vol. 28, no. 7, pp. 415-430, 1995.

[34] Z. P. Bažant, M. H. Hubler and Q. Yu, "Pervasiveness of Excessive Segmental Bridge Deflections: Wake-Up Call for Creep," *ACI Structural Journal,* vol. 108, no. 6, pp. 766-774, 2011.

[35] Z. P. Bažant, Q. Yu and G. Li, "Excessive Long-Time Deflections of Prestressed Box Girders. I: Record-Span Bridge in Palau and Other Paradigms," *Journal of Structural Engineering,* vol. 138, p. 676 – 696, 2012.

[36] L. Su, Y. F. Wang, S. Q. Mei and P. F. Li, "Experimental investigation on the fundamental behaviour of concrete creep," *Construction and Building Materials,* vol. 152, p. 250–258, 2017.

[37] M. Rasoolinejad, S. Rahimi-Aghdam and Z. P. Bažant, "Statistical filtering of useful concrete creep data from imperfect laboratory tests," *Materials and Structures,* vol. 51, p. 1 – 14, 2018.

[38] M. Sichani, M. Hanifehzadeh, J. Padgett and B. Gencturk, "Probabilistic analysis of vertical concrete dry casks subjected to tip-over and aging effects," *Nuclear Engineering and Design,* vol. 343, pp. 232-247, 2019.

[39] W. Liu, S. Zhang, B. Sun and L. Chen, "Creep characteristics and time-dependent creep model of tunnel lining structure concrete," *Mechanics of Time-Dependent Materials,* 2020.

[40] M. Hariri-Ardebili, L. Sanchez and R. Rezakhani, "Aging of Concrete Structures and Infrastructures: Causes, Consequences, and Cures (C3)," *Advances in Materials*

Science and Engineering, p. 9370591, 2020.

[41] ACI, "Prediction of creep, shrinkage and temperature effects in concrete structures. ACI-209R-82, ACI Committee 209," American Concrete Institute (ACI), Detroit, 1982.

[42] CEB-FIP, "Model Code 1990: Design Code 1994," Thomas Telford, London, 1994.

[43] RILEM, "Creep and shrinkage prediction model for analysis and design of concrete structures - model B3," *Materials and Structures,* vol. 28, pp. 357-365, 1995.

[44] RILEM, "Errata: creep and shrinkage prediction model for analysis and design of concrete structures - model B3," *Materials and Structures,* vol. 29, p. 126, 1995.

[45] Z. P. Bažant and S. Prasannam, "Solidification theory for concrete creep I: formulation," *Journal of Engineering Mechanics,* vol. 115, no. 8, pp. 1691-1703, 1989.

[46] Z. P. Bažant, A. Hauggaard, S. Baweja and F. Ulm, "Microprestress-Solidification Theory for Concrete Creep. I: Aging and Drying Effects," *Journal of Engineering Mechanics*, vol. 123, no. 11, pp. 1188-1194, 1997.

[47] Y. Wei, W. Guo and S. Liang, "Microprestress-solidification theory-based tensile creep modeling of early-age concrete: Considering temperature and relative humidity effects," *Construction and Building Materials,* vol. 127, pp. 618-626, 2016.

[48] H. B. Motra, "Uncertainty and Sensitivity Analysis of Cyclic Creep Models of Concrete," *Asian Journal of Applied Sciences*, vol. 8, no. 4, p. 240 – 258, 2015.

[49] H. Keitel and A. Dimmig-Osburg, "Uncertainty and sensitivity analysis of creep models for uncorrelated and correlated input parameters," *Engineering Structures,* vol. 32, no. 11, pp. 3758 -3767, 2010.

[50] H. Keitel, "Quantifying sources of uncertainty for creep models under varying stresses," *Journal of Structural Engineering,* vol. 139, no. 6, pp. 949 -956, 2013.

[51] J. Ortega, G. Ruiz, R. Yu, N. Afanador-García, M. Tarifa, E. Poveda, X. Zhang and F. Evangelista Jr., "Number of tests and corresponding error in concrete fatigue," *International Journal of Fatigue,* vol. 116, pp. 210-219, 2018.

[52] T. A. A. Oliveira, G. Gomes and F. Evangelista Jr., "Multiscale aircraft fuselage fatigue analysis by the dual boundary element method," *Engineering Analysis with Boundary Elements,* vol. 104, pp. 107-119, 2019.

[53] J. F. Borges, V. F. de Paula, F. Evangelista Jr. and L. M. Bezerra, "Reliability and uncertainty quantification of the net section tension capacity of cold-formed steel angles with bolted connections considering shear lag," *Advances in Structural Engineering*, vol. 1, no. 1, pp. 1-30, 2020.

[54] S. S. Jin, S. L. Cha and H. J. Jung, "Improvement of concrete creep prediction with probabilistic forecasting method under model uncertainty," *Construction and Building Materials*, vol. 184, pp. 617 -633, 2018.

[55] H. Sousa, L. O. Santos and M. Chryssanthopoulos, "Quantifying monitoring requirements for predicting creep deformations through Bayesian updating methods," *Structural Safety*, vol. 76, p. 40–50, 2019.

[56] M. Berveiller, Y. L. Pape, B. Sudret and F. Perrin, "Updating the long-term creep strains in concrete containment vessels by using Markov chain Monte Carlo simulation and polynomial chaos expansions," *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance*, vol. 8, no. 5, p. 425 – 440, 2012.

[57] F. Barpi and S. Valente, "Fuzzy parameters analysis of time-dependent fracture of concrete dam models," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 26, no. 10, pp. 1005-1027, 2002.

[58] Z. P. Bažant, J. K. Kim and S. E. Jeon, "Cohesive Fracturing and Stresses Caused by Hydration Heat in Massive Concrete Wall," *Journal of Engineering Mechanics*, vol. 129, pp. 21-30, 2003.

[59] J. Saliba, A. Loukili, F. Grondin and J. P. Regoin, "Experimental study of creep-damage coupling in concrete by acoustic emission technique," *Materials and Structures*, vol. 45, p. 1389–1401, 2012.

[60] J. Torres, Three-dimensional numerical analysis of reflective cracks in airfield pavements. Doctoral dissertation., University of Illinois at Urbana Champaign, UIUC, USA, 2013.

[61] R. H. Myers, D. C. Montgomery and C. M. Anderson-Cook, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 3rd ed., New Jersey: John Wiley & Sons, 2009.

[62] D. Montgomery, Design and Analysis of Experiments, 8th ed., John & Sons, 2012.

[63] C. M. Bishop, Pattern recognition and machine learning, New York: Springer, 2013.

[64] M. D. Buhmann, Radial Basis Functions, Cambridge Press, 2003.

[65] G. E. Fasshauer, Meshfree approximation methods with MATLAB. Interdisciplinary Mathematical Sciences, vol. 6, Illinois Institute of Technology: World Scientific, 2007.

[66] M. D. Buhmann, S. De Marchi and E. Perracchione, "Analysis of a new class of rational RBF expansions," *IMA Journal of Numerical Analysis*, p. 1 – 22, 2019.

[67] J. C. J. Newman and I. S. Raju, "An empirical stress-intensity factor equation for the

surface crack," *Engineering fracture mechanics,* vol. 15, no. 1-2, pp. 185 -192, 1981.

[68] P. Araújo Junior, J. B. Soares, A. Holanda, E. Parente Junior and F. Evangelista Junior, "Dynamic Viscoelastic Analysis of Asphalt Pavements using a Finite Element Formulation," *Road Materials and Pavement Design,* vol. 11, no. 2, pp. 409-433, 2010.

[69] D. S. Atrushi, Tensile and compressive creep of early age concrete: testing and modelling, Department of Civil Engineering, The Norwegian University of Science and technology, Trondheim, Norway: Doctoral thesis, 2003.

[70] JCSS, Joint Committee on Structural Safety. Probabilistic Model Code, Joint Commitee on Structural Safety. ISBN 978-3-909386-79-6, 2001.

[71] B. Efron and R. Tibshirani, An Introduction to the Bootstrap, Boca Raton: FL: Chapman & Hall/CRC, 1993.

[72] Z. Govindarajulu, Elements of Sampling Theory and Methods, Upper Saddle River: NJ: Prentice Hall, 1999.

[73] M. D. Shields, K. Teferra, A. Hapij and R. P. Daddazio, "Refined Stratified Sampling for efficient Monte Carlo based uncertainty quantification," *Reliability Engineering and System Safety,* vol. 142, p. 310–325, 2015.

[74] M. D. Shields and J. Zhang, "The generalization of Latin hypercube sampling," *Reliability Engineering and System Safety,* vol. 148, p. 96–108, 2016.

[75] T. Tarpey, "A note on the Prediction Sum of Squares Statistic for Restricted Least Squares," *The American Statistician,* vol. 54, no. 2, p. 116–118, 2000.

[76] Z. P. Bazant, "Prediction of Concrete Creep Effects Using Age-adjusted Effective Modulus Method," *Journal of the American Concrete Institute,* vol. 69, pp. 212-217, 1972.

[77] F. Evangelista Junior, F. C. Macedo and M. M. Farias, "A methodology to assess the evolution of viscoelastic properties of hardened cement pastes through dynamic mechanical analysis," *Construction and Building Materials,* vol. 226, pp. 849-858, 2019.

[78] Y. S. Ma and Y. F. Wang, "Creep effects on the Reliability of a Concrete-Filled Steel Tube Arch Bridge," *Journal of Bridge Engineering,* vol. 18, pp. 1095-1104, 2013.

[79] B. Han, T. Y. Xiang and H. B. Xie, "A Bayesian inference framework for predicting the long-term deflection of concrete structures caused by creep and shrinkage," *Engineering Structures,* vol. 142, pp. 46-55, 2017.

[80] Z. Zheng, D. Hu, P. Liu, F. Sha, L. Liu and Z. Yu, "Considering the effect of the randomness of concrete strength and relative humidity on concrete creep," *Structural*

Concrete - Journal of the fib, vol. 22, no. S1, pp. E916-E930, 2021.

[81] M. Hinze, S. Xiao, A. Schmidt and W. Nowak, "Experimental evaluation and uncertainty quantification for a fractional viscoelastic model of salt concrete," *Mechanics of Time-Dependent Materials,* 2022.

[82] N. C. Xiao, M. J. Zuo and C. Zhou, "A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis," *Reliability Engineering and System Safety,* vol. 169, pp. 330-338, 2018.

[83] F. Biglari, K. M. Nikbin, I. W. Goodall and G. A. Webster, "Determination of fracture mechanics parameters J and C* by finite element and reference stress methods for a semi-elliptical flaw in a plate," *International Journal of Pressure Vessels and Piping,* vol. 80, no. 7-8, pp. 565-571, 2003.

[84] A. H. Elhewy, E. Mesbahi and Y. Pu, "Reliability analysis of structures using neural network method," *Probabilistic Engineering Mechanics,* vol. 21, p. 44 – 53, 2006.

[85] I. Depina, T. M. H. Le, G. Fenton and G. Eiksund, "Reliability analysis with Metamodel Line Sampling," *Structural Safety,* vol. 60, pp. 1-15, 2016.

[86] A. S. Nowak and K. R. Collins, Reliability of Structures, United States: McGraw-Hill Companies, 2000.

[87] H. B. Nezhad, M. Miri and M. R. Ghasemi, "New neural network-based response surface method for reliability analysis of structures," *Neural Computing and Applications,* vol. 31, pp. 777 -791, 2019.

[88] B. Sudret, "Meta-models for structural reliability and uncertainty quantification," in *Fifth Asian-Pacific Symposium on Structural Reliability and its Applications (5APSSRA)*, 2012.

[89] R. E. Melchers and M. Ahammed, "A fast approximate method for parameter sensitivity estimation in Monte Carlo structural reliability," *Computers and Structures*, vol. 82, pp. 55 - 61, 2004.

[90] R. Rubinstein and D. Kroese, Simulation and the Monte Carlo method, Wiley Series in Probability and Statistics. Wiley, 2008.

[91] M. Hohenbichler and R. Rackwitz, "Non-normal dependent vectors in structural safety," *Journal of the Engineering Mechanics Division*, vol. 107, pp. 1227-1238, 1981.

[92] K. Breitung, "Asymptotic Approximations for Multinormal Integrals," *Journal of Engineering Mechanics,* vol. 110, pp. 357-366, 1984.

[93] L. Tvedt, "Distributions of quadratic forms in normal space – applications to structural reliability," *Journal of Engineering Mechanics*, vol. 116, p. 1183–1197,

1990.

[94] O. Ditlevsen, R. Olesen and G. Mohr, "Solution of a class of load combination problems by directional simulation," *Structural Safety,* vol. 4, pp. 95-109, 1987.

[95] P. Bjerager, "Probability integration by directional simulation," *Journal of Engineering Mechanics,* vol. 114, pp. 1285-1302, 1988.

[96] M. Hohenbichler and R. Rackwitz, "Improvement of second-order reliability estimates by importance sampling," *Journal of Engineering Mechanics,* vol. 114, pp. 2195-2199, 1988.

[97] S. Au and J. Beck, "A new adaptive importance sampling scheme for reliability calculations," *Structural Safety,* vol. 21, pp. 135-158, 1999.

[98] H. Dai and Z. Cao, "A wavelet Support Vector Machine-Based Neural Network Metamodel for Structural Reliability Assessment," *Computer-Aided Civil and Infrastructure Engineering,* vol. 32, pp. 344-357, 2017.

[99] Q. Wang and H. Fang, "Reliability analysis of tunnels using an adaptive RBF and a first-order reliability method," *Computers and Geotechnics,* vol. 98, p. 144–152, 2018.

[100] L. Hong, H. Li and J. Fu, "A novel surrogate-model based active learning method for structural reliability analysis," *Computer Methods in Applied Mechanics and Engineering,* vol. 394, p. 114835, 2022.

[101] C. Zhou, H. Zhang, M. A. Valdebenito and H. Zhao, "A general hierarchical ensemble-learning framework for structural reliability analysis," *Reliability Engineering and System Safety,* vol. 225, p. 108605, 2022.

[102] H. M. Kroetz, R. K. Tessari and A. T. Beck, "Performance of global metamodeling techniques in solution of structural reliability problems," *Advances in Engineering Software,* vol. 114, pp. 394 - 404, 2017.

[103] R. Schöbi, B. Sudret and S. Marelli, "Rare event estimation using Polynomial-Chaos Kriging," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering,* vol. 3, no. 2, 2017.

[104] J. Guo, J. Zhao and S. Zeng, "Structural reliability analysis based on analytical maximum entropy method using polynomial chaos expansion," *Structural and Multidisciplinary Optimization,* vol. 58, pp. 1187-1203, 2018.

[105] L. Hong, H. Li and K. Peng, "A combined radial basis function and adaptive sequential sampling method for structural reliability analysis," *Applied Mathematical Modelling,* vol. 90, pp. 375-393, 2021.

[106] Q. X. Lieu, K. T. Nguyen, K. D. Dang, S. Lee, J. Kang and J. Lee, "An adaptive surrogate model to structural reliability analysis using deep neural network," *Expert*

Systems with Applications, vol. 189, p. 116104, 2022.

[107] J. Deng, "Structural reliability analysis for implicit performance function using radial basis function network," *International Journal of Solids and Structures,* vol. 43, pp. 3255-3291, 2006.

[108] Q. Wang, H. Fang and L. Shen, "Reliability analysis of tunnels using a metamodeling technique based on augmented radial basis function," *Tunnelling and Underground Space Technology,* vol. 56, p. 45–53, 2016.

[109] K. Amouzgar and N. Strömberg, "Radial basis functions as surrogate models with a priori bias in comparison with a posteriori bias," *Structural and Multidisciplinary Optimization,* vol. 55, pp. 1453-1469, 2017.

[110] J. Biazar and M. Hosami, "An interval for the shape parameter in radial basis function approximation," *Applied Mathematics and Computation,* vol. 315, p. 131 – 149, 2017.

[111] S. A. Sarra and Y. Bai, "A rational radial basis function method for accurately resolving discontinuities and steep gradients," *Applied Numerical Mathematics,* vol. 130, pp. 131-142, 2018.

[112] Y. Liu, Z. Yang and X. Li, "Adaptive ensemble learning of radial basis function for efficient geotechnical reliability analysis," *Computers and Geotechnics,* vol. 146, p. 104753, 2022.

[113] M. Griebel and M. A. Schweitzer, "A Particle-Partition of Unity Method-Part II: Efficient Cover Construction and Reliable Integration," *SIAM Journal on Scientific Computing,* vol. 23, no. 5, p. 1655 – 1682, 2002.

[114] Q. Zhang, U. Banerjee and I. Babuška, "Higher order stable generalized finite element method," *Numerische Mathematik,* vol. 128, no. 1, p. 1–29, 2014.

[115] F. M. Sato, D. Piedade Neto and S. P. B. Proença, "Numerical experiments with the Generalized Finite Element Method based on a flat-top Partition of Unity," *Latin American Journal of Solids and Structures,* vol. 15, pp. 1679-7825, 2018.

[116] I. Babuška, U. Banerjee and J. E. Osborn, "Superconvergence in the generalized finite element method," *Numerische Mathematik,* vol. 107, no. 3, p. 353 – 395, 2007.

[117] J. Marcelin, "A metamodel using neural networks and genetic algorithms for an integrated optimal design of mechanisms," *The International Journal of Advanced Manufacturing Technology,* vol. 24, pp. 708 - 714, 2004.

[118] G. Vosniakos, A. Teifakis and P. Benardos, "Neural network simulation metamodels and genetic algorithms in analysis and design of manufacturing cells," *The International Journal of Advanced Manufacturing Technology,* vol. 29, pp. 541 - 550, 2006.

[119] C. Harpham, C. W. Dawson and M. R. Brown, "A review of genetic algorithms applied to training radial basis function networks," *Neural Computing & Applications,* vol. 13, pp. 193 - 201, 2004.

[120] F. Afiatdoust and M. Esmaeilbeigi, "Optimal variable shape parameters using genetic algorithm for radial basis function approximation," *Ain Shams Engineering Journal,* vol. 6, pp. 639-647, 2015.

[121] X. Tan, W. Bi, X. Hou and W. Wang, "Reliability analysis using radial basis function networks and support vector machines," *Computers and Geotechnics,* vol. 38, p. 178–186, 2011.

[122] Y. Zheng and P. K. Das, "Improved response surface method and its application to stiffened plate reliability analysis," *Engineering Structures,* vol. 22, p. 544–551, 2000.

[123] L. Schueremans and D. Van Gemert, "Benefit of Splines and neural networks in simulation based structural reliability analysis," *Structural Safety,* vol. 27, pp. 246-261, 2005.

[124] C. Weidong, X. Chunlong, S. Yaquin, M. Jingxin and L. Shengzhuo, "A hybrid Kriging-based reliability method for small failure probabilities," *Reliability Engineering and System Safety,* vol. 189, pp. 31-41, 2019.

[125] X. Zhang, Z. Lu and K. Cheng, "AK-DS: An adaptive Kriging-based directional sampling method for reliability analysis," *Mechanical Systems and Signal Processing,* vol. 156, p. 107610, 2021.

[126] J. Wang, G. Xu, Y. Li and A. Kareem, "AKSE: A novel adaptive Kriging method combining sampling region scheme and error-based stopping criterion for structural reliability analysis," *Reliability Engineering and System Safety,* vol. 219, p. 108214, 2022.

[127] J. Zhao, C. Jianqiao and L. Xu, "RBF-GA: An adaptive radial basis function metamodeling with genetic algorithm for structural reliability analysis," *Reliability Engineering and System Safety,* vol. 189, pp. 42 - 57, 2019.

[128] K. Hass, "Limiting Machine Learning Overfitting Uncertainties Through Persistent Homology," in *Verification and Validation Symposium (VVS), Proceedings of the ASME*, 2020.

[129] R. K. Chaudhary, R. V. Coile and T. Gernay, "Potential of Surrogate Modelling for Probabilistic Fire Analysis of Structures," *Fire Technology,* vol. 57, pp. 3151 - 3177, 2021.

[130] Z. Pan, W. Lu, Y. Fan and J. Li, "Identification of groundwater contamination sources and hydraulic parameters based on bayesian regularization deep neural network," *Environmental Science and Pollution Research,* vol. 28, pp. 16867 -

16879, 2021.

[131] G. H. Golub, M. Heath and G. Wahba, "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics,* vol. 21, no. 2, pp. 215 - 223, 1979.

[132] R. J. Praga-Alejo, D. S. González-González, M. Cantú-Sifuentes and L. M. Torres-Treviño, "The ridge method in a radial basis function neural network," *The International Journal of Advanced Manufacturing Technology,* vol. 79, pp. 1787-1796, 2015.

[133] X. Wang, J. Kou and W. Zhang, "Unsteady aerodynamic modeling based on fuzzy scalar radial basis function neural networks," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering,* vol. 233, no. 14, pp. 5107 - 5121, 2019.

[134] Y. Han, C. Fan, Z. Geng, B. Ma, D. Cong, K. Chen and B. Yu, "Energy efficient building envelope using novel RBF neural network integrated affinity propagation," *Energy,* vol. 209, p. 118414, 2020.

[135] Z. Wang, H. Xu, L. Xia, Z. Zou and C. Guedes Soares, "Kernel-based support vector regession for nonparametric modeling of ship maneuvering motion," *Ocean Engineering,* vol. 216, p. 107994, 2020.

[136] A. N. Tikhonov and V. Y. Arsenin, Solutions of Ill-Posed Problems, Winston, Washington, 1977.

[137] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics,* vol. 12, no. 1, pp. 55 - 67, 1970.

[138] Q. Zhou, S. Song, C. Wu and G. Huang, "Kernelized LARS-LASSO for constructing radial basis function neural networks," *Neural Computing and Applications,* vol. 23, pp. 1969 - 1976, 2013.

[139] A. Xiong, Y. You and L. Long, "L-RBF: A Customer Churn Prediction Model Based on Lasso+RBF," in *International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData),* Atlanta, GA, USA, 2019.

[140] C. S. Lee and J. S. Jeon, "Adaptive hysteretic model for reinforced concrete columns including variations in axial force and shear span length," *Earthquake Engineering Structural Dynamics,* vol. 50, no. 15, pp. 4001 - 4031, 2021.

[141] Y. Yan, H. Huang and L. Sun, "Multi-parameter seismic fragility and sensitivity analysis of long-span cable-stayed bridge based on multi-task lasso regression," *Structures,* vol. 38, pp. 1515-1528, 2022.

[142] T. Goel and N. Stander, "Comparing three error criteria for selecting radial basis function network topology," *Computer Methods in Applied Mechanics and Engineering*, vol. 198, no. 27-29, pp. 2137 - 2150, 2009.

[143] L. Zhang, M. Zheng, D. Du, Y. Li, M. Fei and Y. Guo, "State-of-Charge Estimation of Lithium-Ion Battery Pack Based on Improved RBF Neural Networks," *Complexity*, p. 88400240, 2020.

[144] C. A. A. Duineveld, A. K. Smilde and D. A. Doornbos, "Comparison of experimental designs combining process and mixture variables - Parte II. Design evaluation on measured data," *Chemometrics and Intelligent Laboratory Systems,* vol. 19, pp. 309 - 318, 1993.

[145] R. Tenzer and R. Klees, "The choice of the spherical radial basis functions in local gravity field modeling," *Studia Geophysica et Geodaetica*, vol. 52, pp. 287 - 304, 2008.

[146] E. Menga, C. López, S. Hernández, A. Baldomir, I. Romero and M. J. Sánchez, "Comparison of metamodelling techniques for uncertainty quantification and global sensitivity analysis in assembled aircraft structures," in *AIAA Non-Deterministic Approaches Conference*, Kissimmee, Florida, 2018.

[147] B. Mu, T. Chen and L. Ljung, "Asymptotic Properties of Generalized Cross Validation Estimators for Regularized System Identification," *IFAC - PapersOnLine,* vol. 51, no. 15, pp. 203-208, 2018.

[148] S. Xu, "Predicted Residual Error Sum of Squares of Mixed Models - An application for Genomic Prediction," *G3 Genes\Genomes\ Genetics,* vol. 7, no. 3, pp. 895-909, 2017.

[149] J. Holland, Adaptation in natural and artificial systems, MI: University of Michigan Press, Ann Arbor, 1975.

[150] S. A. Billings and G. L. Zheng, "Radial Basis Function Network Configuration Using Genetic Algorithms," *Neural Networks,* vol. 8, no. 6, pp. 877 - 890, 1995.

[151] W. Chang, "An RBF Neural Network Combined with OLS Algorithm ang Genetic Algorithm for Short-Term Wind Power Forecasting," *Journal of Applied Mathematics*, vol. 13, p. 971389, 2013.

[152] I. G. Tsoulos, N. Anastasopoulos, G. Ntritsos and A. Tzallas, "Train RBF networks with a hybrid genetic algorithm," *Evolutionary Intelligence,* 2021.

[153] D. F. Cook, C. T. Ragsdale and R. L. Major, "Combining a neural network with a genetic algorithm for process parameter optimization," *Engineering Applications of Artificial Intelligence,* vol. 13, pp. 391-396, 2000.

[154] L. Kuncheva, "Initializing of an RBF network by a genetic algorithm,"

Neurocomputing, vol. 14, pp. 273 - 288, 1997.

[155] O. Akbilgic and H. Bozdogan, "Predictive Subset Selection using Regression Trees and RBF Neural Networks Hybridized with the Genetic Algorithm," *European Journal of Pure and Applied Mathematics,* vol. 4, no. 4, pp. 467 - 486, 2011.

[156] S. Chen, Y. Wu and B. Luk, "Combined Genetic Algorithm Optimization and Regularized Orthogonal Least Squares Learning for Radial Basis Function Networks," *IEEE Transactions on Neural Networks,* vol. 10, no. 5, pp. 1239 - 1243, 1999.

[157] M. Esmaeilbeigi and M. N. Hosseini, "A new approach based on the genetic algorithm for finding a good shape parameter in solving partial differential equations by Kansa's method," *Applied Mathematics and Computation,* vol. 249, pp. 419 - 428, 2014.

[158] C. Yan, L. Wang, W. Liu and M. Qi, "Financial early warning of non-life insurance company based on RBF neural network optimized by genetic algorithm," *Concurrency and Computation - Pratice and Experience,* vol. 30, no. 23, pp. 1-11, 2017.

[159] D. Goldberg, "Genetic Algorithms In Search, Optimization, and Machine Learning," in *Ethnographic Praxis in Industry Conference Proceedings,* 1988.

[160] L. Sahoo, A. Bhunia and P. Kapur, "Genetic algorithm based multi-objective reliability optimization in interval environment," *Computers & Industrial Engineering,* vol. 62, no. 1, pp. 152-160, 2012.

[161] H. Chen, H. Yong and Y. Zhou, "Crack detection in bulk superconductor using Genetic Algorithm," *Engineering Fracture Mechanics,* vol. 265, p. 108372, 2022.

[162] S. N. Sivanandam and S. N. Deepa, "Genetic Algorithm Implementation Using Matlab," in *Introduction to Genetic Algorithms*, Berlin, Heidelberg, Springer, 2008, pp. 211 - 262.

[163] D. Dua and C. Graff, "UCI Machine Learning Repository," Irvine, CA: University of California, School of Information and Computer Science, 3 August 2007. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength. [Accessed 5 May 2021].

[164] I. C. Yeh, "Modeling of strength of high performance concrete using artificial neural networks," *Cement and Concrete Research,* vol. 28, no. 12, pp. 1797 - 1808, 1998.

[165] I. C. Yeh, "Prediction of strength of fly Ash and Slag concrete by the use of artificial neural networks," *Journal of the Chinese Institute of Civil and Hydraulic Engineering,* vol. 15, no. 4, pp. 659 - 663, 2003.

[166] D. Wu, C. T. Lin and J. Huang, "Active Learning for regression using greedy

sampling," *Information Sciences,* vol. 474, pp. 90 - 105, 2019.

[167] H. Yu and S. Kim, "Passive sampling for regression," in *IEEE Int'l. Conf. on Data Mining*, Sydney, Australia, 2010.

[168] D. Dua and C. Graff, "UCI Machine Learning Repository," Irvine, CA: University of California, School of Information and Computer Science, 30 April 2009. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Concrete+Slump+Test. [Accessed 10 May 2021].

[169] I. C. Yeh, "Modeling slump flow of concrete using second-order regressions and artificial neural networks," *Cement & Concrete Composites,* vol. 29, no. 6, pp. 474 - 480, 2007.

[170] I. C. Yeh, "Simulation of concrete slump using neural networks," *Construction Materials,* vol. 162, no. 1, pp. 11 - 18, 2009.

[171] D. Wu, "Pool-Based Sequential Active Learning for Regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1348 - 1359, 2019.

APPENDIX

APPENDIX A

The geometric functions used in Equations (2.10) to (2.13) as obtained by [67] through a finite element approximation:

$$Q(a,c) = 1 + 1.464\left(\frac{a}{c}\right)^{1.65} \quad \text{for } \frac{a}{c} \le 1$$

$$Q(a,c) = 1 + 1.464\left(\frac{c}{a}\right)^{1.65} \quad \text{for } \frac{a}{c} > 1$$

$$F\left(\frac{a}{e_0}, \frac{a}{c}, \frac{c}{W_t}, \phi_a\right) = \left[M_1 + M_2\left(\frac{a}{e_0}\right)^2 + M_3\left(\frac{a}{e_0}\right)^4\right] f_{\phi_a} f_w f_g$$

$$M_1 = 1.13 - 0.09\left(\frac{a}{c}\right)$$

$$M_2 = -0.54 + \frac{0.89}{0.2 + (a/c)} \qquad \text{(A1-A9)}$$

$$M_3 = 0.5 - \frac{1.0}{0.65 + (a/c)} + 14\left(1.0 - \frac{a}{c}\right)^{24}$$

$$f_{\phi_a} = \left[\left(\frac{a}{c}\right)^2 \cos^2\phi_a + \sin^2\phi_a\right]^{1/4}$$

$$f_g = 1 + \left[0.1 + 0.35\left(\frac{a}{e_0}\right)^2\right]\left(1 - \sin\phi_a\right)^2$$

$$f_w = \left[\sec\left(\frac{\pi c}{W_t}\right)\sqrt{\frac{a}{e_0}}\right]^{1/2}$$

The auxiliary functions of the geometric function given in Equation (2.12) are:

$$H_p = H_1 + \left(H_2 - H_1\right)\sin^p\phi_a$$

$$p = 0.20 + \left(\frac{a}{c}\right) + 0.60\left(\frac{a}{e_0}\right) \qquad \text{(A10-A15)}$$

$$H_1 = 1 - 0.34\left(\frac{a}{e_0}\right) - 0.11\left(\frac{a}{c}\right)\left(\frac{a}{e_0}\right)$$

$$H_2 = 1 + G_1\left(\frac{a}{e_0}\right) + G_2\left(\frac{a}{e_0}\right)^2$$

$$G_1 = -1.22 - 0.12\left(\frac{a}{c}\right)$$

$$G_2 = 0.55 - 1.05\left(\frac{a}{c}\right)^{3/4} + 0.47\left(\frac{a}{c}\right)^{3/2}$$

The adjustment of the regressed creep function was indicated by the analysis of the sum of square error (SSE) and the root mean square error (RMSE). The SSE error is defined as:

$$SSE = \sum_{i=1}^{p}\left(\hat{y}_i - y_i\right)^2 \tag{A16}$$

in which, $\hat{y}_i$ is the value predicted by the regressed creep function and y_i is the value of experimental data from Atrushi [69]. The RMSE error is given as:

$$RMSE = \left[\frac{1}{p}\sum_{i=1}^{p}\left(\hat{y}_i - y_i\right)^2\right]^{1/2} \tag{A17}$$

Another verification in the adjustment of the creep function was done by the use of the coefficient of determination:

$$r^2 = \frac{S_t - SSE}{S_t} \tag{A18}$$

with:

$$S_t = \sum_{i=1}^{p}(y_i - \mu(y_i))^2 \tag{A19}$$

in which, S_t is the sum of the total residuals between the experimental data from the literature and their respective mean. The coefficient of determination closest to 1.0 it means a good fit of the creep function.

APPENDIX B

The CDFs predicted by the surrogate models of chapter 03 were statistically tested with the results obtained from the reference solution using the two-samples Kolmogorov-Smirnov (KS) statistical test. In this statistical test, the test variable KS is the maximum vertical difference between the CDFs being compared. The value of KS is given as:

$$KS = \max\left[F(\hat{y}) - F_{MC}(y)\right] \tag{B1}$$

in which $F(\hat{y})$ is the CDF constructed from the values predicted by the surrogate model and $F_{MC}(y)$ is the CDF generated with the values of the MC^{REF} solution. The hypothesis H_0 consists of $F(\hat{y}) = F_{MC}(y)$, being rejected if the two CDFs are not obtained from the same population and, therefore, are not statistically similar for a given level of significance, which for this work follows the same procedure adopted by Evangelista Jr. and Almeida [5]. The results of the KS test are shown in the following table.

Table B – Results of the KS statistical hypothesis test for a significance level of 5% of the five cases studied

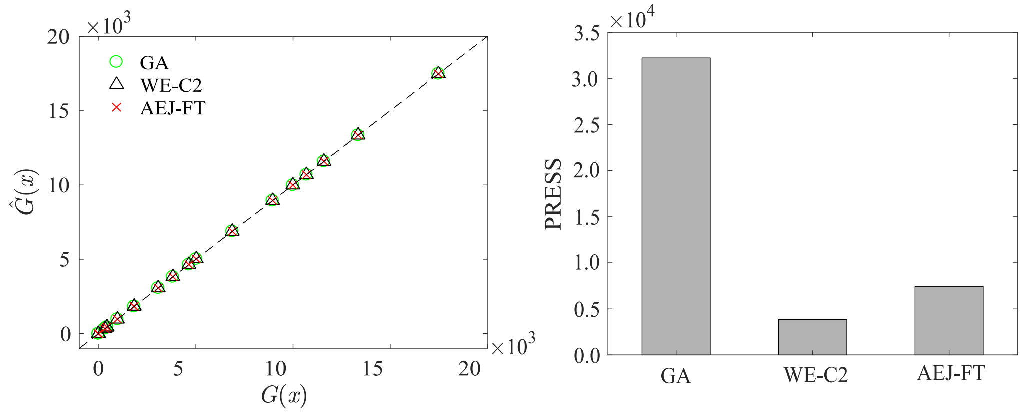
Case	p	Method	Result	KS
t_c = 1 day / t = 11 days	79	RBF (WE-C2)	H_0	0.08
	180	RBF (WE-C2)	H_0	0.01
t_c = 1 day / t = 365 days	79	RBF (WE-C2)	H_0	0.08
	180	RBF (WE-C2)	H_0	0.02
t_c = 7 days / t = 17 days	79	RBF (WE-C2)	H_0	0.06
	180	RBF (WE-C2)	H_0	0.02
t_c = 7 days / t = 365 days	79	RBF (WE-C2)	H_0	0.07
	180	RBF (WE-C2)	H_0	0.03
t_c = 7 days / t = 3650 days	79	RBF (WE-C2)	H_0	0.09
	180	RBF (WE-C2)	H_0	0.04

H_0: the surrogate predictions and the MC reference solution are extracted from the same population
H_1: the surrogate predictions and the MC reference solution are extracted from different populations

In Table B, it is noted that for all cases with different elapsed times, the CDFs constructed from the predicted results both by the surrogate models with training sets p = 79 and for p = 180 are part of the same population as the CDFs constructed with the results of MC^{REF}. This behavior was expected, since some of the results obtained are part of the paper of Evangelista Jr. and Almeida [5].

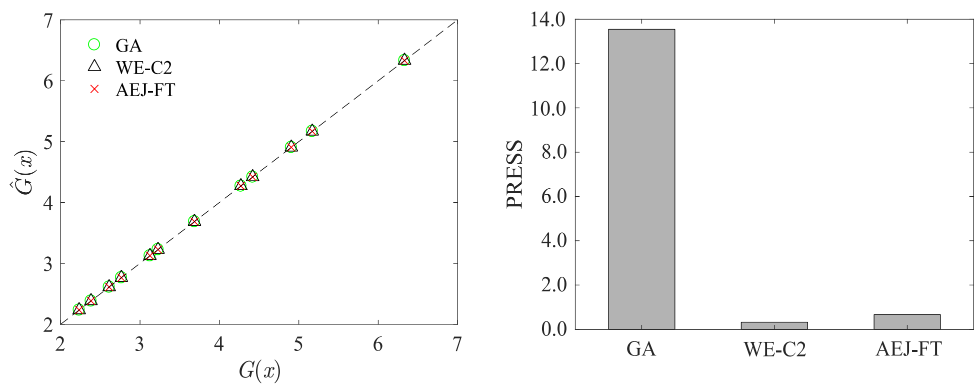
APPENDIX C

The cross-validation (LOOCV) results and the calculated PRESS values are presented in this appendix to the Examples 01 to 04 of Chapter 04. Figures C1 to C4 shows that in PRESS, each prediction of the unseen data is performed using the surrogate trained with the remaining *p - 1* data points. In general, the numerical examples considered three different types of radial basis function surrogate models that are: Gaussian (GA), Wendland-C2 (WE-C2) and the new Almeida-Evangelista Jr. (AEJ-FT).



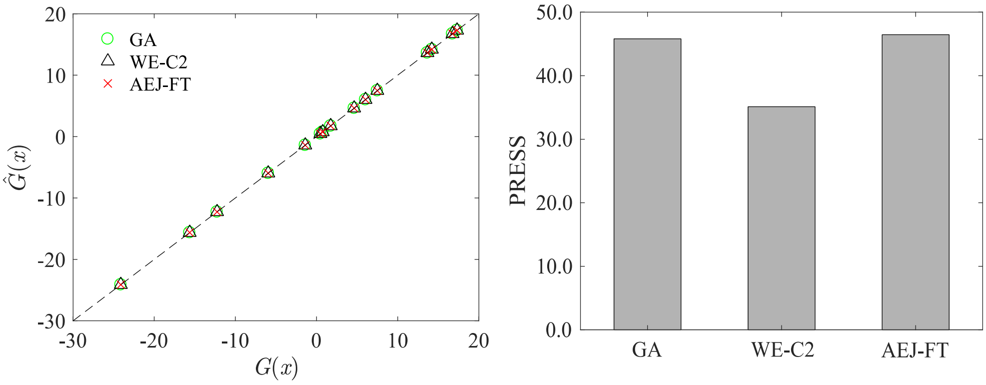
a) LOOCV, p = 19 b) PRESS

Figure C1 - Leave-one-out cross-validation (LOOCV) results for Example 01



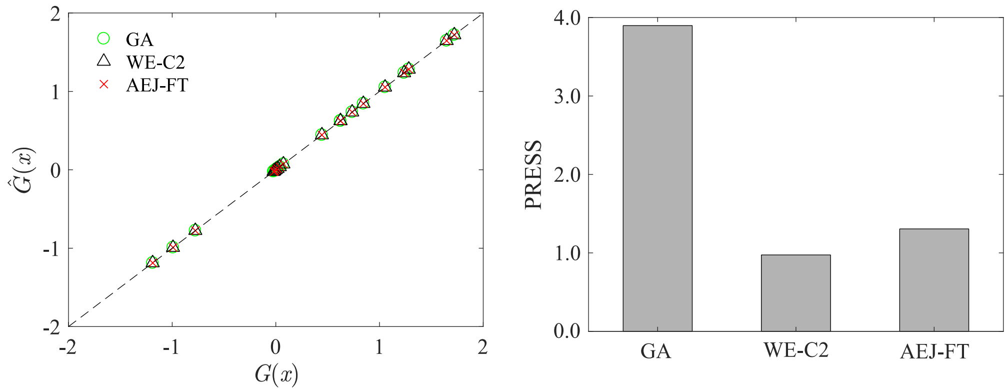
a) LOOCV, p = 12 b) PRESS

Figure C2 - Leave-one-out cross-validation (LOOCV) results for Example 02



a) LOOCV, p = 15 b) PRESS

Figure C3 - Leave-one-out cross-validation (LOOCV) results for Example 03

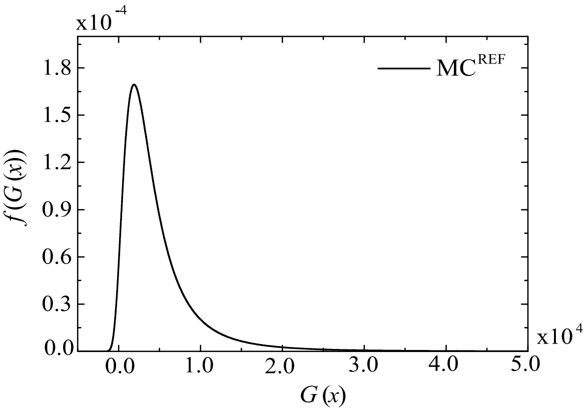


a) LOOCV, p = 19 b) PRESS

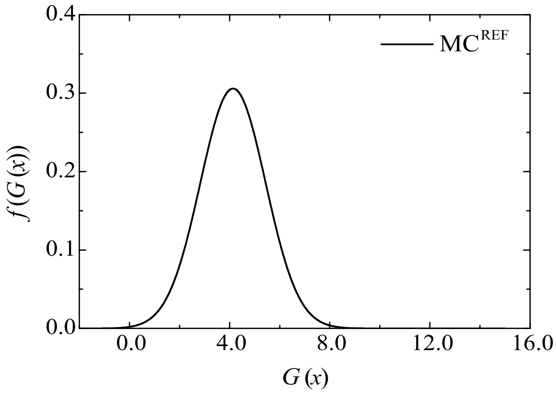
Figure C4 - Leave-one-out cross-validation (LOOCV) results for Example 04

The plots show the better performance and higher generalization capabilities of the RBFs models WE-C2 and AEJ-FT rather than the spreader predictions of GA for the most of examples as can be seen in the PRESS values. The plots of LOOCV shows small errors in the prediction of all the examples. On the other hand, the calculated PRESS shows high magnitudes, principally to the Example 01. This behavior demonstrates the difficult to approximate the selected benchmark problems in which considered fewer training samples than the surrogate models found in the literature.

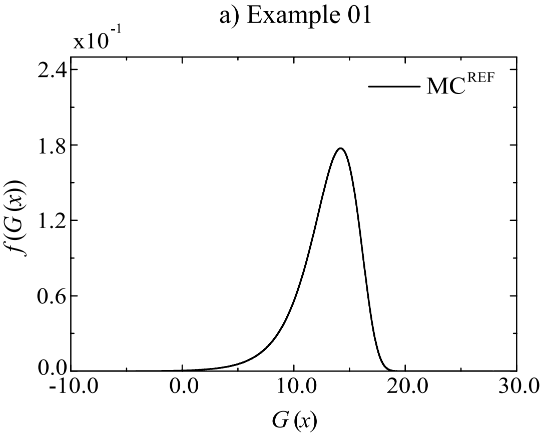
The probability density functions of the MC^{REF} solution of Examples 01 to 04 considered in Chapter 04 are shown in Figure C5.



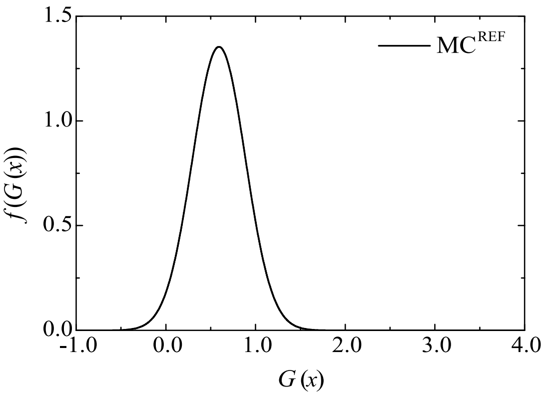
a) Example 01



b) Example 02



c) Example 03



d) Example 04

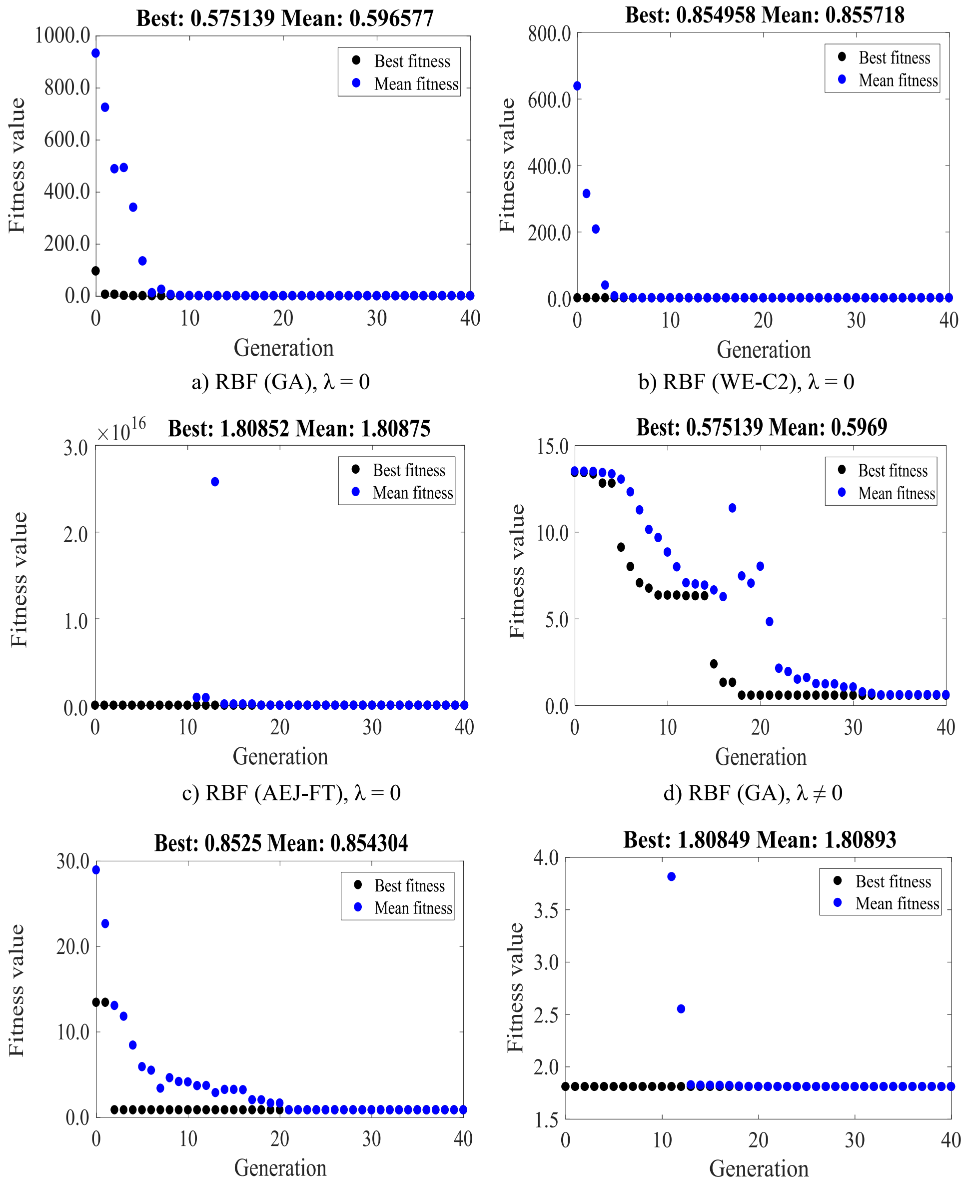
Figure C5 – Probability density functions for MC^{REF} from Examples 01 to 04.

APPENDIX D

The results of the optimization process of the genetic algorithm for the PRESS and the GCV for the Examples 01 to 03 of Chapter 05 are shown in the appendix D. In the optimization of the genetic algorithm, numerical experiments were done to determine the population size. Each numerical example of this paper considered a different population in order to guarantee a lower computational cost. The results of Example 01 are shown below.

a) Example 01:

The graphs of the optimization of the genetic algorithm to minimize the PRESS of Example 01 are shown in Figure D1.



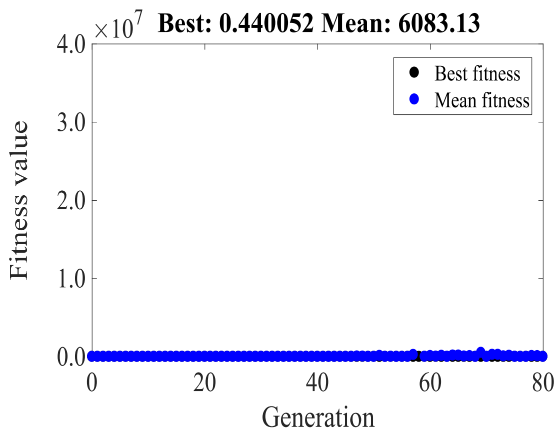
a) RBF (GA), $\lambda = 0$ b) RBF (WE-C2), $\lambda = 0$

c) RBF (AEJ-FT), $\lambda = 0$ d) RBF (GA), $\lambda \neq 0$

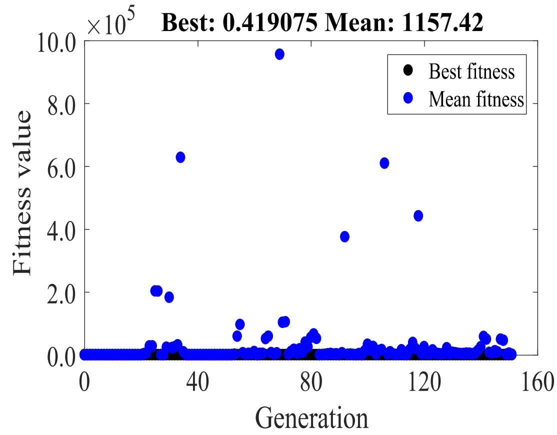
e) RBF (WE-C2), $\lambda \neq 0$ f) RBF (AEJ-FT), $\lambda \neq 0$

Figure D1 – PRESS optimization for case 01 of Example 01

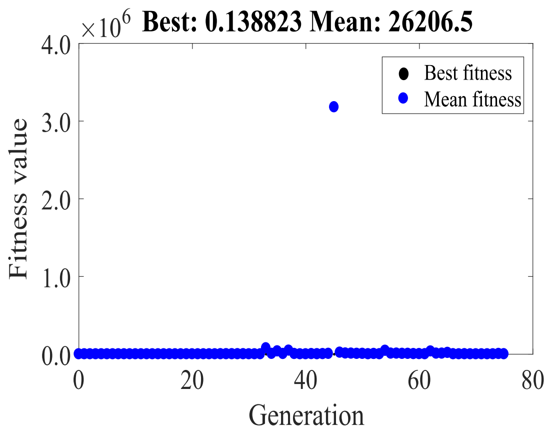
Figure D1 shows a fast convergence to the surrogate model with RBF (GA) before reaching the tenth generation. The surrogate models with the RBF (WE-C2) and RBF (AEJ-FT) showed greater difficult in determining the minimum value of the PRESS. This difficulty in optimization occurs mainly for models with regularized regression. It is also important to note that among all surrogate models, the RBF (AEJ-FT) presented a greater difficulty in optimization because it presents a greater number of shape parameters. The results of the GCV optimization for case 01 are shown in Figure D2.



a) RBF (GA), $\lambda \neq 0$



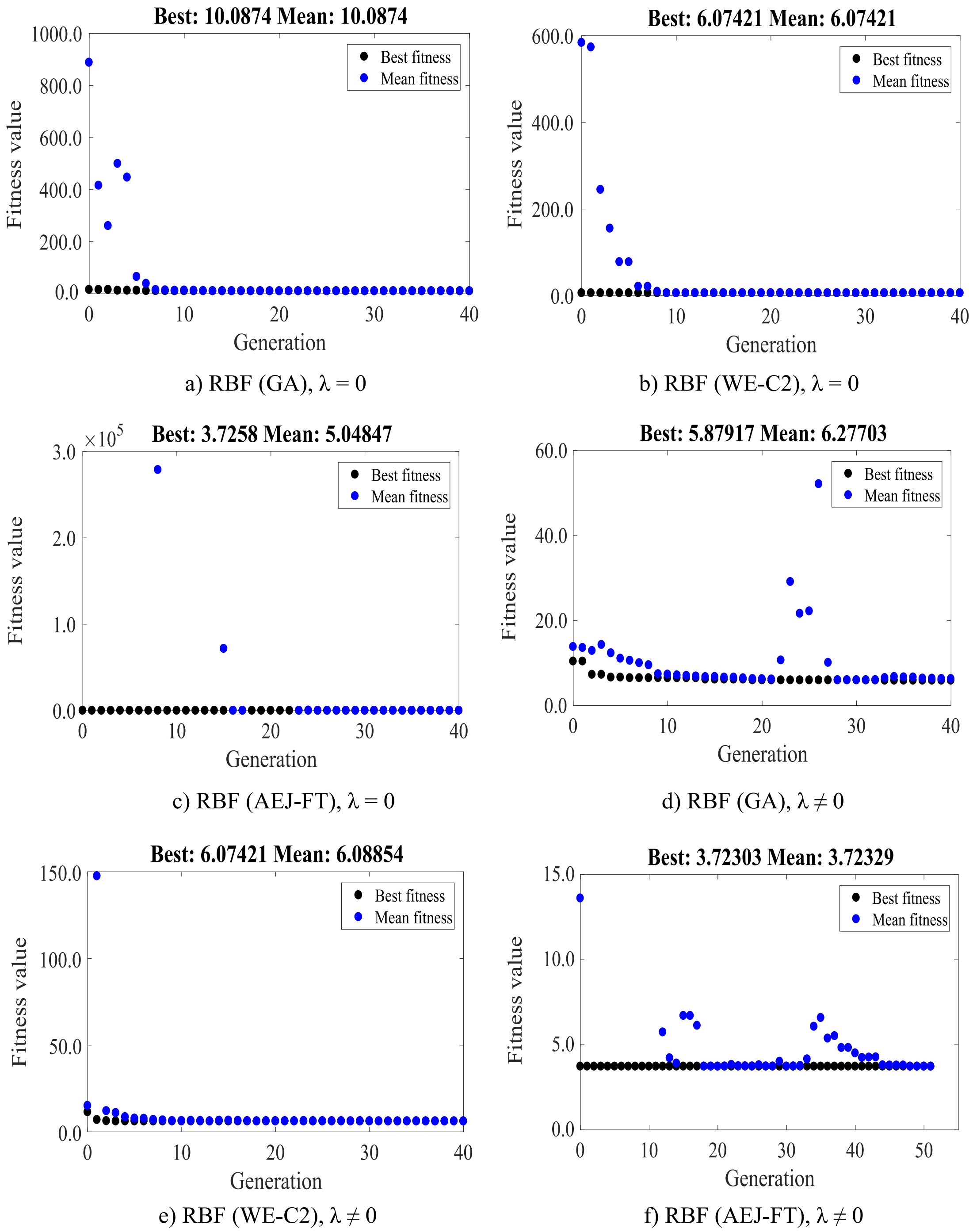
b) RBF (WE-C2), $\lambda \neq 0$



c) RBF (AEJ-FT), $\lambda \neq 0$

Figure D2 – GCV optimization for case 01 of Example 01

The results of the optimization of the GCV show a different behavior than the one presented for the PRESS. The results in Figure D2b,c show that the GCV algorithm ends the optimization process before reaching the maximum number of iterations. This behavior is due to the greater capability of the GCV. Thus, the graphs obtained from the GCV are not well defined in the optimization process. To ensure that overfitting did not occur, an additional optimization process was also done for the GCV implementations. The routines of GCV has an additional scan for some situations with an initial population obtained from an initial scan. In this second scan, the population size and the maximum number of generations are increased. However, the result obtained by these additional scans presented similar parameter values to those obtained initially, and therefore it is the graphs of this second scan that are considered instead of the results of the initial population. The PRESS results from case 02 for Example 01 are shown in Figure D3.



a) RBF (GA), λ = 0

b) RBF (WE-C2), λ = 0

c) RBF (AEJ-FT), λ = 0

d) RBF (GA), λ ≠ 0

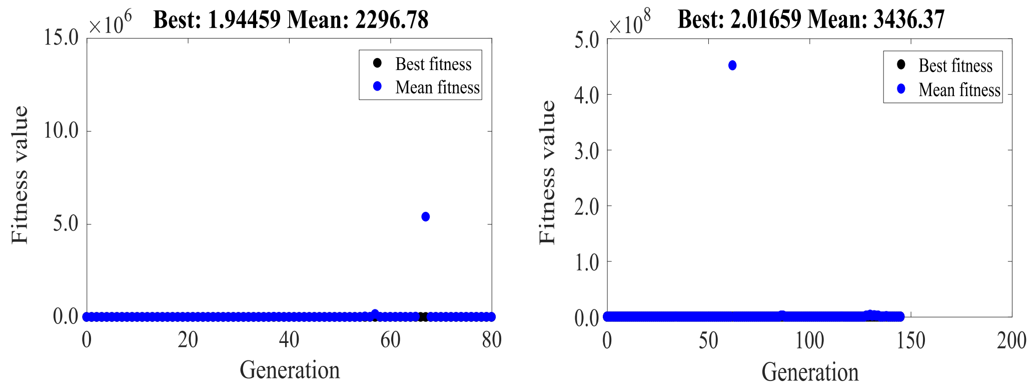
e) RBF (WE-C2), λ ≠ 0

f) RBF (AEJ-FT), λ ≠ 0

Figure D3 – PRESS optimization for case 02 of Example 01

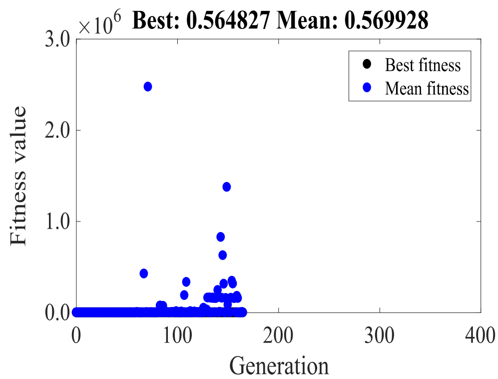
The results show a greater difficulty for the convergence of the regularized surrogate models with RBF (WE-C2) and RBF (AEJ-FT). In the case of the RBF (AEJ-FT) a population of size 60 was used in the genetic algorithm and the result of the minimization of the PRESS presented a termination of the processing before reaching the maximum generation limit considered. This was due to an adaptation in the PRESS routine to terminate the process in a shorter computational time.

On the other hand, this modification in the implementation of the genetic algorithm was not performed in the other surrogate models, as only the RBF (AEJ-FT) presented a greater difficulty in optimizing the PRESS. The GCV optimization for case 02 of Example 01 is shown in Figure D4.



a) RBF (GA), $\lambda \neq 0$

b) RBF (WE-C2), $\lambda \neq 0$



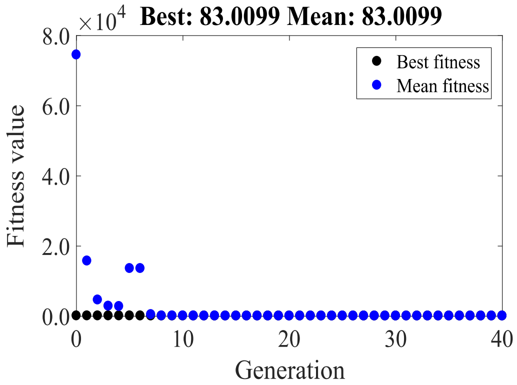
c) RBF (AEJ-FT), $\lambda \neq 0$

Figure D4 – GCV optimization for case 02 of Example 01

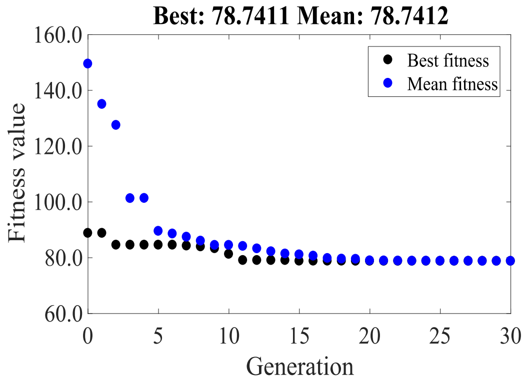
In Figure D4, it is noted that the generation number used to minimize the GCV is higher than that of the PRESS. The population considered for the RBFs has a size of 50 to ensure that the ideal parameters of the surrogate model are obtained. The RBF (WE-C2) and RBF (AEJ-FT) also presented modifications in the computational implementations to guarantee a faster and more efficient optimization and consequently the algorithm finished the optimization process before reaching the maximum number of pre-defined generations. In addition, the RBF (AEJ-FT) showed a greater difficulty in the optimization, being, therefore, similar to the case 01.

b) Example 02:

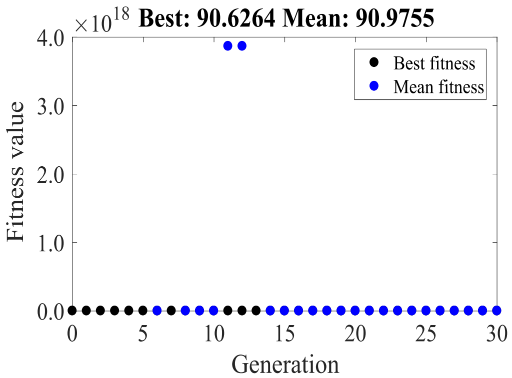
The graphs of the genetic algorithm to minimize the PRESS of Example 02 are shown in Figure D5.



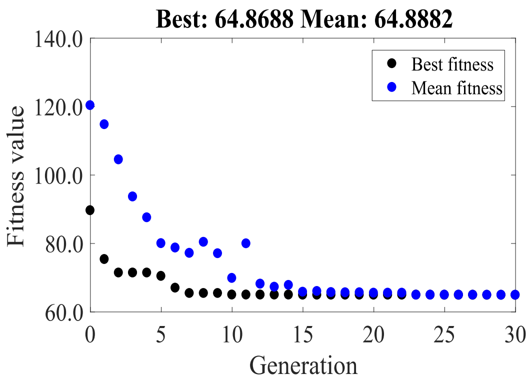
a) RBF (GA), $\lambda = 0$



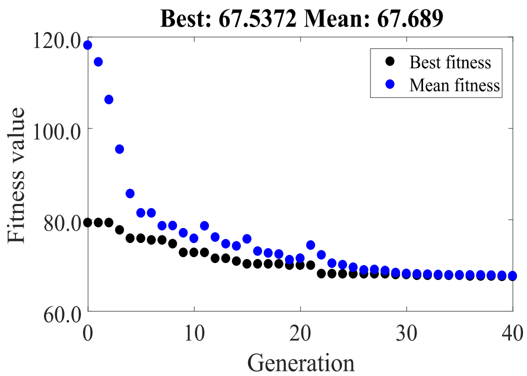
b) RBF (WE-C2), $\lambda = 0$



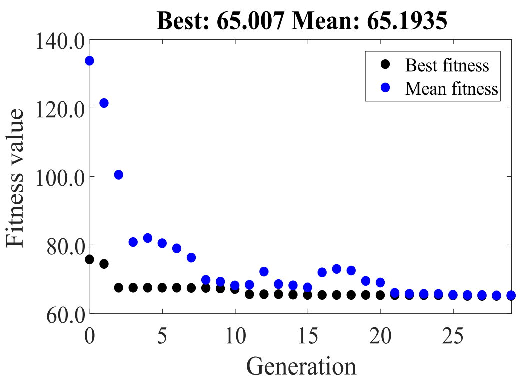
c) RBF (AEJ-FT), $\lambda = 0$



d) RBF (GA), $\lambda \neq 0$



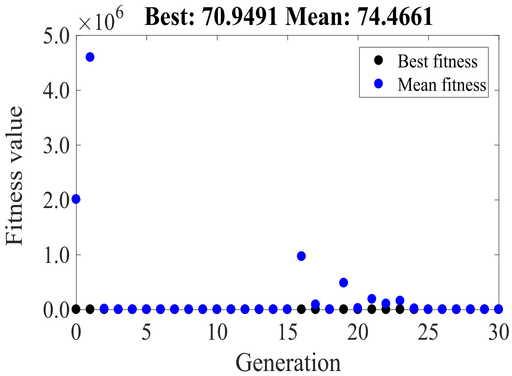
e) RBF (WE-C2), $\lambda \neq 0$



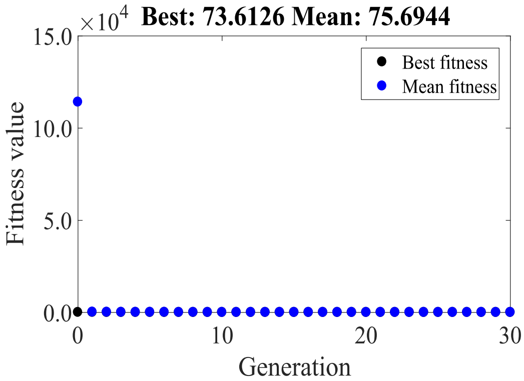
f) RBF (AEJ-FT), $\lambda \neq 0$

Figure D5 – PRESS optimization of Example 02

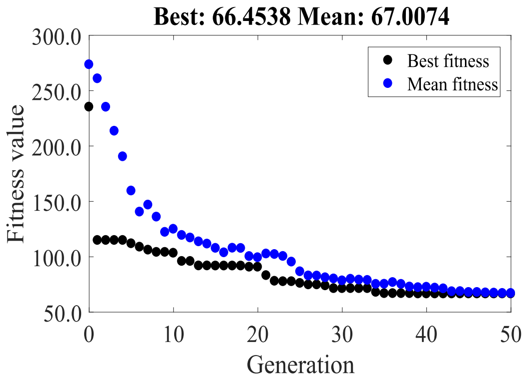
Figure D5 shows a fast convergence to the surrogate model with RBF (GA) and non-regularized PRESS. The convergence of the regularized PRESS of the RBF (GA) and RBF (WE-C2) tends to occur with a similar number of generations, that is, approximately in the fifteenth generation. The optimization of the RBF (AEJ-FT) occurred only for generations greater than 20. This RBF, therefore, was the one that presented the greatest difficulty for convergence. The GCV convergence graph is shown in Figure D6.



a) RBF (GA), $\lambda \neq 0$



b) RBF (WE-C2), $\lambda \neq 0$



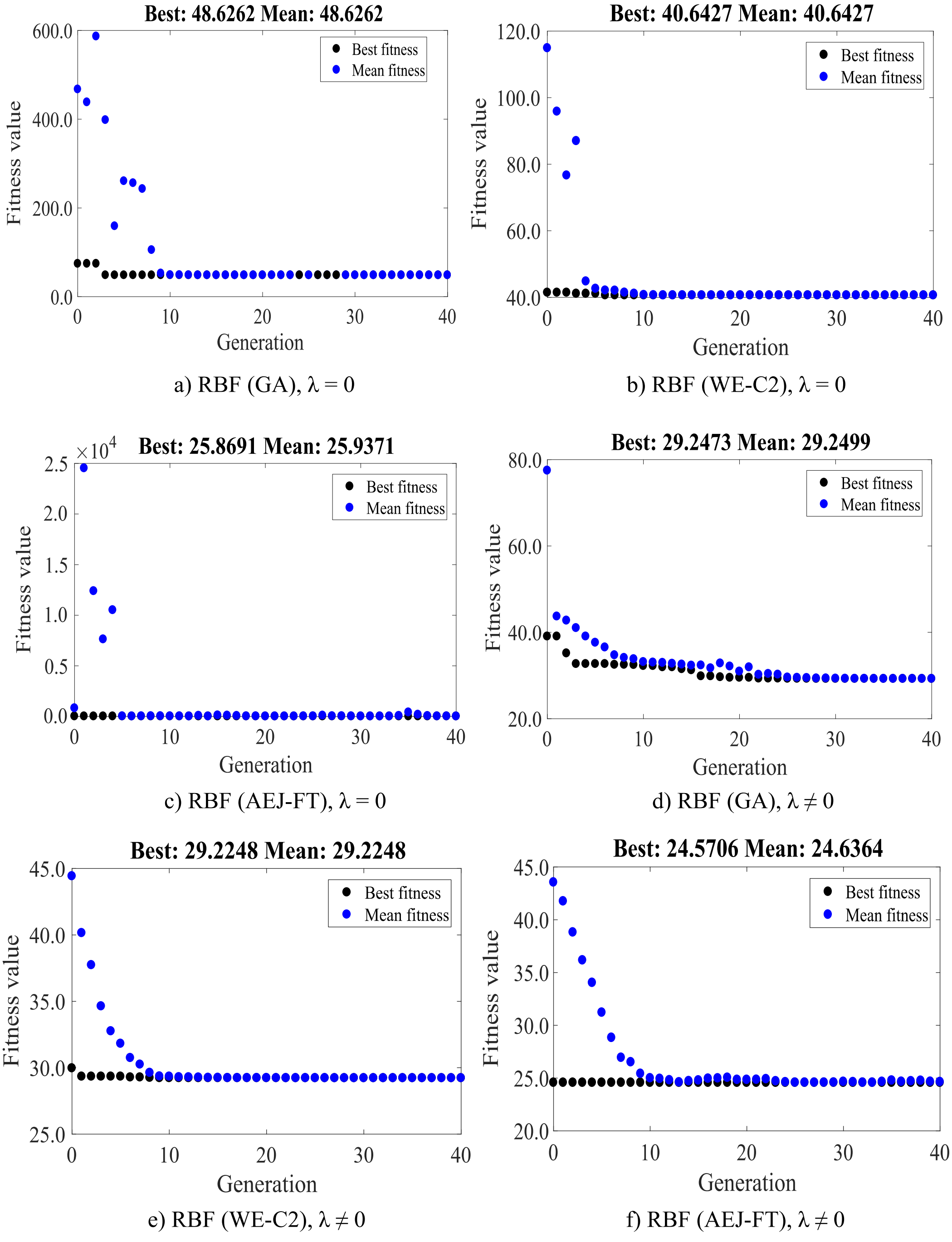
c) RBF (AEJ-FT), $\lambda \neq 0$

Figure D6 – GCV optimization of Example 02

The parameters optimized by the GCV of the RBF (GA) and RBF (AEJ-FT) follow the same behavior as the PRESS, that is a convergence of values is clear in the graphs. It is important to note that the GCV will not always present a clear convergence graph. In the case of RBF (WE-C2), the convergence is reached quickly with the GCV value equal to 73.61.

c) Example 03:

The graphs of the optimization of the genetic algorithm to minimize the PRESS of Example 03 are shown in Figure D7.



a) RBF (GA), λ = 0

b) RBF (WE-C2), λ = 0

c) RBF (AEJ-FT), λ = 0

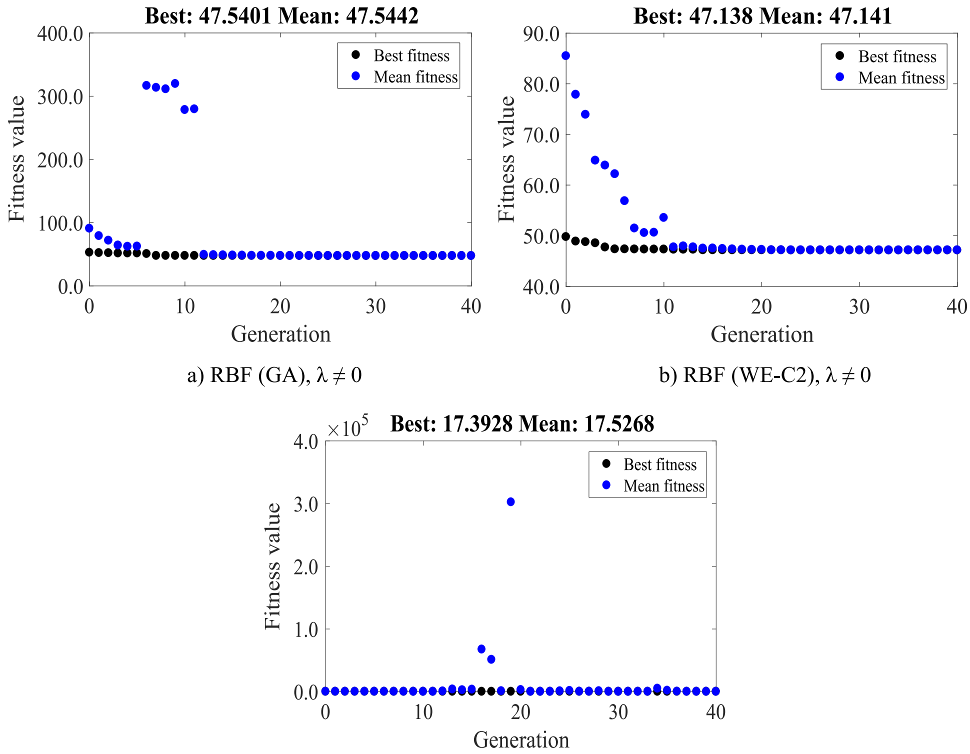
d) RBF (GA), λ ≠ 0

e) RBF (WE-C2), λ ≠ 0

f) RBF (AEJ-FT), λ ≠ 0

Figure D7 – PRESS optimization of Example 03

The PRESS convergence graphs show that the number of generations needed to reach the minimum PRESS value is only 10 generations for most of the surrogate models. The population considered for the surrogate models was 30. It is also noted that there is no need for a large number of population or generation to obtain the optimized values of the PRESS. The only surrogate models that showed a disturbance in the optimization were the RBF (WE-C2) and RBF (AEJ-FT) with regularization. This disturbance in the optimization process occurs in the generation 15 to 20. The convergence graphs of the GCV optimization process are shown in Figure D8.



a) RBF (GA), $\lambda \neq 0$

b) RBF (WE-C2), $\lambda \neq 0$

c) RBF (AEJ-FT), $\lambda \neq 0$

Figure D8 – GCV optimization of Example 03

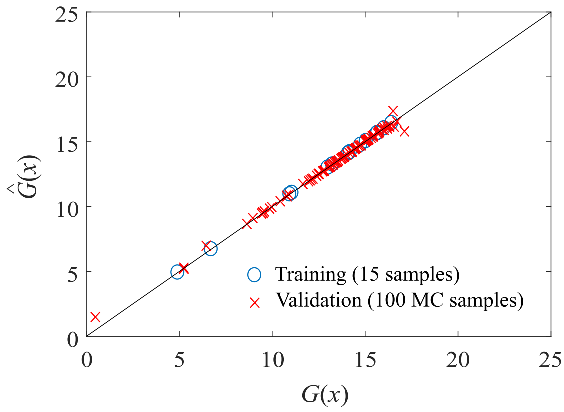
Figure D8 shows a convergence of the GCV with a generation of approximately 11 for the RBF (GA) and a generation of 15 for the RBF (WE-C2). The RBF (AEJ-FT) presents a greater difficulty for optimization being the result of the minimum GCV reached in generation 35.

APPENDIX E

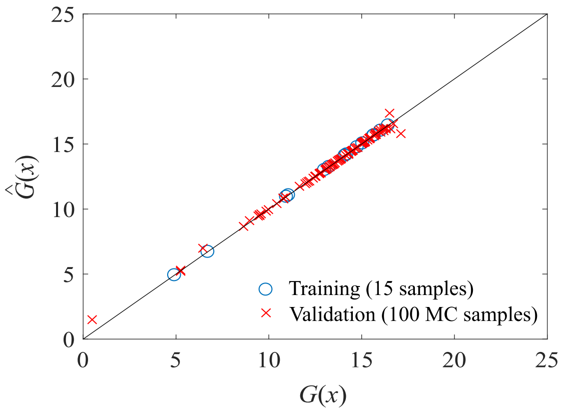
The cross-validation results (LOOCV) for Example 01 of chapter 05 are shown in appendix E. In this example, 15 training samples and only 100 samples of the 5 E+05 MC samples were considered for the validation. The results correspond to cases 01 and 02, being for case 01 with the set of the training samples without the additional noise and for case 02 with the addition of a noise that follows a normal distribution with zero mean and standard deviation of 0.10.

a) Case 01:

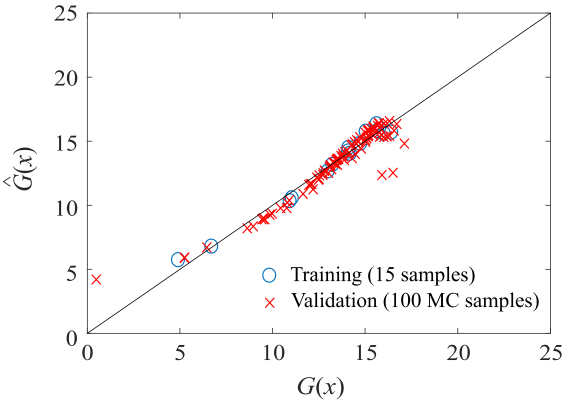
The case 01 corresponds to the training set without an initial perturbation, and these samples are used for the construction of the surrogate models with RBF (GA), RBF (WE-C2) and RBF (AEJ-FT). The cross-validation results for the RBF (GA) constructed with parameters obtained from the minimization of the PRESS and GCV are shown in Figure E1.



a) PRESS, $\lambda = 0$



b) PRESS, $\lambda \neq 0$

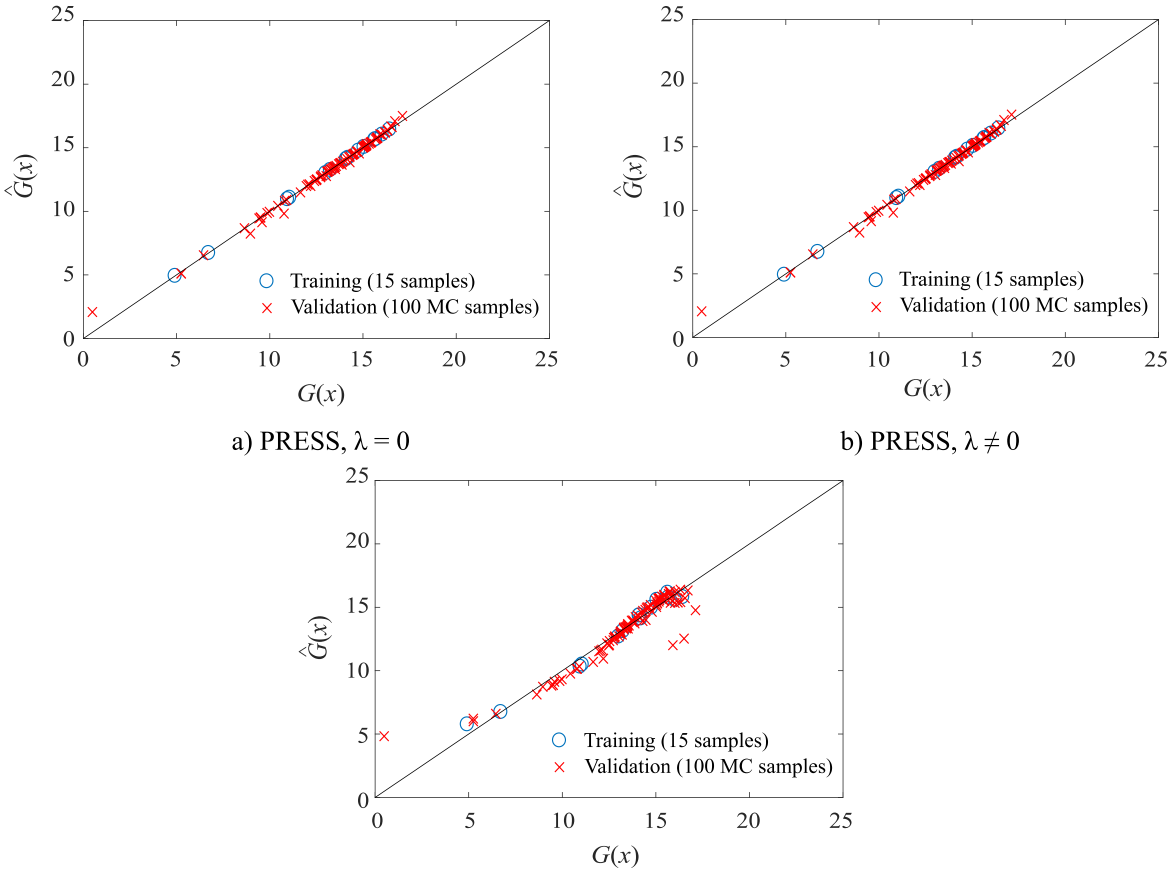


c) GCV, $\lambda \neq 0$

Figure E1 – Results of cross-validation (LOOCV) for RBFs (GA) of case 01 to Example 01

The LOOCV results for the RBF (GA) show that all cases presented a good approximation. This behavior was expected because the results of the error ε and the KS test have already proven this behavior. It is also important to emphasize that Figure E1a and E1b are quite similar due to the

PRESS results having similar magnitudes. For the GCV model, there is a small dispersion of training and validation data. The LOOCV results of the RBF (WE-C2) are shown in Figure E2.



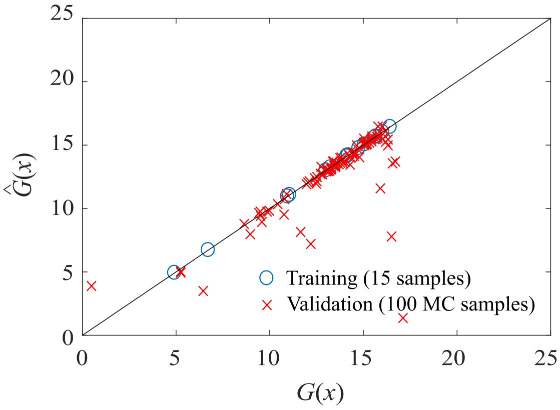
a) PRESS, λ = 0

b) PRESS, λ ≠ 0

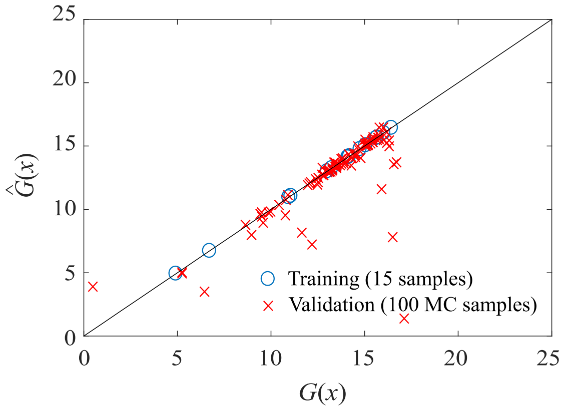
c) GCV, λ ≠ 0

Figure E2 – Results of cross-validation (LOOCV) for RBFs (WE-C2) of case 01 to Example 01

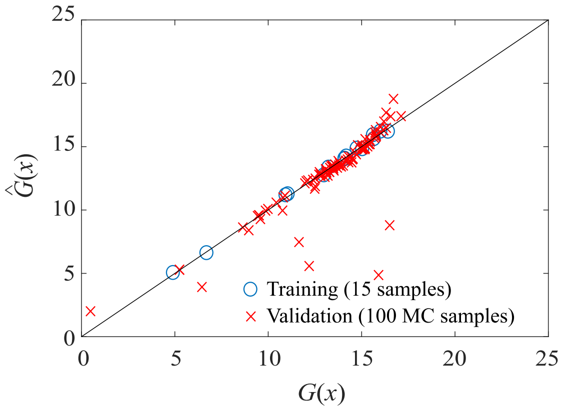
It is observed for the LOOCV of the RBF(WE-C2) a behavior similar to that of the RBF (GA), that is, all surrogate models presented a good approximation with data dispersion in the approximation practically insignificant. The RBF (WE-C2) with parameters obtained from the minimization of the GCV consists of the situation with the presence of greater dispersion in the approximation, but with insignificant magnitudes, as verified in the graph and in the results presented above. The results for the RBFs (AEJ-FT) are shown in Figure E3. It is also important to note that for the LOOCV of Example 01, only 100 samples of MC were considered, and 5 E+05 were used in the uncertainty quantification.



a) PRESS, $\lambda = 0$



b) PRESS, $\lambda \neq 0$



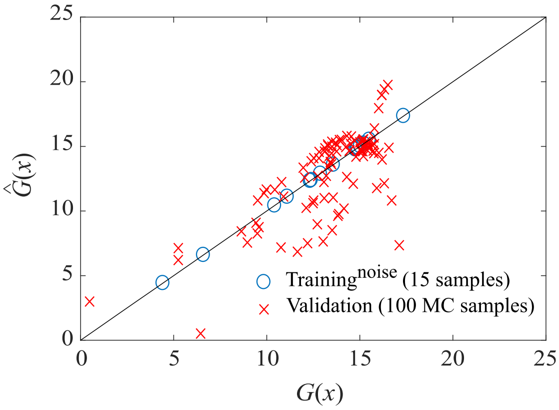
c) GCV, $\lambda \neq 0$

Figure E3 – Results of cross-validation (LOOCV) for RBFs (AEJ-FT) of case 01 to Example 01

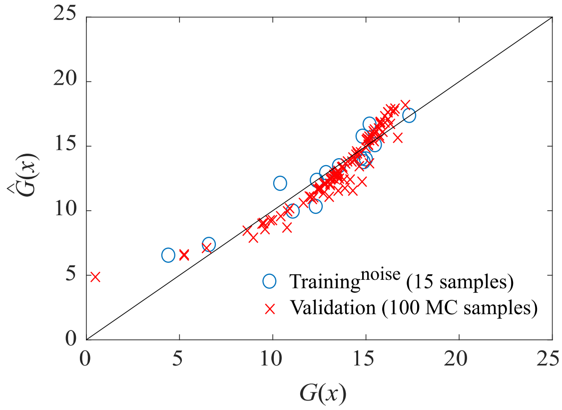
The cross-validation results for the RBFs (AEJ-FT) consist of the models that presented the worst approximation for case 01 with the presence of a dispersion of the validation results. For these RBFs, it is noted that the training samples showed a good approximation, however, for the MC validation data, it was observed that the surrogate models were not able to predict the MC values with the same precision as those of the training samples. However, this behavior does not invalidate the models, as the results of the KS test and the normalized error test presented previously, demonstrated the efficiency of these models.

b) Case 02:

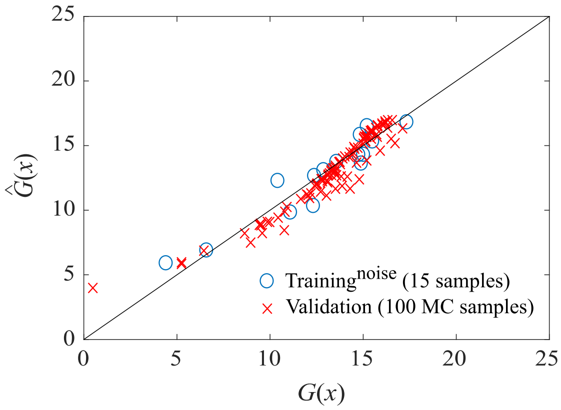
The case 02 corresponds to the set of training samples that present in the dependent variable (output) an additional noise for the construction of surrogate models of the type RBF (GA), RBF (WE-C2) and RBF (AEJ-FT). This noise is added in order to reproduce the behavior of real experimental data, as MC simulation data generally do not have this noise. The cross-validation results for the RBF (GA) constructed with parameters obtained from the minimization of PRESS and GCV are shown in Figure E4.



a) PRESS, λ = 0



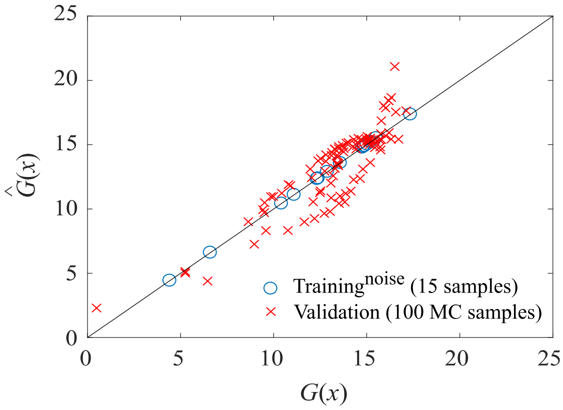
b) PRESS, λ ≠ 0



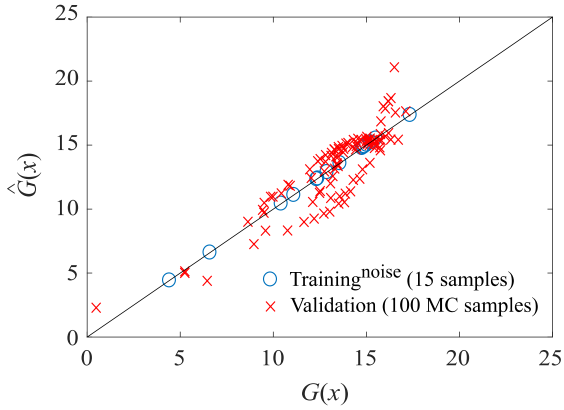
c) GCV, λ ≠ 0

Figure E4 – Results of cross-validation (LOOCV) for RBFs (GA) of case 02 to Example 01

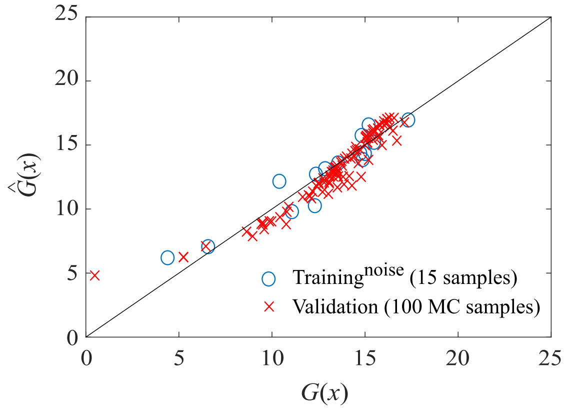
The addition of noise to the training samples caused a dispersion in the prediction of both the 15 training samples (Figure E4b,c) and the 100 MC validation samples (Figure E4a). The dispersion of results is greater for the RBF (GA) model built with parameters obtained from the minimization of PRESS without regularization. This dispersion is smaller for the PRESS model with regularization, being, therefore, observed that for the PRESS, the presence of the regularization parameter causes an improvement in the approximation. For the regularized GCV, a behavior similar to that of the regularized PRESS is observed. The results of the RBFs (WE-C2) are shown in Figure E5.



a) PRESS, λ = 0



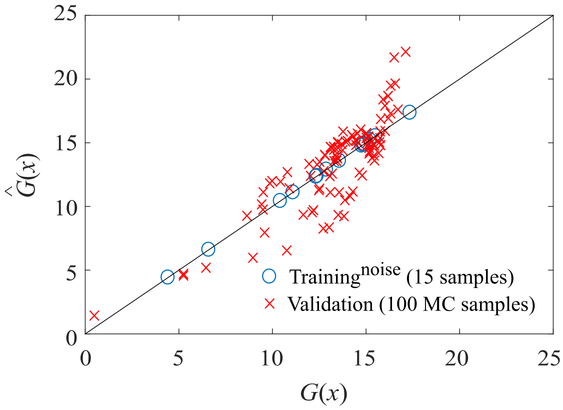
b) PRESS, λ ≠ 0



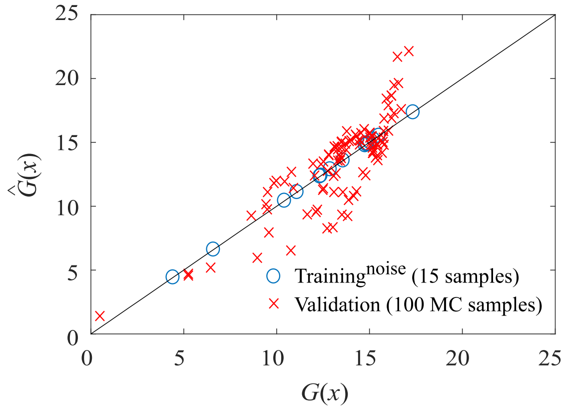
c) GCV, λ ≠ 0

Figure E5 – Results of cross-validation (LOOCV) for RBFs (WE-C2) of case 02 to Example 01

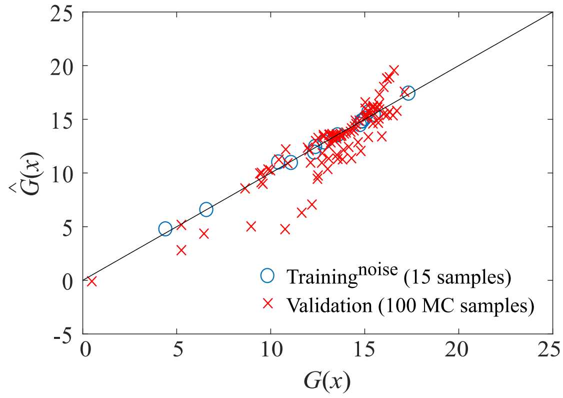
In Figure E5, it is noted that the model built with parameters obtained from the minimization of the GCV presents a better approximation than the models with PRESS. This occurs because the magnitude of the GCV is lower than that of the PRESS. Another important point to be observed in all models is the presence of a small dispersion of the validation samples. For the models with PRESS, there is a similarity between the model without regularization and the model with regularization, and this behavior is expected, since the parameters of the models are quite similar. The results for the RBF (AEJ-FT) models with parameters obtained from the PRESS and the GCV are shown in Figure E6.



a) PRESS, λ = 0



b) PRESS, λ ≠ 0



c) GCV, λ ≠ 0

Figure E6 – Results of cross-validation (LOOCV) for RBFs (AEJ-FT) of case 02 to Example 01

The surrogate models of the RBF (AEJ-FT) showed a high dispersion in both the PRESS and GCV models. This dispersion occurs only in the prediction of the validation data, but this behavior does not invalidate these surrogate models, because as can be seen in the KS test and in the comparison of the CDFs, the surrogate models with RBF (AEJ-FT) are able to predict well the maximum deflection of the beam of Example 01. However, better accuracy can be obtained with a larger set of training samples. Furthermore, it is important to emphasize that the training samples present for case 02 an additional noise that caused the respective dispersion of the validation data.

APPENDIX F

The cross-validation results (LOOCV) for Example 02 of chapter 05 are showed in the appendix F. In this example, 1030 experimental data of concrete compressive strength were considered for the LOOCV, with 55 training samples and 975 validation samples. The experimental data already show a natural noise in the 1030 data.

The cross-validation is performed for the 55 training samples and their respective values predicted by the surrogate model, and also for the 975 samples used in the validation of the surrogate models of the RBF (GA), RBF (WE-C2) and RBF (AEJ-FT). The cross-validation results for the RBFs (GA) constructed with parameters obtained from the minimization of PRESS (with and without regularization) and the GCV (with regularization) are shown in Figure F1.

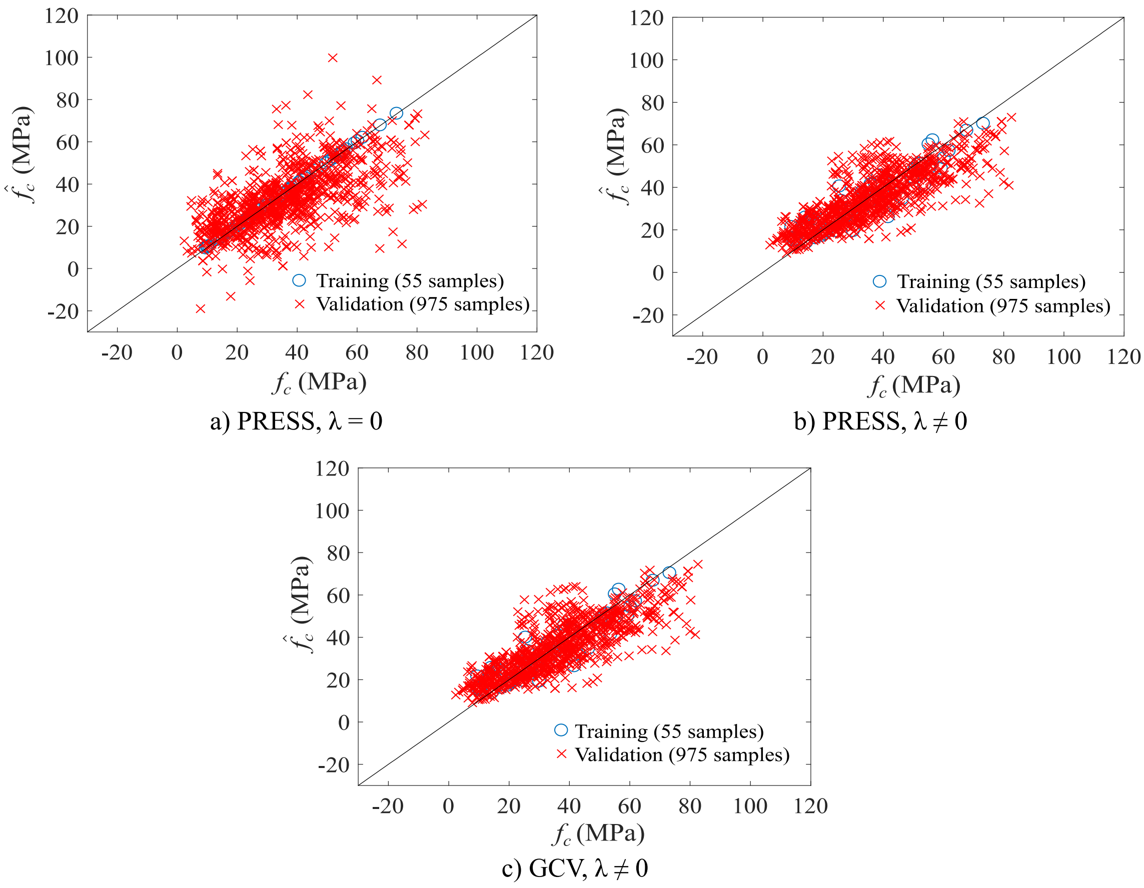
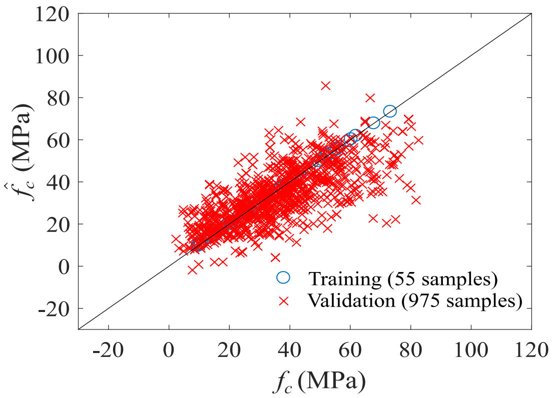
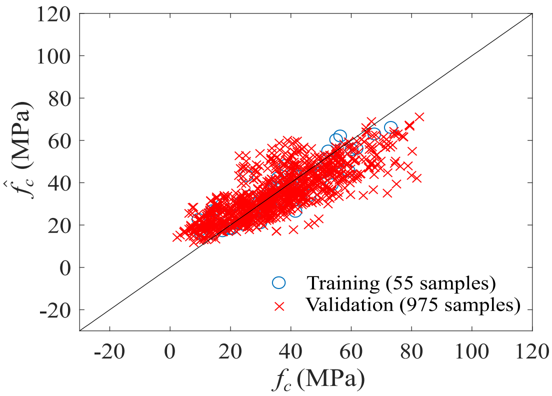


Figure F1 – Results of cross-validation (LOOCV) for RBFs (GA) to Example 02

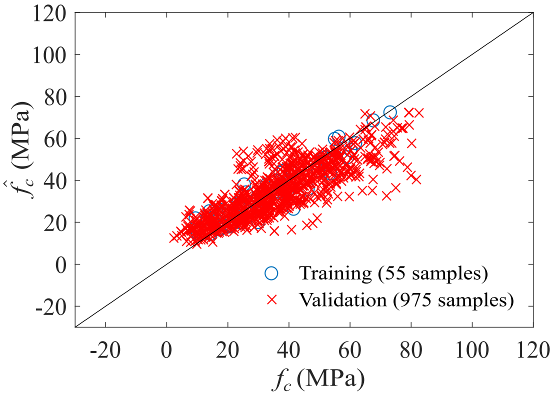
The results indicate a large dispersion of the validation samples, demonstrating difficulty of predicting the compressive strength, and this behavior is expected due to the RMSE values being a little high. For models with regularization, a better accuracy was noted in relation to PRESS without regularization, indicating the necessity of regularization for a better construction of the surrogate models. The results of the RBF (WE-C2) for Example 02 are shown in Figure F2.



a) PRESS, λ = 0



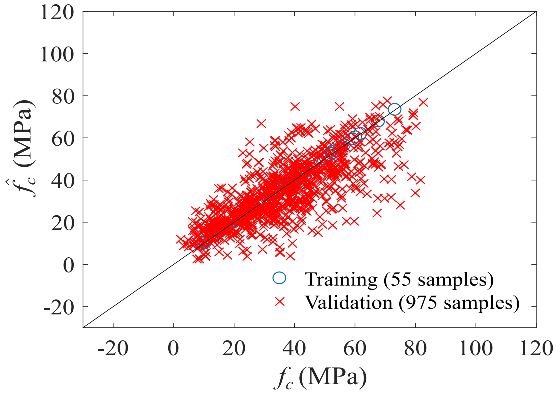
b) PRESS, λ ≠ 0



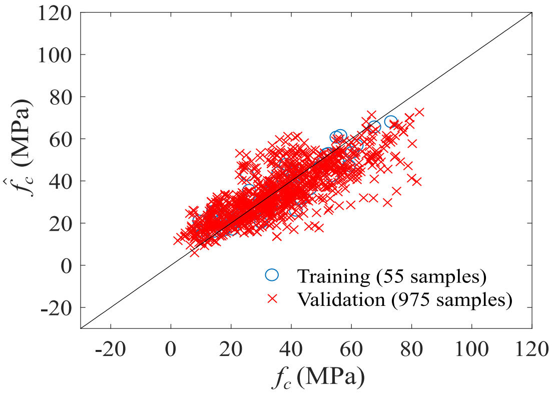
c) GCV, λ ≠ 0

Figure F2 – Results of cross-validation (LOOCV) for RBFs (WE-C2) to Example 02

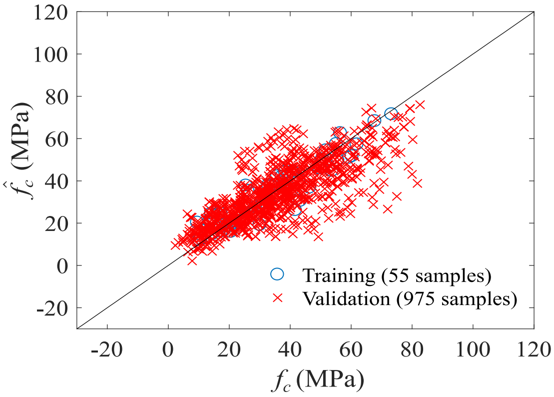
The surrogate model of the RBF (WE-C2) with PRESS and regularization shows a smaller dispersion and consequently a better approximation in relation to the reference model when compared to the surrogate model of the RBF (GA) with parameters obtained from the minimization of the non-regularized PRESS. However, in both models with PRESS, the dispersion of the prediction of the validation data is clear, demonstrating the difficulty of predicting the compressive strength of a set of unknown data. In the case of GCV, the behavior is quite similar to the regularized PRESS. Furthermore, it is shown again that the regularization allows better approximation of the prediction results of the surrogate models in cross-validation (LOOCV). The results of the RBFs (AEJ-FT) are shown in Figure F3.



a) PRESS, λ = 0



b) PRESS, λ ≠ 0



c) GCV, λ ≠ 0

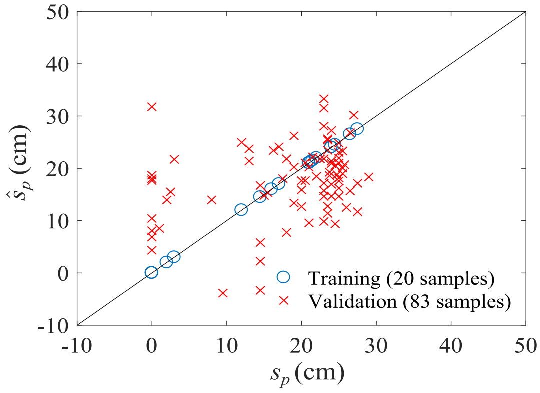
Figure F3 – Results of cross-validation (LOOCV) for RBFs (AEJ-FT) to Example 02

The LOOCV results for the RBF (AEJ-FT) show a behavior similar to those presented for the RBF (GA) in which the presence of the regularization parameter caused an improvement in the prediction of the surrogate models. For the model with GCV, there is a greater dispersion of the prediction of the validation data when compared to the PRESS with regularization.

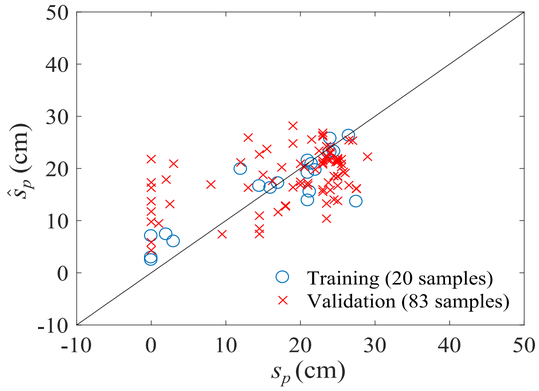
APPENDIX G

The cross-validation (LOOCV) results for Example 03 of Chapter 05 are presented in appendix G. In this example, 103 experimental data from the slump test were considered for the LOOCV, with 20 training samples and 83 validation samples. The experimental data already show a natural noise from the 103 samples.

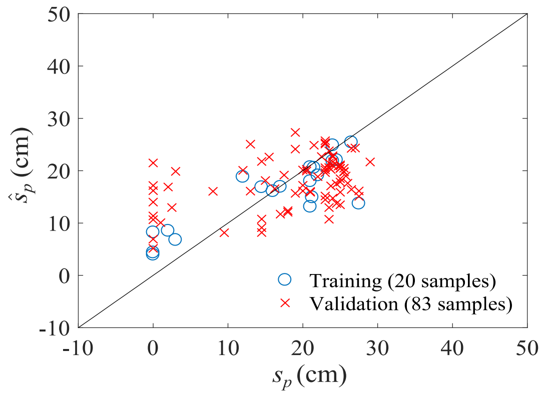
The cross-validation is performed for the training samples and their respective predicted values by the surrogate model, and also for the 83 samples used in the validation of the surrogate models of type RBF (GA), RBF (WE-C2) and RBF (AEJ-FT). The cross-validation results for the RBFs (GA) built with parameters obtained from the minimization of PRESS (with and without regularization) and GCV (with regularization) are shown in Figure G1.



a) PRESS, λ = 0



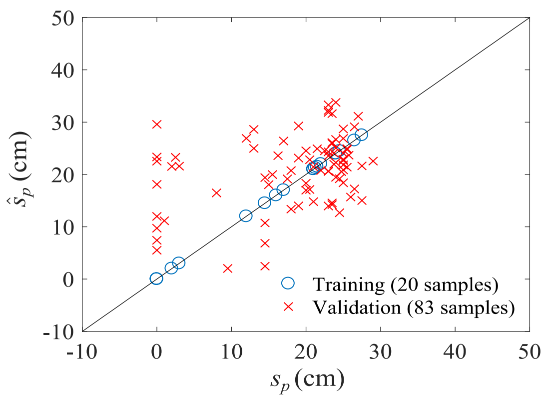
b) PRESS, λ ≠ 0



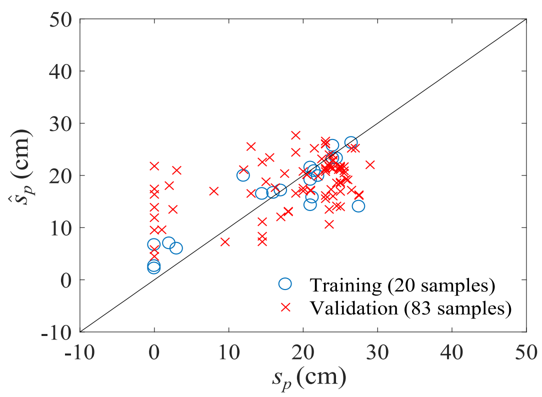
c) GCV, λ ≠ 0

Figure G1 – Results of cross-validation (LOOCV) for RBFs (GA) to Example 03

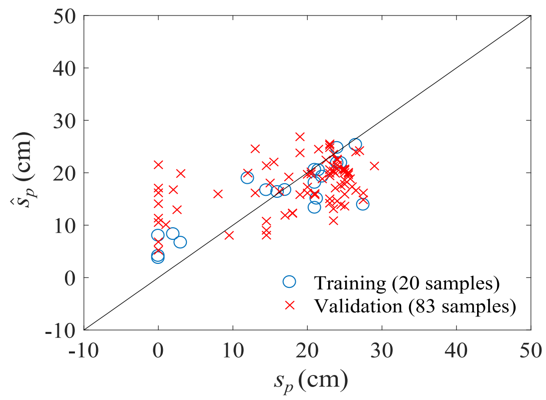
The LOOCV results for the RBF (GA) show that the non-regularized PRESS presents a good approximation of the training set, but in the validation phase, points further away from the 45° straight line (continuous line) are noted. This behavior was expected, since the RMSE and MSE in the validation stage present significant magnitudes and consequently a greater difficulty in predicting unseen data from the already trained model. However, as previously demonstrated, surrogate models with RBF have a good generalization capacity when compared to models in the literature. In models with regularization, a disturbance in the prediction of the samples is shown during training. This occurred due to the introduction of the regularization parameter that aimed to increase the accuracy of the models and reduce generalization errors. The LOOCV results of the RBF (WE-C2) models are shown in Figure G2.



a) PRESS, λ = 0



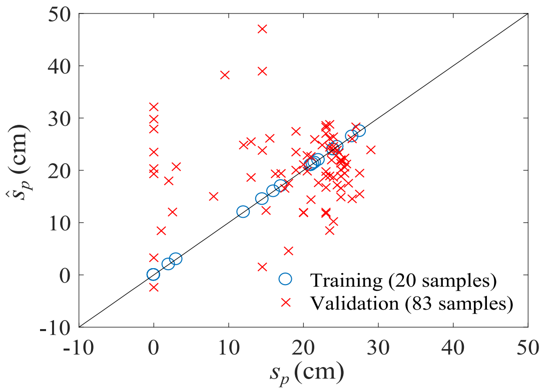
b) PRESS, λ ≠ 0



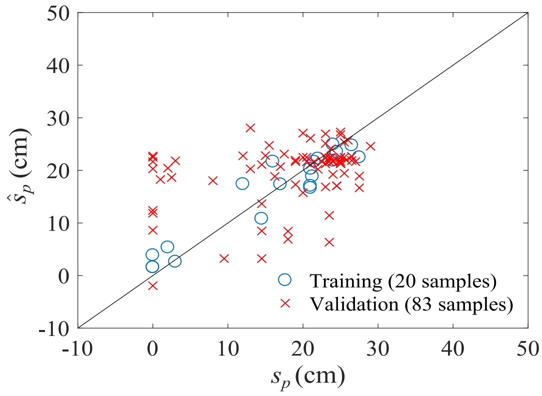
c) GCV, λ ≠ 0

Figure G2 – Results of cross-validation (LOOCV) for RBFs (WE-C2) to Example 03

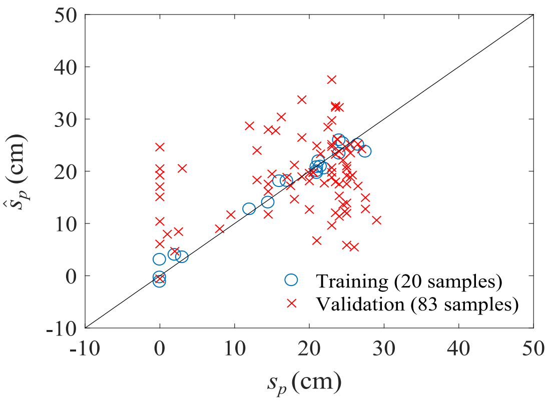
The LOOCV results for the RBF (WE-C2) show a behavior similar to those presented for the RBF (GA) in which the presence of the regularization parameter caused an improvement in the prediction of the surrogate models in which the points are closer to the straight line. The result of the GCV indicates a behavior similar to the PRESS with regularization. The LOOCV results of the RBF (AEJ-FT) are shown in Figure G3.



a) PRESS, $\lambda = 0$



b) PRESS, $\lambda \neq 0$



c) GCV, $\lambda \neq 0$

Figure G3 – Results of cross-validation (LOOCV) for RBFs (AEJ-FT) to Example 03

The greatest dispersion of data during the validation phase occurs for the RBF (AEJ-FT) with parameters optimized by the non-regularized PRESS. The presence of regularization caused a decrease in the dispersion in the regularized PRESS. The GCV shows a greater dispersion of validation data when compared to the PRESS with regularization.